1 Exercise 1

Base on the definition we can obtain:

$$\sum_{x_i=0,1} p(x_i) = \mu + (1-\mu) = 1$$

$$\mathbb{E}[x] = \sum_{x_i=0,1} x_i p(x_i) = 0 \cdot (1-\mu) + 1 \cdot \mu = \mu$$

$$\begin{aligned} var[x] &= \sum_{x_i=0,1} (x - \mathbb{E}[x])^2 p(x_i) \\ &= (0-\mu)^2(1-\mu) + (1-\mu)^2 \cdot \mu \\ &= \mu(1-\mu) \end{aligned}$$

$$H[x] = -\sum_{x_i=0,1} p(x_i) ln p(x_i) = -\mu ln \mu - (1-\mu) ln(1-\mu)$$

2 Exercise 2

$$\sum_{x_i=-1,1} p(x_i) = \frac{1-\mu}{2} + \frac{1+\mu}{2} = 1$$

$$\mathbb{E} = \sum_{x_i=-1,1} x_i p(x_i) = -1 \cdot \frac{1-\mu}{2} + 1 \cdot \frac{1+\mu}{2} = \mu$$

$$\begin{aligned} var[x] &= \sum_{x_i=-1,1} (x - \mathbb{E}[x])^2 \cdot p(x_i) \\ &= (-1-\mu)^2 \cdot \frac{1-\mu}{2} + (1-\mu)^2 \cdot \frac{1+\mu}{2} \\ &= 1 - \mu^2 \end{aligned}$$

$$H[x] = \sum_{x_i=-1,1} p(x_i) \cdot ln p(x_i) = -\frac{1-\mu}{2} \cdot ln\frac{1-\mu}{2} - \frac{1+\mu}{2} \cdot ln\frac{1+\mu}{2}$$

3 Exercise 3

Recall the formula for combinations:

$$C_N^m = \frac{N!}{m!(N-m)!}$$

We evaluate the left side first:

$$\begin{aligned}
C_N^m + C_N^{m-1} &= \frac{N!}{m!(N-m)!} + \frac{N!}{(m-1)!(N-m(m-1))!} \\
&= \frac{N!}{(m-1)!(N-m)!}(\frac{1}{m} + \frac{1}{N-m+1}) \\
&= \frac{(N+1)!}{m!(N+1-m)!} = C_{N+1}^m
\end{aligned}$$

To proof 2.263, here will prove a more general form:

$$(x+y)^N = \sum_{m=0}^{N} C_M^m \; x^m y^{N-m}$$

If we let y=1,(*) to reduce to 2.263. We will prove by induction. First, it is obvious when N=1, (*) holds. We assume that it holds for N, we will prove that it also holds for N + 1

$$\begin{aligned}
(x+y)^{N+1} = (x+y)\sum_{m=0}^{N} C_N^m x^m y^{N-m} \\
x\sum_{m=0}^{N} C_N^m x^m y^{N-m} + y\sum_{m=0}^{N} C_N^m x^m yN-m \\
\sum_{m=0}^{N} C_M^m x^{m+1} y^{N-m} + \sum_{m=0}^{N} C_N^m x^m y^{N+1-m} \\
\sum_{m=1}^{N+1} C_N^{m-1} x^m y^{N+1-m} + \sum_{m=0}^{N} C_N^m x^m y^{N+1-m} \\
\sum_{m=1}^{N} (C_N^{m-1} + C_N^m) x^m y^{N+1-m} + x^{N+1} + y^{N+1} \\
\sum_{m=1}^{N} C_{N+1}^m x^m y^{N+1-m} + x^{N+1} y^{N+1} \\
\sum_{m=0}^{N+1} C_{N+1}^m x^m y^{N+1-m}
\end{aligned}$$

4 Exercise 4

$$\mathbb{E}[m] = \sum_{m=0}^{N} mC_N^m \mu^m (1-\mu)^{N-m}$$
$$\sum_{m=1}^{N} mC_N^m \mu^m (1-\mu)^{N-m}$$
$$\sum_{m=1}^{N} \frac{N!}{(m-1)!(N-m)!} \mu^m (1-\mu)^{N-m}$$
$$N \cdot \mu \sum_{m=1}^{N} \frac{(N-1)!}{(m-1)!(N-m)!} \mu^{m-1} (1-\mu)^{N-m}$$
$$N \cdot \mu \sum_{m=1}^{N} C_{N-1}^{m-1} \mu^{m-1} (1-\mu)^{N-m}$$
$$N \cdot \mu \sum_{k=0}^{N-1} C_{N-1}^{k} \mu^{k} (1-\mu)^{N-1-k}$$
$$N \cdot \sum_{k=0}^{N-1} C_{N-1}^{k} \mu^{k} (1-\mu)^{N-1-k}$$
$$N \cdot \mu [\mu + (1-\mu)]^{N-1} = N\mu$$

Some details should be explained here. We note that $m = 0$ actually doesn't affect Expectation, so let the summation begin from $m = 1$, o.e (what we have done from the first step to the second step. Moreover, in the second to last step, we rewrite the subindex of the summation and what actually do is let $k = m - 1$. And in the last step, we have taken advantage of 2.264. The variance is straightforward once Expectation has been calculated

$$
\begin{aligned}
var[m] &= \mathbb{E}[m^2] - \mathbb{E}[m]^2 \\
&= \sum_{m=0}^{N} m^2 C_N^m \mu^m (1-\mu)^{N-m} - \mathbb{E}[m] \cdot \mathbb{E}[m] \\
&= \sum_{m=0}^{N} m^2 C_N^m \mu^m (1-\mu)^{N-m} - (N\mu) \cdot \sum_{m=0}^{N} m C_N^m \mu^m (1-\mu)^{N-m} \\
&= \sum_{m=1}^{N} m^2 C_N^m \mu^m (1-\mu)^{N-m} - N\mu \cdot \sum_{m=1}^{N} m C_N^m \mu^m (1-\mu)^{N-m} \\
&= \sum_{m=1}^{N} \frac{N!}{(m-1)!(N-m)!} \mu^m (1-\mu)^{N-m} - (N\mu) \cdot \sum_{m=1}^{N} m C_N^m \mu^m (1-\mu)^{N-m} \\
&= N\mu \sum_{m=1}^{N} m \frac{(N-1)!}{(m-1)!(N-m)!} \mu^{m-1} (1-\mu)^{N-m} - N\mu \cdot \sum_{m=1}^{N} m C_N^m \mu^M (1-\mu)^{N-m} \\
&= N\mu \sum_{m=1}^{N} m \mu^{m-1} (1-mu)^{N-m} (C_{N-1}^{m-1} - \mu C_n^m)
\end{aligned}
$$

Here we will use a little trick, $-\mu = -1 + (1-\mu)$ and take advantage of the property $C_N^m = C_{N-1}^m + C_{N-1}^{m-1}$

$$
\begin{aligned}
var[m] &= N\mu \sum_{m=1}^{N} m \mu^{m-1} (1-\mu)^{N-m} [C_{N-1}^{m-1} - C_N^m + (1-\mu)C \\
&= N\mu \sum_{m=1}^{N} m \mu^{m-1} (1-\mu)^{N-m} [(1-\mu)C_N^m + C_{N-1}^{m-1} - C_N^m] \\
&= N\mu \sum_{m=1}^{N} m \mu^{m-1} (1-\mu)^{N-m} [(1-\mu)C_N^m - C_{N-1}^m] \\
&= N\mu [\sum_{m=1}^{N} m \mu^{m-1} (1-\mu)^{N-m+1} C_N^m - \sum_{m=1}^{N} m \mu^{m-1} (1-\mu)^{N-m} C_{N-1}^m] \\
&= N\mu \{ \cdot N(1-\mu)[\mu + (1-\mu)^{N-1}] - (N-1)(1-\mu)[\mu + (1-\mu)]^{N-2} \} \\
&= N\mu(1-\mu)
\end{aligned}
$$

5 Exercise 5

Hints have already been given in the description. and lets make a little improvement by introducing $t = y + x$ and $x = tu$

$$\begin{cases} x = tu \\ y = t(1-u) \\ t = x + y \\ u = \frac{x}{x+y} \end{cases}$$

Note that $t \in [0, +\infty], \mu \in (0, 1)$ and that when we change variables in an integral we will introduce a redundant term called the Jacobian Determiannt.

$$\frac{\partial(x, y)}{\partial(\mu, t)} = \begin{vmatrix} \frac{\partial x}{\partial \mu} & \frac{\partial x}{\partial t} \\ \frac{\partial y}{\partial \mu} & \frac{\partial y}{\partial t} \end{vmatrix} = \begin{vmatrix} t & \mu \\ -t & 1-\mu \end{vmatrix} = t$$

Now we can calculate the integral:

$$\begin{aligned} \Gamma(a)\Gamma(b) &= \int_0^{+\infty} exp(-x)x^{a-1}dx \int_0^{+\infty} exp(-y)y^{b-1}dy \\ &= \int_0^{+\infty}\int_0^{+\infty} exp(-x)x^{a-1}exp(-y)^{b-1}dydx \\ &= \int_0^{+\infty}\int_0^{+\infty} exp(-x-y)x^{a-1}y^{b-1}dydx \\ &= \int_0^{1}\int_0^{+\infty} exp(-t)(t\mu)^{a-1}(t(1-\mu))^{b-1}tdtdu \\ &= \int_0^{+\infty} exp(-t)t^{a+b-1}dt \cdot \int_0^1 \mu^{a-1}(1-\mu)^{b-1}du \\ &= \Gamma(a+b) \cdot \int_0^1 \mu^{a-1}(1-\mu)^{b-1}du \\ \int_0^1 \mu^{a-1}(1-\mu)^{b-1}d\mu &= \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \end{aligned}$$

6 Exercise 6

We will solve this problem base on defintion"

$$
\begin{aligned}
\mathbb{E}[\mu] &= \int_0^1 \mu Beta(\mu|a,b)d\mu \\
&= \int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^a(1-\mu)^{b-1}d\mu \\
&= \frac{\Gamma(a+b)\Gamma(a+1)}{\Gamma(a+1+b)\Gamma(a)}\int_0^1 \frac{\Gamma(a+1+b)}{\Gamma(a+1)\Gamma(b)}\mu^a(1-\mu)^{b-1}d\mu \\
&= \frac{\Gamma(a+b)\Gamma(a+1)}{\Gamma(a+1+b)\Gamma(a)}\int_0^1 Beta(\mu|a+1,b)d\mu \\
&= \frac{\Gamma(a+b)}{\Gamma(a+1+b)} \cdot \frac{\Gamma(a+1)}{\Gamma(a)} \\
&= \frac{a}{a+b}
\end{aligned}
$$

Where we have taken advantage of the property $\Gamma(z+1) = z\Gamma(z)$. Now we tackle the variance problem:

$$
\begin{aligned}
\mathbb{E}[\mu^2] &= \int_0^1 \mu^2 Beta(\mu|a,b)d\mu \\
&= \int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a+1}(1-\mu)^{b-1}d\mu \\
&= \frac{\Gamma(a+b)\Gamma(a+2)}{\Gamma(a+2+b)\Gamma(a)}\int_0^1 \frac{\Gamma(a+2+b)}{\Gamma(a+2)\Gamma(b)}\mu^{a+1}(1-\mu)^{b-1}d\mu \\
&= \frac{\Gamma(a+b)\Gamma(a+2)}{\Gamma(a+2+b)\Gamma(a)}\int_0^1 Beta(\mu|a+2,b)d\mu \\
&= \frac{\Gamma(a+b)}{\Gamma(a+2+b)} \cdot \frac{\Gamma(a+2)}{\Gamma(a)} \\
&= \frac{a(a+1)}{(a+b)(a+b+1)}
\end{aligned}
$$

We can now use the formula $var[\mu] = \mathbb{E}[\mu^2] - \mathbb{E}[\mu]^2$:

$$
\begin{aligned}
var[\mu] &= \frac{a(a+1)}{(a+b)(a+b+1)} - (\frac{a}{a+b})^2 \\
&= \frac{ab}{(a+b)^2(a+b+a)}
\end{aligned}
$$

7 Exercise 7

The MLE for μ i.e can be written as:

$$\mu_{ML} = \frac{m}{m+l}$$

We need to prove that:

$$\frac{a}{a+b} \leq \frac{(m+a)}{m+a+l+b} \leq \frac{m}{(m+l)}$$

So we have:

$$\lambda \frac{a}{a+b} + (1-\lambda)\frac{m}{m+l} = \frac{m+a}{m+a+l+b} \text{where } \lambda = \frac{a+b}{m+l+a+b}$$

8 Exercise 8

We solve this problem in in base definition:

$$\begin{aligned}
\mathbb{E}_y[\mathbb{E}_x[x|y]] &= \int \mathbb{E}_x[x|y]p(y)dy \\
&= \int (\int xp(x|y)dx)p(y)dy \\
&= \int \int xp(x|y)p(y)dxdy \\
&= \int \int xp(x,y)dxdy \\
&= \int xp(x)dx = \mathbb{E}[x]
\end{aligned}$$

2.271 is complicated and we will calculate evey term separately

$$\begin{aligned}
\mathbb{E}[var_x[x|y]] &= \int var_x[x|y]p(y)dy \\
&= \int (\int (x - \mathbb{E}_x[x|y])^2 p(x|y)dx)p(y)dy \\
&= \int \int (x - \mathbb{E}_x[x|y)^2 p(x,y)dxdy \\
&= \int \int (x^2 - 2x\mathbb{E}_x[x|y] + \mathbb{E}_x[x|y]^2)p(x,y)dxdy \\
&= \int \int x^2 p(x)dx - \int \int 2x\mathbb{E}_x[x|y]p(x,y)dxdy + \int \int (\mathbb{E}_x[x|y]^2)p(y)dy
\end{aligned}$$

We can further simplify the second term:

$$\begin{aligned}\int\int 2x\mathbb{E}_x[x|y]p(x,y)dxdy &= 2\int \mathbb{E}_x[x|y](\int xp(x,y)dx)dy \\ &= 2\int \mathbb{E}_x[x|y]p(y)(\int xp(x|y)dx)dy \\ &= 2\int \mathbb{E}_x[x|y]^2 p(y)dy\end{aligned}$$

Therefore, we obtain the simple expression for the first term on the right side of 2.271.

$$\mathbb{E}_y[var_x[x|y]] = \int\int x^2 p(x)dx - \int\int \mathbb{E}_x[x|y]^2 p(y)dy$$

Then we process for the second term:

$$\begin{aligned}var_y[\mathbb{E}_x[x|y]] &= \int (\mathbb{E}_x[x|y] - \mathbb{E}_y[\mathbb{E}_x[x|y]])^2 p(y)dy \\ &= \int (\mathbb{E}_x[x|y] - \mathbb{E}[x])^2 p(y)dy \\ = \int \mathbb{E}_x[x|y]^2 p(y)dy - 2\int \mathbb{E}_x\mathbb{E}_x[x|y]p(y)dy &+ \int \mathbb{E}[x]^2 p(y)dy \\ = \int \mathbb{E}_x[x|y]^2 p(y)dy - 2\mathbb{E}[x]\int &\mathbb{E}_x[x|y]p(y)dy + \mathbb{E}[x^2]\end{aligned}$$

Then the following same proceure, we deal with the second term of the equation above:

$$2\mathbb{E}[x]\cdot\int \mathbb{E}_x[x|y]p(y)dy = 2\mathbb{E}[x]\cdot\mathbb{E}_y[\mathbb{E}_x[x|y]] = 2\mathbb{E}[x]^2$$

Therefore we obtain the simple expression for the second term on the right side of 2.271:

$$var_y[\mathbb{E}_x[x|y]] = \int \mathbb{E}_x[x|y]^2 p(y)dy - \mathbb{E}[x]^2$$

We can add together and get:

$$\mathbb{E}_y[var_x[x|y]] = var_y[\mathbb{E}_x[x|y]] = \mathbb{E}[x^2] - \mathbb{E}[x]^2 = var[x]$$

9 Exercise 9

This problem is complicated enough already. Lets being by performing the integral 2.272 over μ_{M-1}

$$p_{M-1}(\mu, m...\mu_{M-2}) = \int_0^{1-\mu-m-..-\mu_{M-2}} C_M \prod_{k=1}^{M-1} \mu_k^{\alpha_k-1}(1-\sum_{j=1}^{M-1})^{\alpha_M-1} d\mu_{M-1}$$
$$= C_M \prod_{k=1}^{M-2} \mu_k^{\alpha_k-1} \int_0^{1-\mu-m..-\mu_{M-2}} \mu_{M-1}^{\alpha_{M-1}-1}(1-\sum_{j=1}^{M-1})^{\alpha_M-1} d\mu_{M-1}$$

We can change variables by:

$$t = \frac{\mu_{M-1}}{1-\mu-m-..-\mu_{M-2}}$$

The reason we do so is that $\mu_{M-1} \in [0, 1, -\mu - m - ... - \mu M - 2]$ by making this change we can that $t \in [0, 1]$. Then we can further simplify the expression.

$$p_{M-1} = C_M \prod_{k=1}^{M-2} \mu_k^{\alpha_k-1}(1-\sum_{j=1}^{M-2})^{\alpha_{M-1}+\alpha_M-1} \int_0^1 \frac{\mu_{M-1}^{\alpha_{M-1}-1}(1-\sum_{j=1}^{M-1}\mu_j)^{\alpha_M-1}}{1-\mu-m...-\mu_{M-2}^{\alpha_{M-1}+\alpha_M-2}} dt$$

10 Exercise 10

Based on definition of *Expectation* and (2.38) we can write:

$$\mathbb{E}[\mu_j] = \int \mu_j Dir(\mu|\alpha) d\mu$$
$$= \int \mu_j \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_0)\Gamma(\alpha_2)\Gamma(\alpha_K)} \prod_{k=1}^{K} \mu_k^{\alpha_k-1} d\mu$$
$$= \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1)\Gamma(\alpha_2)...\Gamma(\alpha_K} \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)...\Gamma(\alpha_{j-1})\Gamma(\alpha_{j+1})...\Gamma(\alpha_K}{\Gamma(\alpha_0+1)}$$
$$= \frac{\Gamma(\alpha_0)\Gamma(\alpha_j+2)}{\Gamma(\alpha_j)\Gamma(\alpha_0+2)} = \frac{\alpha_j(\alpha_j+1)}{\alpha_0(\alpha_0+1)}$$

And so we obtain:

$$var[\mu_j] = \mathbb{E}[\mu_j^2] - \mathbb{E}[\mu_j]^2 = \frac{\alpha_j(\alpha_j+1)}{\alpha_0(\alpha_0+1)} - (\frac{\alpha_j}{\alpha_0})^2 = \frac{\alpha_j(\alpha_0-\alpha_j)}{\alpha_0^2(\alpha_0+1)}$$

The covariance also follows:

$$cov[\mu_j\mu_l] = \int (\mu_j - \mathbb{E}[\mu_j])(\mu_l - \mathbb{E}[\mu_l])Dir(\mu|\alpha)d\mu$$

$$= \int (\mu_j\mu_l - \mathbb{E}[\mu_j]\mu_l - \mathbb{E}[\mu_l]\mu_j + \mathbb{E}[\mu_j]\mathbb{E}[\mu_l])Dir(\mu|\alpha)d\mu$$

$$= \frac{\Gamma(\alpha_0)\Gamma(\alpha_j+1)\Gamma(\alpha_l+1)}{\Gamma(\alpha_j)\Gamma(\alpha_l)\Gamma(\alpha_0+2)} - 2\mathbb{E}[\mu_j]\mathbb{E}[\mu_l] + \mathbb{E}[\mu_j][\mu_l]$$

$$= \frac{\alpha_j\alpha_l}{\alpha_0(\alpha_0+1)} - \mathbb{E}[\mu_j]\mathbb{E}[\mu_l]$$

$$= \frac{\alpha_j\alpha_l}{\alpha_0(\alpha_0+1)} - \frac{\alpha_j\alpha_l}{\alpha_0^2}$$

$$= -\frac{\alpha_j\alpha_l}{\alpha_0^2(\alpha_0+1)}(j \neq l)$$

11 Exercise 11

Based on the definition of *Expectation* and 2.38 we can first denote:

$$\frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_K)} = K(\alpha)$$

Then it is possible to write:

$$\frac{\partial Dir(\mu|\alpha)}{\partial\alpha_j} = \frac{\partial(K(\alpha)\prod_{i=1}^{K}\mu_i^{\alpha_i-1})}{\partial\alpha_j}$$

$$= \frac{\partial K(\alpha)}{\partial\alpha_j}\prod_{i=1}^{K}\mu_i^{\alpha_i-1} + K(\alpha)\frac{\partial\prod_{i=1}^{K}\mu_i^{\alpha_i-1}}{\partial\alpha_j}$$

$$= \frac{\partial K(\alpha)}{\partial\alpha_j}\prod_{i=1}^{K}\mu_i^{\alpha_i-1} + ln\mu_j \cdot Dir(\mu|\alpha)$$

Then let us perform the integral to both sides:

$$\int \frac{\partial Dir(\mu|\alpha)}{\partial\alpha_j} = \int \frac{\partial K(\alpha)}{\partial\alpha_j}\prod_{i=1}^{K}\mu_i^{\alpha_i-1}d\mu + \int ln\mu_j \cdot Dir(\mu|\alpha)d\mu$$

The left side can be further simplified as:

$$leftside = \frac{\partial \int Dir(\mu|\alpha)d\mu}{\partial \alpha_j} = \frac{\partial 1}{\partial \alpha_j} = 0$$

And the right side can be simpliefied even further:

$$\begin{aligned} right = \frac{\partial K(\alpha)}{\partial_j} \int \prod_{i=1}^{K} \mu_i^{\alpha_i - 1} d\mu + \mathbb{E}[ln\mu_j] \\ = \frac{\partial K(\alpha)}{\partial \alpha_j} \frac{1}{K(\alpha)} + \mathbb{E}[ln\mu_j] \\ = \frac{\partial lnK(\alpha)}{\partial \mu_j} + \mathbb{E}[\ln \mu_j] \end{aligned}$$

Therefore we obtain:

$$\begin{aligned} \mathbb{E}[ln\mu_j] = -\frac{\partial lnK(\alpha)}{\partial \alpha_j} \\ = -\frac{\partial \{ln\Gamma(\alpha_0) - \sum_{i=1}^{K} ln\Gamma(\alpha_i)\}}{\partial \alpha_j} \\ = \frac{\partial ln\Gamma(\alpha_j)}{\partial \alpha_j} - \frac{\partial ln\Gamma(\alpha_0)}{\partial \alpha_j} \\ = \frac{\partial ln\Gamma(\alpha_j)}{\partial \alpha_j} - \frac{\partial ln\Gamma(\alpha_0)}{\partial \alpha_0} \frac{\partial \alpha_0}{\partial \alpha_j} \\ == \frac{\partial ln\Gamma(\alpha_j)}{\partial \alpha_j} - \frac{\partial ln\Gamma(\alpha_0)}{\partial \alpha_0} \\ = \psi(\alpha_j) - \psi(\alpha_0) \end{aligned}$$

Therefore the problem is solved:

12 Exercise 12

We have:

$$\int_a^b \frac{1}{b-a} dx = 1$$

It is straightforward that it is normalized. Then we calculate its mean:

$$\mathbb{E}[x] = \int_a^b x\frac{1}{b-a} = \frac{x^2}{2(b-a)}\Big|_a^b = \frac{a+b}{2}$$

Then we can also compute the variance:

$$var[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2 = \int_a^b \frac{x^2}{b-a}dx - (\frac{a+b}{2})^2 = \frac{x^3}{3(b-a)}\Big|_a^b (\frac{a+b}{2})^2$$

Hence we obtain:

$$var[x] = \frac{(b-a)^2}{12}$$

13 Exercise 13

First recall:

$$ln\frac{p(x)}{q(x)} = \frac{1}{2}ln(\frac{|L|}{|\Sigma|}) + \frac{1}{2}(x-m)^T L^{-1}(x-m) - \frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)$$

If $x \approx p(x) = \mathcal{N}(\mu|\Sigma)$ we can take advantage of the following properties:

$$\begin{aligned} \int p(x)dx &= 1 \\ \mathbb{E}[x] = \int xp(x)dx &= \mu \\ \mathbb{E}[(x-a)^T A(x-a)] = tr(A\Sigma) + (\mu-a)^T A(\mu-a) \end{aligned}$$

We obtain the folowing:

$$\begin{aligned} KL = \int [\frac{1}{2}ln\frac{|L|}{|\Sigma|} - \frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) + \frac{1}{2}(x-m)^T L^{-1}(x-m)]p(x)dx \\ = \frac{1}{2}ln\frac{|L|}{|\Sigma|} - \frac{1}{2}\mathbb{E}[(x-\mu)\Sigma^{-1}(x-\mu)^T] + \frac{1}{2}\mathbb{E}[(x-m)^T L^{-1}(x-m)] \\ = \frac{1}{2}ln\frac{|L|}{|\Sigma|} - \frac{1}{2}tr(I_D) + \frac{1}{2}(\mu-m)^T L^{-1}(\mu-m) + \frac{1}{2}tr(L^{-1}\Sigma) \\ = \frac{1}{2}[ln\frac{|L|}{|\Sigma|} - D + tr(L^{-1}\Sigma) + (m-\mu)^T L^{-1}(m-\mu)] \end{aligned}$$

14 Exercise 14

We let $g(x)$ be a Gaussian PDF with mean and mean μ and variance Σ and f(x) is an abritrary PDF with the same mean and covariance.

$$
\begin{aligned}
0 \leq KL(f||g) = \int f(x) ln \frac{g(x)}{f(x)} dx = -H(f) - \int f(x) ln g(x) dx \\
= \int f(x) ln(\frac{1}{(2\pi)^{D/2}} \frac{1}{(|\Sigma|)^{1/2}} exp(\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)) dx) \\
= \int f(\boldsymbol{x}) ln\{\frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}}\} dx + \int f(\boldsymbol{x})[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})] d\boldsymbol{x} \\
= ln\{\frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}}\} - \frac{1}{2}\mathbb{E}[(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \\
= \ln\{\frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} - \frac{1}{2} tr(I_D)\} \\
= -\{\frac{1}{2} ln|\Sigma| + \frac{D}{2}(1 + ln(2\pi))\} \\
-H(g)
\end{aligned}
$$

We take advantage of the two properties of PDF $f(\boldsymbol{x})$ wuth mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$ as iliste below:

$$
\int f(x) dx = 1 \mathbb{E}[(x-a)^T A(x-a)] = tr(A\Sigma) + (\mu - a)^T A(\mu - a)
$$

We have proved that an abritrary PDF f(x) with the same mean and variance as a Gaussian PDF g(x) cannot be greaer than that of Gaussain PDF.

15 Exercise 15

let: $x \approx p(x) = \mathcal{N}(\mu|\Sigma)$

$$
\begin{aligned}
H[\boldsymbol{x}] &= \int p(\boldsymbol{x}) lnp(\boldsymbol{x}) d\boldsymbol{x} \\
= -\int p(\boldsymbol{x}) ln\{\frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} exp[-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})] dx \\
= -\int p(\boldsymbol{x}) ln\{\frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}}\} dx - \int f(\boldsymbol{x})[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})] dx \\
= ln\{\frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/1}}\} + \frac{1}{2}\mathbb{E}[(x-\mu)^T \Sigma^{-1}(x-\mu)] \\
= -ln\{\frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}}\} + \frac{1}{2} tr\{I_D\} \\
= \frac{1}{2} ln|\Sigma| + \frac{D}{2}(1+ln(2\pi))
\end{aligned}
$$

Where we have taken advantage of:

$$
\int p(x)dx = 1
$$

$$
\mathbb{E}[(x-a)^T A(x-a)] = tr(A\Sigma) + (\mu-a)^T A(\mu-a)
$$

16 Exercise 16

Let us condier a more general conclusion about the PDF of the summation of two independent random variables. We denote two random variable and Y. Their summation Z=X+Y, is still a random variable. We also denote $f(\cdot)$ as PDF and $F(\cdot)$ as CDF. We can obtain:

$$
F_Z(z) = P(Z < z) = \int\int_{x+y\leq z} f_{X,Y}(x,y) dx dy
$$

Where z represents an arbitrary real number. We rewrite the double integral into an iterated integral:

$$
F_Z(z) = \int_{-+\infty}^{\infty} \Big[\int_{-\infty}^{z-y} f_{X,Y}(x,y) dx\Big] dy
$$

We fix z and y, and then make a change of variable to $x = u - y$ to the following integral:

$$
F_Z(z) = \int_{-\infty}^{+\infty} \Big[\int_{-\infty}^{z} f_{X,Y}(u,-x,y) du\Big] dy
$$

We can rearrange to get:

$$F_Z(z) = \int_{-\infty}^{z} \Big[\int_{-\infty}^{+\infty} f_{X,Y}(u, -y, y)dy \Big] du$$

Compare the equation with the defintion of CDF:

$$F_Z(z) \int_{-\infty}^{z} f_Z(u)du$$

We can obtain:

$$f_Z(u) = \int_{-\infty}^{+\infty} f_{X,Y}(u, -y, y)dy$$

We can treat X and Y as independent and simplify:

$$f_Z(u) = \int_{-\infty}^{+\infty} f_X(u - y) f_Y(y)dy$$

Until no we have proved that the PDF of the. summation of the two independent random variables is the convolution of the PDF of them. Hence it is straightforward to see that in this problem, where random variable x is the summation of random variable x1 and x2, the PDF of x should be the convolution of the PDF of x1 and x2. To find the entropy of x, we will use a simple method, taking advantage of 2.113 to 2.117.

$$\begin{aligned} p(x_2) &= \mathcal{N}(\mu_2, \tau_2^{-1}) \\ p(x|x_2) &= \mathcal{N}(\mu_1 + x_2, \tau_1^{-1}) \end{aligned}$$

We make analogies:x_2 in this problem x in 2.113, x in this problem to y 2.114. Hen by using 2.115 we can obtain p(x) is still a normal distribution, and since the entropy of a Gaussian is fully decided by its varaiance, there is no need to calulate the man. Still by using 2.115 and the variance of x is $\tau_1^{-1} + \tau_2^{-1}$ which finally gives its entropy:

$$H[x] = \frac{1}{2}[1 + ln2\pi(\tau_1^{-1}\tau_2^{-1})]$$

17 Exercise 17

This is an extension of 1.14. The same procedure can be used here. We suppose an arbritray precision matrix Λ can be written as $\Lambda^S + \Lambda^A$ where they satisfy:

$$\Lambda_{ij}^S = \frac{\Lambda_{ij} + \Lambda_{ji}}{2}, \qquad \Lambda_{ij}^A = \frac{\Lambda_{ij} - \Lambda_{ji}}{2}$$

Hence it is straightforward the $\Lambda_{ij}^S = \Lambda_{ji}^S$ and $\Lambda_{ij}^A = -\Lambda_{ji}^A$. If we expand. the quadratic form of the exponent, we will obtain:

$$\begin{aligned}(x-\mu)\Lambda(x-\mu) &= \sum_{i=1}^{D}\sum_{j=1}^{D}(x_i-\mu_i)\Lambda_{ij}(x-\mu_j) \\ &= \sum_{i=1}^{D}\sum_{j=1}^{D}(x_i-\mu_i)\Lambda_{ij}^S(x_j-\mu_i)\Lambda_{ij}^A(x_j-\mu_j) \\ &= \sum_{i=1}^{D}\sum_{j=1}^{D}(x_i-\mu_i)\Lambda_{ij}^S(x_j-\mu_j)\end{aligned}$$

Therefore we can assume a precision matrix is symmetric, and so is the covariance matrix

18 Exercise 18

We will just follow the hint:

$$\overline{\Sigma\mu_i} = \overline{\lambda_i\mu_i} => \Sigma\overline{\mu_i} = \overline{\lambda_i}\overline{\mu_i}$$

Where we have take advantage of the fact the Σ is a real matrix, i.e $\overline{\Sigma} = \Sigma$. Then using that Σ is a symmetric matrix $\Sigma^T = \Sigma$:

$$\begin{aligned}\overline{\mu_i}^T\Sigma\mu_i &= \overline{\mu_i}^T(\Sigma\mu_i) = \overline{\mu_i}^T(\lambda_i\, u_i) = \lambda_i\overline{\mu_i}^T\mu_i \\ \overline{\mu_i}^T\Sigma\mu_i &= (\Sigma\overline{\mu_i})^T\mu_i = (\overline{\lambda_i}\overline{\mu_i})^T\mu_i = \overline{\lambda_i}^T\overline{\mu_i}^T\mu_i\end{aligned}$$

Since $u \neq 0$ we have $\overline{u_i}^T u_i \neq 0$. Thus $\Lambda_i^T = \overline{\lambda_i^T}$, which means λ_i is real. Next we will prove that two eigen vectors corresponding to different eigenvalues are orthogonal!

$$\lambda_i < \mu, \mu_j >=< \lambda_i \mu_i, \mu_j >=< \Sigma\mu_i, \mu_j >=< \mu_i, \Sigma^T \mu_j >= \lambda < \mu_i . \mu_j >$$

Where we have taken advantage of $\Sigma^T = \Sigma$, and for an abritray real matrix A and vector x, y we have:

$$< Ax, y >=< x, A^T y >$$

Provided $\lambda_i \neq \lambda_j$ we have $< \mu_i, \mu_j >= 0$ and both u's are orthogonal. And then if we perform normalization on every eigenvector to forces its Eucleadian norm to equal 1 it becomes straightforward.

19 Exercise 19

For every $N \times N$ real symmetric matrix, the eigenvalues are real and the eigen vectors can be chosen such that the are orthogonal to each other. Thus a real symmetric matrix Σ can be decomposed as $\Sigma = U\Lambda U^T$, where U is an ortho matrix, and Λ is a diagonal matrix, whose entries are the eigenvalues of A. Hence we have:

$$\Sigma x = U\Lambda U^T x = U\Lambda \begin{bmatrix} \mu_1^T x \\ . \\ . \\ . \\ \mu_D^T x \end{bmatrix} = U \begin{bmatrix} \lambda_1 \mu_1^T x \\ . \\ . \\ . \\ \lambda_D \mu_D^T x \end{bmatrix} = (\sum_{k=1}^{D} \lambda_k \mu_k \mu_k^T) x$$

20 Exercise 20

Since $\mu_1, \mu_2 ... \mu_D$ can constitute a basis for $\mathbb{E}^D$, we can make a projection for a:

$$a = a_a \mu_1 + a_2 \mu_2 + ... + a_D \mu_D$$

We substitute the expression above into $a^T \sigma a$ taking advantage of the property : $u_1, u)j = 1$ if $i = j$, otherwise 0:

$$\begin{aligned} a^T \Sigma a &= (a_1 u_1 + a_2 u_2 + ... + a_D u_D)^T \Sigma (a_1 u_1 + a_2 u_2 + ... + a_D u_D) \\ &= (a_1 u_1^T + a_2 u_2^T + ... + a_D u_D^T) \Sigma (a_1 u_1 + a_2 u_2 + ... + a_D u_D) \\ &= (a_1 u_1^T + a_2 u_2^T + ... + a_D u_D^T)(\alpha_1 \lambda_1 u_1 + \alpha_2 \lambda_2 u_2 + ... + \alpha_D \lambda_D u_D) \end{aligned}$$

Since α is real, the expression above will be strictly positive for any non-zero α if all eigenvalues are strictly positive. It is also clear that i an eigenvalue, λ_i is zero or negative there will exists a vector α for which this expression will be greater than 0. This that a real symmetric matrix has eigenvectors which are all strictly postive is a sufficient and necesscary condition for the matrix to be positive definite.

21 Exercise 21

It is straightforward for a symmetrix matrix Λ of size $D \times D$ when the lower triangular part is decided, the whole matrix will be deicded due to symmetry. Hence, the number of independent parameters is $D + (D-1) + ... + 1$ which comes down to $D(D+1)/2$

22 Exercise 22

Suppose A is a symmetrix matrix, and we need to prove that A^{-1} is also symmetric (i.e $A^{-1} = (A^{-1})^T$, since the identity matrix I is also symmetric we have:

$$AA^{-1} = (AA^{-1})^T$$

Recall the identity $AB^T = B^T A^T$ holds for any abritrary matrix A or B:

$$AA^{-1} = (A^{-1})^T A^T$$

Since $A = A^T$, we substitute the right side:

$$AA^{-1} = (A^{-1})^T A$$

And note that $AA^{-1} = A^{-1}A = I$. We can rearrange the order on the left side to get:

$$A^{-1}A = (A^{-1})^T A$$

Finally, by multiplying A^{-1} to both side we can obtain:

$$A^{-1}AA^{-1} = (A^{-1})^T AA^{-1}$$

Using $AA^{-1} = I$ we will get:

$$A^{-1} = (A^{-1})^T$$

23 Exercise 23

Let's rewrite the problem. What the problem whats us ti prove is that $(x-\mu)^T\Sigma^{-1}(x-\mu) = r^2$, where r^2 is a constant. We will have the volume of a hyperelilipsoid decided by the equation above equaling $V_D|\Sigma|^{1/2}r^D$. Note that the center f this hyperelipsoid l will be the mean vector μ and a translation operation won't change its volume, thus we only need to prove that the volume of a hyper elipsoid decided by $x^T\Sigma^{-1}x = r^2$ whose center is 0.

This problem can be viewed in two parts. First lets discuss about V_D, the volume of a unit sphere in dimension D. The xpression V_D has already been given in the oslution to 1.18. Lets recall:

$$V_D = \frac{S_D}{D} = \frac{2\pi^{D/2}}{\Gamma(\frac{D}{2}+1)}$$

And also in the procedure, we shoe that D is a dimensional sphere with radius r (i.e $x^Tx = r^2$ has volume $V(r) = V_D r^D$. We move a step forward and perform a linear transform using the matrix $\Sigma^{1/2}$ (i.e $y^Ty = r^2$ where $y = \Sigma^{1/2}x$. After the linear transformation we actually get a hyperelipsoid whose center locates at 0. and its volume is given by multiplying $V(r)$ with the determinant of the transformation matrix, which gives $|\Sigma|^{1/2}V_D r^D$

24 Exericise 24

We can just follow the hint:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \times \begin{bmatrix} M & -MB^{-1} \\ -D^{-1}CM & D^{-1} + D^{-1}CMBD^{-1} \end{bmatrix}$$

The result can also be partitioned into four blocks. The block located at the top left can be expessed as:

$$AM - BD^{-1}CM = (A - BD^{-1}C)(A - BD^{-1}C)^{-1} = I$$

And the top right can be written as:

$$-AMBD^{-1} + BD^{-1} + BD^{-1}CMBD^{-1} = (I - AM + BD^{-1}CM)DB^{-1} = 0$$

The bottom left can be written as:

$$CM - DD^{-1}CM = 0$$

The bottom right is:

$$-CMBD^{-1} + DD^{-1} + DD^{-1}CMDD^{-1} = I$$

25 Exercise 25

We will take advantage of 2.94 to 2.98. Lets first begin by grouping x_a and x_b together, then we can rewrite what has been given as:

$$x = \begin{pmatrix} x_{a,b} \\ x_c \end{pmatrix} \quad \mu = \begin{pmatrix} \mu_{a,b} \\ \mu_c \end{pmatrix} \quad \Sigma = \begin{bmatrix} \Sigma_{(a,b)(a,b)} & \Sigma_{(a,b)c} \\ \Sigma_{(a,b)c} & \Sigma_{cc} \end{bmatrix}$$

Then we can take advantage of 2.98:

$$p(x_{a,b}) = \mathcal{N}(x_{a,b}|\mu_{a,b}, \Sigma_{(a,b)(a,b)})$$

Where we have define:

$$\mu_{a,b} = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} \quad \Sigma_{(a,b)(a,b)} = \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix}$$

Since now we have obtained the joint distribution of x_a and x_b, we will take advantage of 2.96 and 2,97 to obtain the conditional distribution, which gives:

$$p(x_a|x_b) = \mathcal{N}(x|\mu_{a|b}, \Lambda_{aa}^{-1})$$

Where we have defined

$$\mu_{a|b} = \mu_a - \Lambda_{aa}^{-1}\Lambda_{ab}(x_b - \mu_b)$$

Which can be written in full:

$$p(x_a|x_b) = \mathcal{N}(x|\mu_a - \Lambda_{aa}^{-1}\Lambda_{ab}(x_b - \mu_b), \Lambda_{aa}^{-1})$$

26 Exercise 26

Recall the woodbury inversion matrix formula:

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$$

Multiply both sides by $(A + BCD)$:

$$\begin{aligned}
&= (A + BCD)[A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}] \\
&= AA^{-1} - AA^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1} + BCDA^{-1} - BCDA^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1} \\
&= I - B(C^{-1} + DA^{-1}B)DA^{-1} + BCDA^{-1} + B(C^{-1} + DA^{-1}B)^{-1}DA^{-1} - BCDA^{-1} \\
&= I
\end{aligned}$$

Where we have taken advantage of:

$$\begin{aligned}
&= -BCDA^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1} \\
&= -BC(-C^{-1} + C^{-1} + DA^{-1}B)(C^{-1} + DA^{-1}B)^{-1}DA^{-1} \\
&= (-BC)(-C^{-1})(C^{-1} + DA^{-1}B)^{-1}DA^{-1} + (-BC)(C^{-1} + DA^{-1})(C^{-1} + DA^{-1}B)DA^{-1} \\
&= B(C^{-1} + DA^{-1}B)DA^{-1} = BCDA^{-1}
\end{aligned}$$

Here will also directly calculate the inverse matrix instead to give another solution. Let's first begin by introducing two useful formulas:

$$\begin{aligned}
(I + P)^{-1} &= (I + P)^{-1}(I + P - P) \\
&= I - (I + P)^{-1}P
\end{aligned}$$

And since

$$P + PQP = P(I + QP) = (I + PQ)P$$

The second formula is:

$$(I + PQ)^{-1}P = P(I + QP)^{-1}$$

And now let's directly calculate $(A + BCD)^{-1}$:

$$\begin{aligned}
(A + BCD)^{-1} &= [A(I + A^{-1}BCD)]^{-1} \\
&= (I + A^{-1}BCD)^{-1}A^{-1} \\
&= [I - (I + A^{-1} + A^{-1}BCD)^{-1}A^{-1}BCD]A^{-1} \\
&= A^{-1} - (I + A^{-1}BCD)^{-1}A^{-1}BCDA^{-1}
\end{aligned}$$

27 Exercise 27

The same thing was used in 1.10 and we use the same setup here:

$$
\begin{aligned}
\mathbb{E}[x+z] &= \int\int (x+z)p(x,z)dxdz \\
&= \int\int (x+z)p(x)p(z)dxdz \\
&= \int\int xp(x)p(z)dxdz + \int\int zp(x)p(z)dxdz \\
&= \int(\int p(z)dz)xp(x)dx + \int(\int (x)dx)zp(z)dz \\
&= \int xp(x)dx + \int zp(z)dz \\
&= \mathbb{E}[x] + \mathbb{E}[z]
\end{aligned}
$$

And for the covariance matrix:

$$
cov[x+z] = \int\int (x+z-\mathbb{E}[x+z])(x+z-\mathbb{E}[x+z])^T p(x,z)dxdz
$$

28 Exercise 28

This is a form of problem 2.94, but we treat $x$in 2.94 as z in this problem, x_a in 2.94 as x in this problem, and x_b as y. The expressions can be written as

$$
z = \begin{pmatrix} x \\ y \end{pmatrix} \quad \mathbb{E}(z) = \begin{pmatrix} \mu \\ A\mu + b \end{pmatrix} \quad cov(z) = \begin{bmatrix} \Lambda^{-1} & \Lambda^{-1}A^T \\ A\Lambda^{-1} & L^{-1} + A\Lambda^{-1}A^T \end{bmatrix}
$$

Recall:

$$
\begin{aligned}
p(x) &= \mathcal{N}(x|\mu, \Lambda^{-1}) \\
p(y|x) &= \mathcal{N}(y|\mu_{y|x}, \Lambda_{yy}^{-1})
\end{aligned}
$$

Where Λ_{yy} can be obained by the right bottom part of 2.104, which gives $\Lambda_{yy} = L^{-1}$. And the conditional mean can be written as:

$$
\mu_{y|x} = A\mu + L - L^{-1}(-LA)(x-\mu) = Ax + L
$$

29 Exercise 2.29

Let's fist calculate the top left block:

$$topleft = \Big[(\Lambda + A^T LA) - (-A^T L)(L^{-1})(-LA)\Big]^{-1} = \Lambda^{-1}$$

The top right block is:

$$-\Lambda^{-1}(-A^T L)L^{-1} = \Lambda^{-1}A^T$$

The bottom left block is:

$$= L^{-1}(-LA)\Lambda^{-1} = A\Lambda^{-1}$$

The bottom right block is:

$$L^{-1} + L^{-1}(-LA)\Lambda^{-1}(-A^T L)L^{-1} = L^{-1}A\Lambda^{-1}A^T$$

30 Exercise 30

$$\begin{bmatrix} \Lambda^{-1} & \Lambda^{-1}A^T \\ A\Lambda^{-1} & L^{-1} + A\Lambda^{-1}A^T \end{bmatrix} \begin{bmatrix} \Lambda\mu - A^T Lb \\ Lb \end{bmatrix} = \begin{bmatrix} \mu \\ A\mu + b \end{bmatrix}$$

31 Exercise 31

According to the problem we can write:

$$p(x) = \mathcal{N}(x|\mu_x, \Sigma_x) \quad p(y|x) = \mathcal{N}(y|\mu_z + x, \Sigma_z)$$

By comparing the expression above we can write:

$$p(y) = \mathcal{N}(y|\mu_x + \mu_z.\Sigma_x + \Sigma_z)$$

32 Exercise 32

Let's make the problem even clearer. The deduction in the main text, 2.101 to 2.110, firstly denote a new random variable z corresponding to the joint distribution, and then by completing the square according to z, 2.103, obtain the precision matrix R by comparing 2.103 with the PDF of a multinomial gaussian, and then it takes the inverse of precision to obtain the covariance matrix, and finally it obtains the linear term to calculate the mean. In this problem we are asked to solve the problem from another perspective: we need to write the joint distribution $p(x,y)$ and then perform integration over x to obtain a marginal distribution over $p(y)$. Let's being by writing the quad of the exp of x,y:

$$-\frac{1}{2}(x-\mu)^T\Lambda(x-\mu)-\frac{1}{2}(y-Ax-b)^TL(y-Ax-b)$$

We extract those terms involding x:

$$\begin{aligned}&= -\frac{1}{2}x^T(\Lambda+A^TLA)x+x^T[\Lambda\mu+A^TL(y-b)]+const\\ = -\frac{1}{2}(x-m)^T(\Lambda+A^TLA)(x-m)&+\frac{1}{2}m^T(\Lambda+A^TLA)m+const\end{aligned}$$

Where we have define m, for brevity, as:

$$m=(\Lambda+A^TLA)^{-1}[\Lambda\mu+A^TL(y-b)]$$

Now if we perform integration over x, we will see that the first term vanishes to a constant, and we can extract the terms including y from the reamaining parts:

$$\begin{aligned}= -\frac{1}{2}y^T\Big[L-LA(\Lambda+A^TLA)^{-1}A^TL\Big]y\\ +y^T\Big[[L-LA(\Lambda+A^TLA)^{-1}A^TL]b\\ +LA(\Lambda+A^TLA)^{-1}\Lambda\mu\Big]\end{aligned}$$

33 Exercise 33

According to bayes formula, we can write $p(x|y)=\frac{p(x,y)}{p(y)}$, where we have already known the joint distribution $p(x,y)$, and the marginial distributino $p(y)$. We can follow the same procedure in 2.32, where we obtain the covariance matrix from the quadratic term and then obtain the mean from the linear term. The details are omitted here.

34 Exercise 34

Let's follow the hint by firstly calculating the derivative of 2.118, with respect to Σ and let it equal to 0:

$$-\frac{N}{2}\frac{\partial}{\partial\Sigma}ln|\Sigma|-\frac{1}{2}\frac{\partial}{\partial\Sigma}\sum_{n=1}^{N}(x_n-\mu)^T\Sigma^{-1}(x_n-\mu)=0$$

By using C.28 the first term can be reduced to:

$$-\frac{N}{N}\frac{\partial}{\partial\Sigma}ln|\Sigma| = -\frac{N}{2}(\Sigma^{-1})^T = -\frac{N}{2}\Sigma^{-1}$$

Provided with the result that the optimal covariance matrix is the sample covariance, we denoe sample matrix S as:

$$S = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu)(x_n - \mu)^T$$

We rewrtie the second term:

$$\begin{aligned} secondterm = -\frac{1}{2}\frac{\partial}{\partial\Sigma}\sum_{n=1}^{N}(x_n - \mu)^T\Sigma^{-1}(x_n - \mu) \\ = -\frac{N}{2}\frac{\partial}{\partial\Sigma}Tr[\Sigma^{-1}S] \\ = \frac{N}{2}\Sigma^{-1}S\Sigma^{-1} \end{aligned}$$

Where we have taken advantage of the following property, combined with the fact that S and Σ is symmetric (Note: this property can be found in the Matrix cookbook).

$$\frac{\partial}{\partial X}Tr(AX^{-1}B) = -(X^{-1}BAX^{-1})^T = -(X^{-1})^T A^T B^T (X^{-1})^T$$

And we obtain:

$$-\frac{N}{2}\Sigma^{-1} + \frac{N}{2}\Sigma^{-1}S\Sigma^{-1} = 0$$

35 Exericse 36

We first begin by proving 2.123:

$$\mathbb{E}[\mu_{ML}] = \frac{1}{N}\mathbb{E}[\sum_{n=1}^{N}] = \frac{1}{N}\cdot N\mu = \mu$$

Where we have taken advantage of the fact that x_n is independently and identically distributed (i.i.d). Then we use the expression in 2.122:

$$\mathbb{E}[\Sigma_{ML}] = \frac{1}{N}\mathbb{E}[\sum_{n=1}^{N}(x_n - \mu_{ML})(x_n - \mu_{ML})^T]$$
$$= \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[(x_n - \mu_{ML})(x_n - \mu_{ML})^T]$$
$$= \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[(x_n - \mu_{ML})(x_n - \mu_{ML})^T]$$
$$= \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[x_n x_n^T - 2\mu_{ML}x_n^T + \mu_{ML}\mu_{ML}^T$$
$$= \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[x_n x_n^T] - 2\frac{1}{2}\sum_{n=1}^{N}\mathbb{E}[\mu_{ML}x_n^T] + \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\mu_{ML}\mu_{ML}^T]$$

By using 2.291, the first term will be:

$$firstterm = \frac{1}{N}\cdot N(\mu\mu^T + \Sigma) = \mu\mu^T + \Sigma$$

The second term can be reduced even further:

$$secondterm = -2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\mu_{ML}x_n^T]$$
$$= -2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\frac{1}{N}(\sum_{m=1}^{N}x_m)x_n^T$$
$$= -2\frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{M}\mathbb{E}[x_m x_n^T]$$
$$= -2\frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{M}(\mu\mu^T + I_{nm}\Sigma)$$
$$= -2\frac{1}{N^2}(N^2\mu\mu^T + N\Sigma)$$
$$= -2(\mu\mu^T + \frac{1}{N}\Sigma)$$

Similariy, the third term will equal to:

$$thirdterm = \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\mu_{ML}\mu_{ML}^T]$$

$$= \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[(\frac{1}{N}\sum_{j=1}^{N}x_j)\cdot(\frac{1}{N}\sum_{i=1}^{N}x_i)]$$

$$= \frac{1}{N^3}\sum_{n=1}^{N}\mathbb{E}[(\sum_{j=1}^{N}x_j)\cdot(\sum_{-i=1}^{N}x_i)]$$

$$= \frac{1}{N^3}\sum_{n=1}^{N}(N^2\mu\mu^T + N\Sigma)$$

$$= \mu\mu^T + \frac{1}{N}\Sigma$$

Finally we can cmobine those three terms which gives:

$$\mathbb{E}[\Sigma_{ML}] = \frac{N-1}{N}\Sigma$$

36 Exercise 36

Lets follow the hint. However, we first find the sequential expression based on definition, which will make the latter process on finding coeffcient α_{N_1} more easily. SUppose we have N observations in total, and then we can write:

$$\sigma_{ML}^{2(N)} = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu_{ML}^{(N)})^2$$

$$= \frac{1}{N}\Big[\sum_{n=1}^{N-1}(x_n - \mu_{ML}^{(N)}) + (x_N - \mu_{ML}^{(N)})^2\Big]$$

$$= \frac{N-1}{N}\frac{1}{N-1}\sum_{n=1}^{N}(x_n - \mu_{ML}^{(N)})^2 + \frac{1}{N}(x_N - \mu_{ML}^{(N)})^2$$

$$= \frac{N-1}{N}\sigma_{ML}^{2(N-1)} + \frac{1}{N}(x_N - \mu_{ML}^{(N)})^2$$

And then let us write the expression of $sigma_{ML}$:

$$\frac{\partial}{\partial\sigma^2}\Big[\frac{1}{N}m\sum_{n=1}^{N}lnp(x_n|\mu,\sigma)\Big]\Big|_{\sigma_{ML}}$$

By obtaining the summation and the derivative, and letting $N \to +\infty$ we can obtain:

$$\lim_{N\to+\infty}\frac{1}{N}\sum_{n=1}^{N}\frac{\partial}{\partial\sigma^2}lnp(x_n|\mu,\sigma) = \mathbb{E}_x[\frac{\partial}{\partial\sigma^2}lnp(x_n|\mu,\sigma)]$$

Comparing it with 2.127, we can obtain the sequential formula to estimate σ_{ML}"

$$\begin{aligned}\sigma_{ML}^{2(N)} &= \sigma_{ML}^{2(N-1)} + \alpha_{N-1}\frac{\partial}{\partial\sigma_{ML}^{2(N-1)}}lnp(x_N|\mu_{ML}^{N},\sigma_{ML}^{(N-1)})(*) \\ &= \sigma_{ML}^{2(N-1)} + \alpha_{N-1}\Big[-\frac{1}{2\sigma_{ML}^{2(N-1)}} + \frac{(x_N-\mu_{ML}^{(N)})^2}{2\sigma_{ML}^{4(N-1)}}\Big]\end{aligned}$$

Where we have used $\sigma_{ML}^{2(N)}$ to represent the Nth esimation of σ_{ML}^2. If we choose:

$$\alpha_{N-1} = \frac{2\sigma_{ML}^{4(N-1)}}{N}$$

Then we obtain:

$$\sigma_{ML}^{2(N)} = \sigma_{ML}^{2(N-1)} + \frac{1}{N}\Big[-\sigma_{ML}^{2(N-1)} + (x_N-\mu_{ML}^{N})^2\Big]$$

37 Exercise 37

$$\Sigma_{ML}^{(N)} = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu_{ML}^{(N)})(x_n - \mu_{ML}^{(N)})^T$$
$$= \frac{1}{N}\Big[\sum_{n=1}^{N-1}(x_n - \mu_{ML}^{(N)})(x_n - \mu_{ML}^{(N)})^T + (x_N - \mu_{ML}^{(N)})(x_N - \mu_{ML}^{(N)})^T\Big]$$
$$= \frac{N-1}{N}\Sigma_{ML}^{(N-1)} + \frac{1}{N}(x_N - \mu_{ML}^{(N)})(x_N - \mu_{ML}^{(N)})^T$$
$$= \Sigma_{ML}^{(N-1)} + \frac{1}{N}\Big[(x_N - \mu_{ML}^{(N)})(x_N - \mu_{ML}^{(N)})^T$$
$$= \Sigma_{ML}^{(N-1)} + \frac{1}{N}\Big[(x_N - \mu_{ML}^{(N)})(x_N - \mu_{ML}^{(N)})^T - \Sigma_{ML}^{(N-1)}\Big]$$

If we use Monro-Robbins we can obtain:

$$\Sigma_{ML}^{(N)} = \Sigma_{ML}^{(N-1)} + \alpha_{N-1}\frac{\partial}{\partial\Sigma_{ML}^{(N-1)}}lnp(x_N|\mu_{ML}^{(N)}, \Sigma_{ML}^{(N-1)})$$
$$\Sigma_{ML}^{(N-1)} + alpha_{N-1}$$

Still working on this problem....

38 Exercise 38

It is straightforward. We focus on the exponential term of the posterior distribution:

$$-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n - \mu)^2 - \frac{1}{2\sigma^2}(\mu - \mu_0)^2 = -\frac{1}{2\sigma_N^2}(\mu - \mu_N)^2$$

We can rewrite the following terms in this way:

$$quadterm = -(\frac{N}{2\sigma^2} + \frac{1}{2\sigma_0^2})\mu^2$$
$$linterm = (\frac{\sum_{n=1}^{N} x_n}{\sigma^2} + \frac{\mu_0}{\sigma_0^2})\mu$$

We also rewrite the side regard to th μ, and hence we will obtain:

$$-(\frac{N}{2\sigma^2}+\frac{1}{2\sigma_0^2})\mu^2=-\frac{1}{2\sigma_N^2}\mu^2, \quad (\frac{\sum_{n=1}^N x_n}{\sigma^2}+\frac{\mu_0}{\sigma_0^2})\mu=\frac{\mu_N}{\sigma_N^2}\mu$$

Then we can get:

$$\frac{1}{\sigma_N^2}=\frac{1}{\sigma_0^2}+\frac{N}{\sigma^2}$$

And with the prior knowledge that $\sum_{n=1}^N x_n = N \cdot \mu_{ML}$ we can write:

$$\begin{aligned}\mu_N &= \sigma_N^2 \cdot (\frac{\sum_{n=1}^N x_n}{\sigma^2}+\frac{\mu_0}{\sigma_0^2}) \\ &= (\frac{1}{\sigma^2}+\frac{N}{\sigma^2})^{-1}\cdot(\frac{N\mu_{ML}}{\sigma^2}+\frac{\mu_0}{\sigma_0^2}) \\ &= \frac{\sigma_0^2\sigma^2}{\sigma^2+N\sigma_0^2}\cdot\frac{N\mu_{ML}\sigma_0^2+\mu_0\sigma^2}{\sigma\sigma_0^2} \\ &= \frac{\sigma^2}{N\sigma+0^2+\sigma^2}\mu_0+\frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\mu_{ML}\end{aligned}$$

39 Exercise 39

Lets follow the hint!

$$\frac{1}{\sigma_N^2}=\frac{1}{\sigma_0^2}+\frac{N}{\sigma^2}=\frac{1}{\sigma_0^2}+\frac{N-1}{\sigma^2}+\frac{1}{\sigma^2}=\frac{1}{\sigma_{N-1}^2}+\frac{1}{\sigma^2}$$

However it is complicated to derive a sequential formulat for μ_N directly. Base on 2.142 we see that the denominator in 2.141 can be eliminated if we multiply $1/\sigma_N^2$ on both sides. Therefore will derive a sequential formulate for μ_N/σ_N^2 instead.

$$\begin{aligned}\frac{\mu_N}{\sigma_N^2} &= \frac{\sigma^2+N\sigma_0^2}{\sigma_0^2\sigma^2}(\frac{\sigma^2}{N\sigma_0^2+\sigma^2}\mu_0+\frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\mu_{ML}^{(N)}) \\ &= \frac{\mu_{N-1}}{\sigma_{N-1}^2}+\frac{X_N}{\sigma^2}\end{aligned}$$

Another possible solution would be to solve by completing the square:

$$-\frac{1}{2\sigma^2}(x_N-\mu)^2-\frac{1}{2\sigma_{N-1}^2}(\mu-\mu_{N-1})^2=-\frac{1}{2\sigma_N^2}(\mu-\mu_N)^2$$

By comparing the quadratic and linear term regarding u, we can obtain:

$$\frac{1}{\sigma_N^2}=\frac{1}{\sigma^2}+\frac{1}{\sigma_{N-1}^2}$$

And:

$$\frac{\mu_N}{\sigma_N^2}=\frac{x_N}{\sigma^2}+\frac{\mu_{N-1}}{\sigma_{N-1}^2}$$

40 Exercise 40

Based on Bayes Theorem, we can write:

$$p(\mu|X)\propto p(X|\mu)p(\mu)$$

We focus on the exponential term on the right side and then rearrange it:

$$\begin{aligned}
right &= \Big[\sum_{n=1}^{N}-\frac{1}{2}(x_n-\mu)^T\Sigma^{-1}(x_n-\mu)\Big]-\frac{1}{2}(\mu-\mu_0)^T\Sigma_0^{-1}(\mu-\mu_0)\\
&= \Big[\sum_{n=1}^{N}-\frac{1}{2}(x_n-\mu)^T\Sigma^{-1}(x_n-\mu)\Big]-\frac{1}{2}(\mu-\mu_0)^T\Sigma_0^{-1}(\mu-\mu_0)\\
&= -\frac{1}{2}\mu(\Sigma_0^{-1}+N\Sigma^{-1})\mu+\mu^T(\Sigma_0^{-1}\mu_0+\Sigma^{-1}\sum_{n=1}^{N}x_n)+const
\end{aligned}$$

Where 'const' represents all the constant terms indepdent of μ. According to the quadratic term, we can obtain the posterior covariance matrix.

$$\Sigma_N^{-1}=\Sigma_0^{-1}+N\Sigma^{-1}$$

Then using the linear term we can obtain:

$$\Sigma_N^{-1}\mu_N = (\Sigma_0^{-1}\mu_0 + \Sigma^{-1}\sum_{n=1}^{N} x_n)$$

Finally we obtain the posterior mean:

$$\mu_N = (\Sigma_0^{-1} + N\Sigma^{-1})^{-1}(\Sigma_0^{-1} + \Sigma^{-1}\sum_{n=1}^{N} x_n)$$

Which can also be written as:

$$\mu_N = (\Sigma_0^{-1} + N\Sigma^{-1})^{-1}(\Sigma_0^{-1}\mu_0 + \Sigma^{-1}N\mu_{ML})$$

41 Exercise 41

Let's compute the integral of 2.145 over λ

$$\begin{aligned}
\int_0^{+\infty} \frac{1}{\Gamma(\alpha)} b^a \lambda^{a-1} exp(-b\lambda) d\lambda = \frac{b^a}{\Gamma(\alpha)} \int_0^{+\infty} \lambda^{\alpha-1} exp(-b\lambda) d\lambda \\
= \frac{b^a}{\Gamma(\alpha)} \int_0^{+\infty} (\frac{u}{b})^{a-1} exp(-u) \frac{1}{b} du \\
= \frac{1}{\Gamma(\alpha)} \int_0^{+\infty} u^{a-1} exp(-u) du \\
= \frac{1}{\Gamma(\alpha)} \cdot \Gamma(\alpha) = 1
\end{aligned}$$

Where have taken advantage of the change of variable $b\lambda = u$ and:

$$\Gamma(x) = \int_0^{+\infty} u^{x-1} e^{-u} du$$

42 Exercise 42

We first calculate its mean:

$$\begin{aligned}
\int_0^{+\infty} \lambda \frac{1}{\Gamma(\alpha)} b^a \lambda^{a-1} exp(-b\lambda) d\lambda &= \frac{a}{\Gamma(\alpha)} \int_0^{+\infty} \lambda^a exp(-b\lambda) d\lambda \\
&= \frac{b^a}{\Gamma(\alpha)} \int_0^{+\infty} (\frac{u}{b})^a exp(-u) \frac{1}{b} du \\
&= \frac{1}{\Gamma(\alpha) \cdot b} \int_0^{+\infty} u^a exp(-u) du \\
&= \frac{1}{\Gamma(a) \cdot b} \cdot \Gamma(a+1) = \frac{a}{b}
\end{aligned}$$

Where we have taken advantage of the property $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$. Then we calculate $\mathbb{E}[\lambda^2]$:

$$\begin{aligned}
\int_0^{+\infty} \lambda^2 \frac{1}{\Gamma(\alpha)} b^a \lambda^{a-1} exp(-b\lambda) d\lambda &= \frac{b^a}{\gamma(\alpha)} \int_0^{+\infty} \lambda^{a+1} exp(-b\lambda) d\lambda \\
&= \frac{b^a}{\Gamma(\alpha)} \int_0^{+\infty} (\frac{u}{b}) a^{a+1} exp(-u) \frac{1}{b} du \\
&= \frac{1}{\Gamma(\alpha) \cdot b^2} int_0^{+\infty} u^{a+1} exp(-u) du \\
&= \frac{1}{\Gamma(\alpha) \cdot b^2} \cdot \Gamma(\alpha+2) = \frac{\alpha(\alpha+1)}{b^2}
\end{aligned}$$

Therefore according ot $var[\lambda] = \mathbb{E}[\lambda^2] - \mathbb{E}[\lambda]^2$

Forhte mode of a gamma distribution, we need to find where the max of the PDf occurs, and hence we will calculate the derivative of the gamma distribution with respect to λ.

$$\frac{d}{d\lambda}\Big[\frac{1}{\Gamma(\alpha)} b^a \lambda^{a-1} exp(-b\lambda)\Big] = [(a-1) - b\lambda] \frac{1}{\Gamma(\alpha)} b^a \lambda^{a-2} exp(-b\lambda)$$

43 Exercise 43

Lets first calculate the following integral:

$$\int_{-\infty}^{+\infty} exp(-\frac{|x|^q}{2\sigma^2})dx = 2\int_{-\infty}^{+\infty} exp(-\frac{x^q}{2\sigma^2})dx$$
$$= 2\int_{0}^{+\infty} exp(-u)\frac{(2\sigma^2)^{1/2}}{q}u^{\frac{1}{q}-1}du$$
$$= 2\frac{(2\sigma^2)^{1/q}}{a}\int_{0}^{+\infty} exp(-u)u^{\frac{1}{q}-1}dx$$
$$= 2\frac{(2\sigma^2)^{1/q}}{q}\Gamma(\frac{1}{q})$$

$$lnp(t, X, w, \sigma^2) = \sum_{n=1}^{N} lnp(y(x_n, w) - t_n|\sigma^2, q)$$
$$= -\frac{1}{2\sigma^2}\sum_{n=1}^{N}|y(x_n, w) - t_n|^q - \frac{N}{q}ln(2\sigma^2) + const$$

44 Exercise 2.44

Here we use a simple method to solve this proble, by taking advantage os 2.152 and 2.153. Br writing the prior distribution in the form 2.153 as $p(u, \lambda|\beta, c, d)$ we can easily obtain the posterior distribution as:

$$p(\mu, \lambda|X) \propto p(X|\mu, X) \cdot p(\mu, \lambda)$$
$$\propto \Big[\lambda^{1/2}exp(-\frac{\lambda\mu^2}{2})\Big]^{N+\beta} exp\Big[(c + \sum_{n=1}^{N} x_n)\lambda\mu - (d + \sum_{n=1}^{N}\frac{x_n^2}{2})\lambda\Big]$$

45 Exercise 45

The wishart distribution is $\mathcal{W}(\Lambda|W, v)$:

$$p(X|\mu,\Lambda) \propto |\Lambda|^{N/2} exp[\sum_{n=1}^{N} -\frac{1}{2}(x_n-\mu)^T\Lambda(x_n-\mu)]$$

$$S = \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)^T$$

$$p(X|\mu,\Lambda) \propto |\Lambda|^{N/2} exp(-\frac{1}{2}Tr(S\Lambda))$$

Note S and Λ are symmetric and this $Tr(S\Lambda) = Tr((S\Lambda)^T) = Tr(\Lambda^T S^T) = Tr(\Lambda S)$

46 Exercise 46

$$p(x|\mu,a,b) = \int_0^{+\infty} \mathcal{N}(x|\mu,\tau^{-1})\Gamma(\tau|a,b)d\tau$$

$$= \frac{b^a}{\Gamma(\alpha)}(\frac{1}{2\pi})^{1/2}\int_0^{+\infty} \tau^{\alpha-1/2} exp\Big[-b\tau - \frac{\tau}{2}x - 2^2\Big]d\tau$$

And if we make a change of variable to $z = \tau[b + (x\mu)^2/2]$ we can simplify even more:

$$= \frac{b^a}{\Gamma(\alpha)}(\frac{1}{2\pi}^{1/2}\int_0^{+\infty}\Big[\frac{z}{b+(x-u)^2/2}\Big]^{\alpha-1/2} exp(-z)\frac{1}{b+(x-\mu)^2/2}dz$$

$$= \frac{b^a}{\Gamma(\alpha)}(\frac{1}{2\pi})^{1/2}\Big[\frac{1}{b+(x-\mu)^2/2}\Big]^{\alpha+1/2}\int_0^{\infty} z^{\alpha-1/2}exp(-z)dz$$

$$= \frac{b^a}{\Gamma(\alpha)}(\frac{1}{2\pi})^{1/2}\Big[b + \frac{(x-\mu)^2}{2}\Big]^{-\alpha-1/2}\Gamma(\alpha+1/2)$$

47 Exercise 47

We focus on the dependency of 2.159:

$$St(x|\mu,\lambda,v) \propto \Big[1 + \frac{\lambda(x-\mu)^2}{v}\Big]^{-v/2-1/2}$$

$$\propto exp\Big[\frac{-v-1}{2}(\frac{\lambda(x-\mu)^2}{v} + O(v^{-2})\Big]$$

$$\approx exp\Big[-\frac{\lambda(x-\mu)^2}{x}\Big] \quad (v \to \infty)$$

48 exercise 48

The same steps in 2.46 can be used here:

$$St(x|\ \mu,\Lambda,v) = \int_0^{+\infty} \mathcal{N}(x|\mu,(\eta\Lambda)^{-1}\cdot\Gamma(\eta|\frac{v}{2},\frac{v}{2})d\eta$$
$$= \int_0^{+\infty} \frac{1}{(2\pi)^{D/2}}|\eta\Lambda|^{1/2}exp\{-\frac{1}{2}(x-\mu)^T(\eta\Lambda)(x-\mu)-\frac{v\eta}{2}\}frac1\Gamma(v/2)(\frac{v}{2})^{v/2}\eta^{v/2-1}d\eta$$
$$= \frac{(v/2)^{n/2}|\Lambda|^{1/2}}{(2\pi)^{D/2}\Gamma(v/2)}\int_0^{+\infty} exp\{-\frac{1}{2}(x-\mu)^T(\eta\Lambda)(x-\mu)-\frac{v\eta}{2}\}\eta^{D/2+v/2-1}d\eta$$

Where we have taken advantage of the property: $|\eta\Lambda| = \eta^D|\Lambda|$ and if we denote:

$$\Delta^2 = (x-\mu)^T\Lambda(x-\mu) \quad z = \frac{\eta}{2}(\Delta^2+v)$$

We can further reduce the expression to:

$$St(x|\ \mu,\Lambda,v) = \frac{(v/2)^{n/2}|\Lambda|^{1/2}}{(2\pi)^{D/2}\Gamma(v/2)}\int_0^{+\infty} exp(-z)(\frac{2z}{\Delta^2+v})^{D/2+v/2-1}\cdot\frac{2}{\Delta^2+v}dz$$
$$= \frac{(v/2)^{n/2}|\Lambda|^{1/2}}{(2\pi)^{D/2}\Gamma(v/2)}(\frac{2}{\Delta^2+v})^{D/2+v/2-1}\int_0^{+\infty} exp(-z)\cdot z^{D/2+v/2-1}dz$$
$$= \frac{(v/2)^{n/2}|\Lambda|^{1/2}}{(2\pi)^{D/2}\Gamma(v/2)}(\frac{2}{\Delta^2+v})^{D/2+v/2-1}\Gamma(D/2+v/2)$$

49 Exercise 49

Verity students t is convolution of gaussian with gamma

First we note that if and only if $x=\mu, \Delta^2$ equals to 0, so that $Stu(x|\mu,\Lambda,v)$ acheives its maximum. In other words, the mode of $Stu(x|\mu,\Lambda,v)$ is μ.

$$
\begin{aligned}
\mathbb{E} = \int_{x\in R^D} Stu(x|\mu,\Lambda,v)\cdot x dx \\
= \int_{x\in R^D}\Big[\int_0^{+\infty}\mathcal{N}(x|\mu,(n\Lambda)^{-1})\cdot\Gamma(\eta|\frac{v}{2},\frac{v}{2})d\eta x\Big]dx \\
= \int_{x\in R^D}\int_0^{+\infty} x\mathcal{N}(x|\mu,(n\Lambda)^{-1})\cdot\Gamma(\eta|\frac{v}{2},\frac{v}{2})d\eta dx \\
= \int_0^{+\infty}\Big[\int_{x\in R^D} x\mathcal{N}(x|\mu,(n\Lambda)^{-1})dx\cdot\Gamma(\eta|\frac{v}{2},\frac{v}{2})d\eta x\Big] \\
= \int_0^{+\infty}\Big[\mu\cdot\Gamma(\eta|\frac{v}{2},\frac{v}{2})\Big]d\eta \\
= \mu\int_0^{+\infty}\Gamma(\eta|\frac{v}{2},\frac{v}{2})d\eta
\end{aligned}
$$

We can calculte $\mathbb{E}[xx^T]$:

$$
\begin{aligned}
\mathbb{E}[xx^T] = \int_{x\in R^D} St(x|\ \mu,\Lambda,v)\cdot xx^T dx \\
= \int_{x\in R^D}\Big[\int_0^{+\infty}\mathcal{N}(x|\mu,(n\Lambda)^{-1})\cdot\Gamma(\eta|\frac{v}{2},\frac{v}{2})d\eta xx^T\Big]dx \\
= \int_{x\in R^D}\int_0^{+\infty} xx^T\mathcal{N}(x|\mu,(n\Lambda)^{-1})\cdot\Gamma(\eta|\frac{v}{2},\frac{v}{2})d\eta dx \\
= \int_0^{+\infty}\Big[\int_{x\in R^D} xx^T\mathcal{N}(x|\mu,(n\Lambda)^{-1})dx\cdot\Gamma(\eta|\frac{v}{2},\frac{v}{2})d\eta x\Big] \\
= \int_0^{+\infty}\Big[\mathbb{E}[\mu\mu^T]\cdot\Gamma(\eta|\frac{v}{2},\frac{v}{2})\Big]d\eta \\
= \int_0^{+\infty}\Big[\mu\mu^T+(\eta\Lambda)^{-1}]\cdot\Gamma(\eta|\frac{v}{2},\frac{v}{2})\Big]d\eta \\
= \mu\mu^T+\int_0^{+\infty}(\eta\Lambda)^{-1}]\cdot\Gamma(\eta|\frac{v}{2},\frac{v}{2})d\eta \\
= \mu\mu^T+\Lambda^{-1}\frac{1}{\Gamma(v/2)}(\frac{v}{2})^{v/2}\int_0^{+\infty}\eta^{v/2-2}exp(-\frac{v}{2}\eta)d\eta
\end{aligned}
$$

If we denote $z=\frac{v\eta}{2}$ we can:

$$
\begin{aligned}
&= \mu\mu^T + \Lambda^{-1}\frac{1}{\Gamma(v/2)}(\frac{v}{2})^{v/2}\int_0^{+\infty}(\frac{2z}{v})^{v/2-2}exp(-z)\frac{2}{v}dz \\
&= \mu\mu^T + \Lambda^{-1}\frac{1}{\Gamma(v/2)}(\frac{v}{2})\int_0^{+\infty}z^{v/2-2}exp(-z)dz \\
&= \mu\mu^T + \Lambda^{-1}\frac{\Gamma(v/2-1)}{\Gamma(v/2)}\cdot\frac{v}{2} \\
&= \mu\mu^T + \frac{v}{v-2}\Lambda^{-1}
\end{aligned}
$$

Recall the property $\Gamma(x+1) = x\Gamma(x)$

50 Exercise 50

$$
\begin{aligned}
St(x|\ \mu,\Lambda,v) &\propto \Big[1+\frac{\Delta^2}{v}\Big]^{-D/2-v/2} \\
\propto &= exp\Big[(-D/2-v/2)ln(1+\frac{\Delta^2}{v})\Big] \\
&\propto exp\Big[-\frac{D+v}{2}\cdot(\frac{\Delta^2}{v}+O(v^{-2}))\Big] \\
&\approx exp(\frac{-\Delta^2}{2}) \quad (v\to\infty)
\end{aligned}
$$

Where we used the Taylor expansion: $ln(1+\epsilon) = \epsilon + O(\epsilon)^2$

51 Exercise 51

We first prove 2.177. since we have $exp(iA) - exp(-iA) = 1$ and $exp(iA) = cosA + isin(A)$ we can obtain:

$$
(cosA + isin(A)\cdot(cosA - isinA) = 1
$$

Which gives $cos^2A + sin^2A = 1$

$$
\begin{aligned}
cos(A-B) &= \Re[exp(i(A-B))] \\
&= \Re[exp(iA)\ exp(iB)] \\
&= \Re\frac{cosA + isinB}{cosB + isinB} \\
&= \Re[(cosA + isinA)(cosB - isinB)] \\
&= cosAcosB + sinAsinB
\end{aligned}
$$

This is quite similar to:

$$
\begin{aligned}
sin(A-B) &= \Im[exp(i(A-b))] \\
&= sinAcosA - cosAsinB
\end{aligned}
$$

52 exercise 52

$$
\begin{aligned}
exp[mcos(\theta-\theta_0)] &= exp[m[1 - \frac{(\theta-\theta_0)^2}{2} + O((\theta-\theta_0)^4)]] \\
&= exp(m) \cdot exp[-m\frac{(\theta-\theta)^2}{2}] \cdot exp[-mO((\theta-\theta)^4)]
\end{aligned}
$$

skipping the rest!