2.13. Evaluate the Kullback-Leibler divergence between two Gaussians $p(x) = \mathcal{N}(x|\mu, \Sigma^2)$ and $q(x) = \mathcal{N}(x|m, L^2)$

Sol. we know $KL(p\|q) = -\int p(x)\ln q(x)\,dx + \int p(x)\ln p(x)\,dx$

$$-\int p(x)\ln q(x)\,dx = -\int \mathcal{N}(\mu,\Sigma^2)\ln q(x)\,dx$$

$$q(x) = \frac{1}{(2\pi)^{\frac{D}{2}}}\cdot\frac{1}{|L|^{\frac12}} e^{-\frac12(x-m)^T L^{-1}(x-m)}$$

$$\therefore \ln q(x) = -\frac12(x-m)^T L^{-1}(x-m) - \frac{D}{2}\ln(2\pi) - \frac12\ln|L|$$
$$= -\frac12\left[(x-m)^T L^{-1}(x-m) + D\ln 2\pi + \ln|L|\right]$$

$$\therefore -\int \mathcal{N}(\mu,\Sigma^2)\ln q(x)\,dx = \frac12\int \mathcal{N}(\mu,\Sigma^2)\left[[(x-m)^T L^{-1}(x-m)] + D\ln 2\pi + \ln|L|\right]dx$$
$$= \frac12\int \mathcal{N}(\mu,\Sigma^2)(x-m)^T L^{-1}(x-m)\,dx + \frac12\int \mathcal{N}(\mu,\Sigma^2)\,D\cdot\ln 2\pi\,dx + \frac12\int \mathcal{N}(\mu,\Sigma^2)\ln|L|\,dx$$
$$= \frac12\int \mathcal{N}(\mu,\Sigma^2)(x-m)^T L^{-1}(x-m)\,dx + \frac{D}{2}\ln 2\pi + \frac12\ln|L|$$
$$= \frac12\int \mathcal{N}(\mu,\Sigma^2)\left[x^T L^{-1}x - m^T L^{-1}x - x^T L^{-1}m + m^T L^{-1}m\right]dx + \frac{D}{2}\ln 2\pi + \frac12\ln|L|$$
$$= \frac12\int \boxed{\mathcal{N}(\mu,\Sigma^2)\,x^T L^{-1}x}\,dx - \frac12 m^T L^{-1}\mu - \frac12\mu^T L^{-1}m + \frac12 m^T L^{-1}m + \frac{D}{2}\ln 2\pi + \frac12\ln|L|$$

Tr(a) = a
a is a constant

$$= \frac12\boxed{Tr\left[L^{-1}(\mu\mu^T+\Sigma)\right]} - \frac12 m^T L^{-1}\mu - \frac12\mu^T L^{-1}m + \frac12 m^T L^{-1}m + \frac{D}{2}\ln 2\pi + \frac12\ln|L|$$

$$* \int \mathcal{N}(\mu,\Sigma^2)\ln p(x)\,dx = -\frac12 Tr\left[\Sigma^{-1}(\mu\mu^T+\Sigma)\right] + \frac12\mu^T\Sigma^{-1}\mu + \frac12\mu^T\Sigma^{-1}\mu - \frac12\mu^T\Sigma^{-1}\mu - \frac{D}{2}\ln 2\pi - \frac12\ln|\Sigma|$$

$$\therefore KL(p\|q) = \left[-\int p(x)\ln q(x) + \int p(x)\ln p(x)\right]dx = \frac12\ln\frac{|L|}{|\Sigma|} + \frac12\left(Tr\left[L^{-1}(\mu\mu^T+\Sigma)\right] - m^T L^{-1}\mu - \mu^T L^{-1}m + m^T L^{-1}m\right) - \frac{D}{2}$$
$$= \frac12\left(\ln\frac{|L|}{|\Sigma|} + Tr\left[L^{-1}(\mu\mu^T+\Sigma)\right] - \mu^T L^{-1}m - m^T L^{-1}\mu + m^T L^{-1}m - D\right)$$

2.25. In Section 2.3.1 and 2.3.2, we considered the conditional and marginal distributions for a multivariate Gaussian. More generally, we can consider a partitioning of the components of X into three parts X_a, X_b and X_c, with a ~~cros~~ corresponding partitioning of the mean vector μ and of the covariance matrix Σ in the form.

$$\mu = \begin{bmatrix}\mu_a\\ \mu_b\\ \mu_c\end{bmatrix} \qquad \Sigma = \begin{bmatrix}\Sigma_{aa} & \Sigma_{ab} & \Sigma_{ac}\\ \Sigma_{ba} & \Sigma_{bb} & \Sigma_{bc}\\ \Sigma_{ca} & \Sigma_{cb} & \Sigma_{cc}\end{bmatrix}$$

By making use of the results of Section 2.3, find an experession for the conditional distribution $p(X_a|X_b)$ in watch X_c has been marginalized out.

Sol. $p(x_a, x_b) = \int p(x_a, x_b, x_c)\, dx_c$
$= \mathcal{N}(x_a, x_b \mid \mu, \Sigma)$
where $\mu = \begin{bmatrix} \mu_a \\ \mu_b \end{bmatrix}$ $\Sigma = \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix}$

~~$p(x_a|x_b) = \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(x_b - \mu_b)$~~

$p(x_a|x_b) = \mathcal{N}(x_a|x_b \mid \mu_{a|b}, \Sigma_{a|b})$
where. $\mu_{a|b} = \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(x_b - \mu_b)$
$\Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$

2.27. Let x and z be two independent random vectors. so that $p(x, z) = p(x)p(z)$. Show that the mean of their sum $y = x + z$ is given by the sum of means of each of the variable separately. Similarly, show that the covariance matrix of y is given by the sum of the covariance matrics of x and z. Confirm that this result agrees with that of Exercise 1.10

Sol. $y = x + z$. Obviously. $E[y] = E[x] + E[z]$

$Cov(y) = \begin{bmatrix} \sigma_1^2 & \cdots & Cov(y_1, y_n) \\ & \ddots & \\ Cov(y_n, y_1) & & \sigma_n^2 \end{bmatrix}$

$D(y_i) = D(x_i + z_i) = D(x_i) + D(z_i) + 2Cov(x_i, z_i)$
$= D(x_i) + D(z_i)$

$Cov(y_i, y_j) = E(y_i y_j) - E(y_i)E(y_j) = E[(x_i + z_i)(x_j + z_j)] - E(x_i + z_i)E(x_j + z_j)$
$= E(x_i x_j) + E(x_i z_j) + E(x_j z_i) + E(z_i z_j) - E(x_i)E(x_j) - E(z_i)E(x_j) - E(x_i)E(z_j) - E(z_i)E(z_j)$
$= Cov(x_i, x_j) + Cov(z_i, z_j) + Cov(x_i, z_j) + Cov(x_j, z_i)$
$= Cov(x_i, x_j) + Cov(z_i, z_j)$

$\therefore Cov(y) = \begin{bmatrix} \sigma_1^2 & \cdots & Cov(y_1, y_n) \\ & \ddots & \\ Cov(y_n, y_1) & \cdots & \sigma_n^2 \end{bmatrix} = \begin{bmatrix} D(y_1) & \cdots & Cov(y_1, y_n) \\ \vdots & \ddots & \vdots \\ Cov(y_n, y_1) & \cdots & D(y_n) \end{bmatrix} = \begin{bmatrix} D(x_1)+D(z_1) & \cdots & Cov(x_1, x_n) + Cov(z_1, z_n) \\ \vdots & \ddots & \vdots \\ Cov(x_n, x_1) + Cov(z_n, z_1) & \cdots & D(x_n)+D(z_n) \end{bmatrix}$

$= Cov(x) + Cov(z)$

$\therefore Cov(y) = Cov(x) + Cov(z)$

2.28. Consider a joint distribution over the variable $z = \begin{bmatrix} x \\ y \end{bmatrix}$, whose mean and covariance are given by (2.108) and (2.105) respectively. By making use of the results (2.92) and (2.93) show that the marginal distribution $p(x)$ is given (2.99) Similarly, by making use of the results (2.81) and (2.82) show that the conditional distribution $p(y|x)$ is given by (2.100).

Sol. $E(z) = \begin{pmatrix} \mu \\ A\mu + b \end{pmatrix}$ $\quad Cov(z) = R^{-1} = \begin{pmatrix} \Lambda^{-1} & \Lambda^{-1}A^T \\ A\Lambda^{-1} & L^{-1} + A\Lambda^{-1}A^T \end{pmatrix}$

$z = \begin{bmatrix} x \\ y \end{bmatrix}$ $\therefore E(z) = \begin{bmatrix} E(x) \\ E(y) \end{bmatrix}$ $\therefore E(x) = \mu$ $\qquad \therefore p(x) = \mathcal{N}(x \mid \mu, \Lambda^{-1})$

$Cov(z) = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}$ $\therefore \Sigma_{xx} = \Lambda^{-1}$

From $\mu_{a|b} = \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(x_b - \mu_b)$ (2.81)

$\mu_{y|x} = \mu_y + \Sigma_{yx}\Sigma_{xx}^{-1}(x - \mu_x) = A\mu + b + A\Lambda^{-1} \cdot \Lambda(x - \mu) = A\mu + b + A(x - \mu) = Ax + b$

From $\Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$ (2.82)

$\Sigma_{y|x} = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy} = L^{-1} + A\Lambda^{-1}A^T - A\Lambda^{-1}\Lambda\Lambda^{-1}A^T = L^{-1}$

$\therefore p(y|x) = \mathcal{N}(y \mid Ax + b, L^{-1})$

2.31. Consider two ~~multimodel~~ multidimensional random vectors x and z having Gaussian distributions $p(x) = \mathcal{N}(x \mid \mu_x, \Sigma_x)$ and $p(z) = \mathcal{N}(z \mid \mu_z, \Sigma_z)$ respectively, together with their sum $y = x + z$. Use the results (2.109) and (2.110) to find an expression for the marginal distribution $p(y)$ by considering the linear-Gaussian model comprising the product of the marginal distriblution $p(x)$ and the conditional distribution $p(y|x)$.

Sol.: $p(x) = \mathcal{N}(x \mid \mu_x, \Sigma_x)$ $\quad p(z) = \mathcal{N}(z \mid \mu_z, \Sigma_z)$ $\qquad y = z + x$

$\therefore p(y|x) = p(x + z \mid x) = \mathcal{N}(y \mid \mu_z + x, \Sigma_z)$

from 2.109 & 2.110 $\begin{cases} E[y] = A\mu + b \\ Cov[y] = L^{-1} + A\Lambda^{-1}A^T \end{cases}$ and 2.99 & 2.100 $\begin{cases} p(x) = \mathcal{N}(x \mid \mu, \Lambda^{-1}) \\ p(y|x) = \mathcal{N}(y \mid Ax + b, L^{-1}) \end{cases}$

$\therefore \mu \to \mu_x, \ \Sigma_x \to \Lambda^{-1}, \ \mu_z \to b, \ A = I, \ L^{-1} = \Sigma_z$

$\therefore \begin{cases} E[y] = \mu_z + \mu_x \\ Cov[y] = \Sigma_z + \Sigma_x \end{cases}$

2.36. Using an analogous procedure to that used to obtain 2.126, derive an expression for the sequential estimation of the estimation of the variance of a univariate Gaussian distribution, by starting with the maximum likelihood expression

$$\sigma^2_{ML} = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu)^2$$

Verify that substituting the expression for a Gaussian distribution into Robbins-Monro sequential estimation formula (2.135) gives a result of the same form, and hence obtain an expression for the corresponding coefficients a_N.

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

$$
\begin{aligned}
\sigma_{ML}^{2(N)} &= \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)^2 \\
&= \frac{1}{N}(x_N-\mu)^2 + \frac{1}{N}\sum_{n=1}^{N-1}(x_n-\mu)^2 \\
&= \frac{1}{N}(x_n-\mu)^2 + \frac{N-1}{N}\cdot\frac{1}{N-1}\sum_{n=1}^{N-1}(x_n-\mu)^2 \\
&= \frac{1}{N}(x_n-\mu)^2 + \frac{N-1}{N}\cdot\sigma_{ML}^{2(N-1)} \\
&= \sigma_{ML}^{2(N-1)} + \frac{1}{N}\left[(x_n-\mu)^2 - \sigma_{ML}^{2(N-1)}\right]
\end{aligned}
$$

$$p(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$\therefore -\ln p(x|\mu,\sigma) = -\ln\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} = -\ln\frac{1}{\sigma\sqrt{2\pi}} - \ln e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{2}\ln 2\pi\sigma^2 + \frac{(x-\mu)^2}{2\sigma^2}$

$\therefore \frac{\partial}{\partial\sigma^2}(-\ln p(x|\mu,\sigma)) = \frac{1}{2\sigma^2} - \frac{(x-\mu)^2}{2\sigma^4} = -\frac{1}{2\sigma^4}\left[(x-\mu)^2 - \sigma^2\right]$

$\therefore \sigma^{2(N)} = \sigma^{2(N-1)} + a_{N-1}\cdot\frac{1}{2\sigma^{4(N-1)}}\left[(x^N-\mu)^2 - \sigma^{2(N-1)}\right]$

$\therefore a_{N-1} = \frac{2\sigma^{4(N-1)}}{N}$

2.37. Using an analogous to that used to obtain (2.126), derive an expression for the sequential estimation of the covariance of a multivariate Gaussian distribution, by starting with the maximum likelihood expression (2.122). Verify that substituting the expression for a Gaussian distribution into Robbins-Monro sequential estimation formula (2.135) gives a result of the same form, and hence obtain an expression for the ~~Gro~~ corresponding coefficients a_N.

Sol. $\Sigma_{ML}^{N} = \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu_{ML})(x_n-\mu_{ML})^T$

$$
\begin{aligned}
&= \frac{1}{N}\sum_{n=1}^{N-1}(x_n-\mu_{ML})(x_n-\mu_{ML})^T + \frac{1}{N}(x_N-\mu_{ML})(x_N-\mu_{ML})^T \\
&= \frac{N-1}{N}\Sigma_{ML}^{N-1} + \frac{1}{N}(x_N-\mu_{ML})(x_N-\mu_{ML})^T \\
&= \Sigma_{ML}^{N-1} + \frac{1}{N}\left[(x_N-\mu_{ML})(x_N-\mu_{ML})^T - \Sigma_{ML}^{N-1}\right]
\end{aligned}
$$

$$P(x|\mu,\Sigma)=\frac{1}{(2\pi)^{\frac{D}{2}}}\cdot\frac{1}{|\Sigma|^{\frac{1}{2}}}e^{-\frac{1}{2}\{(x-\mu)^T\Sigma^{-1}(x-\mu)\}}$$

$$\frac{\partial \ln P(x_N|\mu,\Sigma)^{N-1}}{\partial\Sigma^{N-1}}=\frac{\partial}{\partial\Sigma^{N-1}}\left(-\frac{D}{2}\ln 2\pi-\frac{1}{2}\ln|\Sigma|-\frac{1}{2}(x_N-\mu)^T\Sigma_{N-1}^{-1}(x_N-\mu)\right)$$

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

$$=-\frac{1}{2}\Sigma_{N-1}^{-1}-\frac{1}{2}\frac{\partial}{\partial\Sigma^{N-1}}(x_N-\mu)^T\Sigma_{N-1}^{-1}(x_N-\mu)$$

From the formula. $\frac{\partial Tr(AX^{-1}B)}{\partial X}=-(X^{-1}BAX^{-1})^T$

$\therefore \frac{\partial}{\partial\Sigma_{N-1}}(x-\mu)^T\Sigma_{N-1}^{-1}(x-\mu)=-[\Sigma_{N-1}^{-1}(x_N-\mu)(x_N-\mu)^T\cdot\Sigma_{N-1}^{-1}]^T$

$\because ((x-\mu)(x-\mu)^T)^T=(x_N-\mu)(x_N-\mu)^T.\quad(\Sigma_{N-1}^{-1})^T=\Sigma_{N-1}^{-1}$

$\therefore \frac{\partial}{\partial\Sigma_{N-1}}\left[(x_N-\mu)^T\Sigma_{N-1}^{-1}(x_N-\mu)\right]=-\Sigma_{N-1}^{-1}(x_N-\mu)(x_N-\mu)^T\Sigma_{N-1}^{-1}$

$\therefore \frac{\partial\ln P(x_N|\mu,\Sigma_{N-1})}{\partial\Sigma_{N-1}}=-\frac{1}{2}\Sigma_{N-1}^{-1}+\frac{1}{2}\Sigma_{N-1}^{-1}(x_N-\mu)(x_N-\mu)^T\Sigma_{N-1}^{-1}$

$\because (x-\mu)\cdot(x-\mu)^T\cdot\Sigma_{N-1}^{-1}=\Sigma_{N-1}^{-1}(x_N-\mu)\cdot(x_N-\mu)^T$. (two matrix are both symmetic matrix)

$\therefore \frac{\partial\ln P(x_N|\mu,\Sigma_{N-1})}{\partial\Sigma_{N-1}}=-\frac{1}{2}\Sigma_{N-1}^{-1}+\frac{1}{2}\Sigma_{N-1}^{-2}(x_N-\mu)(x_N-\mu)^T=\frac{1}{2}\Sigma_{N-1}^{-2}[(x_N-\mu)(x_N-\mu)^T-\Sigma_{N-1}]$

$\therefore \Sigma_N=\Sigma_N^{N-1}+a_{N-1}\cdot\frac{1}{2}\Sigma_{N-1}^{-2}[(x_N-\mu)(x_N-\mu)^T-\Sigma_{N-1}]$

$\therefore a_{N-1}\cdot\frac{1}{2}\Sigma_{N-1}^{-2}=\frac{1}{N}\qquad\therefore a_{N-1}=\frac{2}{N}\Sigma_{N-1}^{2}$

2.39. Starting from the results 2.141 and 2.142 for the posterior distribution of the mean of Gaussian random variable, dissect out the contributions from the first $N-1$ data points and hence obtain expressions for the sequential update of μ_N and σ_N^2. Now derive the same results starting from the posterior distribution $p(\mu|x_1,\dots,x_{N-1})=N(\mu|\mu_{N-1},\sigma_{N-1}^2)$ and multiplying by the likelihood function $P(x_N|\mu)=N(x_N|\mu,\sigma^2)$ and then completing the square and normalizing to obtain the posterior distribution after N observations.

Sol. $P(\mu)=P(\mu|\mu_{N-1},\sigma_{N-1}^2)\times P(x_N|\mu)$

$=N(\mu|\mu_{N-1},\sigma_{N-1}^2)\cdot N(x_N|\mu,\sigma^2)$

$\propto e^{-\frac{1}{2}\left(\frac{(\mu-\mu_{N-1})^2}{\sigma_{N-1}^2}+\frac{(x_N-\mu)^2}{\sigma^2}\right)}$

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

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

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

$$\because\ P(\mu|\mu_N,\sigma_N^2) = N(\mu|\mu_N,\sigma_N^2) \propto e^{-\frac{1}{2}\frac{(\mu-\mu_N)^2}{\sigma_N^2}} = e^{-\frac{1}{2}\frac{\mu^2-2\mu\mu_N+\mu_N^2}{\sigma_N^2}}$$

$$\therefore\ \frac{\mu^2(\sigma^2+\sigma_{N-1}^2)-2\mu(\mu_{N-1}\sigma^2+x_N\sigma_{N-1}^2)+(\mu_{N-1}^2\sigma^2+x_N\sigma_{N-1}^2)}{\sigma^2\sigma_{N-1}^2}$$

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

From 2.142

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

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

From 2.141

$$\mu_N = \frac{\sigma^2}{N\sigma_0^2+\sigma^2}\mu_0 + \frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\mu_{ML} \qquad \frac{\sigma^2}{N\sigma_0^2+\sigma^2} = \frac{1}{\frac{\sigma_0^2}{\sigma^2}N+1} = \frac{1}{\sigma_0^2\left(\frac{N}{\sigma^2}+\frac{1}{\sigma_0^2}\right)} = \frac{1}{\sigma_0^2\cdot\frac{1}{\sigma_N^2}} = \frac{\sigma_N^2}{\sigma_0^2}$$

$$\therefore\ \mu_N = \frac{\sigma_N^2}{\sigma_0^2}\mu_0 + \frac{N\sigma_N^2}{\sigma^2}\mu_{ML}$$

$$\therefore\ \cancel{\mu_{N-1}=}\ \frac{\mu_N}{\sigma_N^2} = \frac{1}{\sigma_0^2}\mu_0 + \frac{N\mu_{ML}}{\sigma^2} = \frac{\mu_0}{\sigma_0^2} + \frac{\sum_{n=1}^{N} x_n}{\sigma^2}$$

$$\therefore\ \frac{\mu_{N-1}}{\sigma_{N-1}^2} = \frac{\mu_0}{\sigma_0^2} + \frac{\sum_{n=1}^{N-1} x_n}{\sigma^2}$$

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

and. $$\mu_N = \frac{\mu_{N-1}\sigma^2+x_N\sigma_{N-1}^2}{\sigma^2+\sigma_{N-1}^2} = \frac{\mu_{N-1}\sigma^2+x_N\sigma_{N-1}^2}{\frac{\sigma^2\sigma_{N-1}^2}{\sigma_N^2}} = \left(\frac{\mu_{N-1}}{\sigma_{N-1}^2}+\frac{x_N}{\sigma^2}\right)\sigma_N^2$$

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

2.40 Consider a D-dimensional Gaussian random variable x with distribution $\mathcal{N}(x|\mu, \Sigma)$ in which the covariance Σ is known and for which we wish to infer the mean μ from a set of observations $X = \{x_1, \dots, x_N\}$. Given a prior distribution $p(\mu) = \mathcal{N}(\mu|\mu_0, \Sigma_0)$, find the corresponding posterior distribution $p(\mu|X)$.

Sol. $p(\mu|X) \propto p(\mu) \prod_{n=1}^{N} p(x_n|\mu, \Sigma)$

$$p(\mu) = \mathcal{N}(\mu|\mu_0, \Sigma_0) \propto e^{-\frac{1}{2}(\mu-\mu_0)^T \Sigma_0^{-1} (\mu-\mu_0)}$$

$$\sum_{n=1}^{N} p(x_n|\mu, \Sigma) \propto e^{-\frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^T \Sigma^{-1} (x-\mu)}$$

$$\therefore \; p(\mu) \cdot \prod_{n=1}^{N} p(x_n|\mu, \Sigma) \propto e^{-\frac{1}{2}(\mu-\mu_0)^T \Sigma_0^{-1}(\mu-\mu_0) - \frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^T \Sigma^{-1}(x-\mu)}$$

$$-\frac{1}{2}(\mu-\mu_0)^T \Sigma_0^{-1}(\mu-\mu_0) - \frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^T \Sigma^{-1}(x_n-\mu)$$

$$= -\frac{1}{2}\mu^T\Sigma_0^{-1}\mu + \frac{1}{2}\mu_0^T\Sigma_0^{-1}\mu + \frac{1}{2}\mu\Sigma_0^{-1}\mu_0 - \frac{1}{2}\mu_0^T\Sigma_0^{-1}\mu_0 - \frac{1}{2}\sum_{n=1}^{N}(x_n\Sigma^{-1}x_n - \mu^T\Sigma^{-1}x_n - x_n^T\Sigma^{-1}\mu + \mu^T\Sigma^{-1}\mu)$$

$$= -\frac{1}{2}\mu^T\Sigma_0^{-1}\mu + \mu^T\Sigma_0^{-1}\mu_0 - \frac{1}{2}\mu_0^T\Sigma_0^{-1}\mu_0 - \frac{1}{2}\sum_{n=1}^{N}x_n^T\Sigma^{-1}x_n + \sum_{n=1}^{N}\mu^T\Sigma^{-1}x_n - \frac{1}{2}\sum_{n=1}^{N}\mu^T\Sigma^{-1}\mu$$

$$= -\frac{1}{2}\mu^T\Sigma_0^{-1}\mu + \mu^T\Sigma_0^{-1}\mu_0 + \sum_{n=1}^{N}\mu^T\Sigma^{-1}x_n - \frac{N}{2}\mu^T\Sigma^{-1}\mu + \text{Const}$$

$$= -\frac{1}{2}\mu^T(\Sigma_0^{-1} + N\Sigma^{-1})\mu + \mu^T(\Sigma_0^{-1}\mu_0 + \Sigma^{-1}\sum_{n=1}^{N}x_n) + \text{Const}$$

from 2.71. a normal distribution can be ~~separe~~ splited

$$-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) = -\frac{1}{2}x^T\Sigma^{-1}x + x^T\Sigma^{-1}\mu + \text{Const}$$

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

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

and the maximum likelihood solution of μ is $\mu_{ML} = \frac{1}{N}\sum_{n=1}^{N}x_n$

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