4.11

Intro

In this exercise, we will derive the NIW posterior for the joint distribution of μ and Σ. **Don't worry**, despite the involvement of a exotic distribution, the resolution is quite straightfoward. The exercise will follow our old established recipe: state the prior and the likelihood, combine them and update the parameters, determine the posterior less the normalization constant.

Solution

Firts let's state the prior and the likelihood. From section 4.6.3, we know that the likelihood is:

$$p(D|\mu,\Sigma) \propto |\Sigma|^{-\frac{N}{2}} exp\left(-\frac{N}{2}(\mu-\bar{x})^T\Sigma^{-1}(\mu-\bar{x})\right) exp\left(-\frac{1}{2}tr(\Sigma^{-1}S_{\bar{x}})\right) \quad (1)$$

Also, from the same section, we know that the conjugate prior is:

$$\begin{aligned} &p(\mu,\Sigma) = NIW(\mu,\Sigma|m_0,\kappa_0,\nu_0,S_0) \propto \\ &|\Sigma|^{-\frac{\nu_0+D+2}{2}} exp\left(-\frac{\kappa_0}{2}(\mu-m_0)^T\Sigma^{-1}(\mu-m_0)-\frac{1}{2}tr(\Sigma^{-1}S_0)\right) \end{aligned} \quad (2)$$

Therefo, the posterior is proportional to:

$$\begin{aligned} &p(\mu,\Sigma|D) \propto p(D|\mu,\Sigma)p(\mu,\Sigma) \propto \\ &|\Sigma|^{-\frac{N}{2}} exp\left(-\frac{N}{2}(\mu-\bar{x})^T\Sigma^{-1}(\mu-\bar{x})\right) exp\left(-\frac{1}{2}tr(\Sigma^{-1}S_{\bar{x}})\right) \\ &|\Sigma|^{-\frac{\nu_0+D+2}{2}} exp\left(-\frac{\kappa_0}{2}(\mu-m_0)^T\Sigma^{-1}(\mu-m_0)-\frac{1}{2}tr(\Sigma^{-1}S_0)\right) = \\ &|\Sigma|^{-\frac{\nu_0+D+2+N}{2}} exp\left(-\frac{N}{2}(\mu-\bar{x})^T\Sigma^{-1}(\mu-\bar{x})-\frac{\kappa_0}{2}(\mu-m_0)^T\Sigma^{-1}(\mu-m_0)\right) \\ &exp\left(-\frac{1}{2}tr(\Sigma^{-1}S_{\bar{x}})-\frac{1}{2}tr(\Sigma^{-1}S_0)\right) \end{aligned} \quad (3)$$

Now, we need to work our way around this expression to find the updated NIW. As a said before, don't worry, it is very straightfoward. We only need to divide and conquer the problem. **First**, Let's start with the determinant of the covariance matrix. In the posterior, we have $|\Sigma|^{-\frac{\nu_0+D+2+N}{2}}$. Since D is fixed during the inference, the only parameter left to update is ν_0. Thus, $\nu_N = \nu_0 + N$.

Second, let's work with the quadratic form expressions. We will be woried only with the inner terms of the expression

$exp\left[-\frac{1}{2}\left(N(\mu-\bar{x})^T\Sigma^{-1}(\mu-\bar{x})+\kappa_0(\mu-m_0)^T\Sigma^{-1}(\mu-m_0)\right)\right]$.

From section 4.6.3.1, we know that:

$$\sum_{i=0}^{N}(x_i-\mu)\Sigma^{-1}(x_i-\mu) = tr(\Sigma^{-1}S_{\bar{x}}) + N(\bar{x}-\mu)^T\Sigma^{-1}(\bar{x}-\mu) \quad (4)$$

Note that our quadratic form sum fits the equation above, because we have a summation of N equal terms of one kind and κ_0 terms of the other kind. Thus:

$$\begin{aligned}
&N(\mu-\bar{x})^T\Sigma^{-1}(\mu-\bar{x})+\kappa_0(\mu-m_0)^T\Sigma^{-1}(\mu-m_0)= \\
&tr(\Sigma^{-1}S_{m_N})+(N+\kappa_0)(m_N-\mu)^T\Sigma^{-1}(m_N-\mu) \\
&m_N=\frac{sumAllTerms}{numberOfTerms}=\frac{\kappa_0 m_0+N\bar{x}}{\kappa_0+N} \\
&S_{m_N}=N(\bar{x}-m_N)(\bar{x}-m_N)^T+\kappa_0(m_0-m_N)(m_0-m_N)^T= \\
&\frac{N\kappa_0}{N+\kappa_0}(\bar{x}-m_0)(\bar{x}-m_0)^T
\end{aligned} \tag{5}$$

In the scatter matrix calculation, in order to arrive at the final result you just need to expand the previous expression and cancel out some terms.

Third, we need to combine the two initial traces with the new one produced in the previous step. In order to do that, we only need to use the property that the some of traces is equal to the trace of sums. Thus

$tr(\Sigma^{-1}S_0)+tr(\Sigma^{-1}S_{barx})+tr(\Sigma^{-1}S_{m_N})=tr(\Sigma^{-1}(S_0+S_{\bar{x}}+S_{m_N}))$.

As the last step, we need to combine all the new terms and update the remaining parameters:

$$\begin{aligned}
&p(\mu,\Sigma)\propto|\Sigma|^{-\frac{\nu_N+D+2}{2}}exp\left(-\frac{N+\kappa_0}{2}(m_N-\mu)^T\Sigma^{-1}(m_N-\mu)\right) \\
&exp(-\frac{1}{2}tr(\Sigma^{-1}(S_0+S_{\bar{x}}+S_{m_N})))= \\
&|\Sigma|^{-\frac{\nu_N+D+2}{2}}exp\left(-\frac{\kappa_N}{2}(\mu-m_N)^T\Sigma^{-1}(\mu-m_N)-\frac{1}{2}tr(\Sigma^{-1}S_N)\right)
\end{aligned} \tag{6}$$

So, our posterior is given by the following distribution:

$$\begin{aligned}
&p(\mu,\Sigma)=NIW(\mu,\Sigma|m_N,\kappa_N,\nu_N,S_N) \\
&m_N=\frac{\kappa_0 m_0+N\bar{x}}{\kappa_0+N} \\
&\kappa_N=\kappa_0+N \\
&\nu_N=\nu_0+N \\
&S_N=S_0+S_{\bar{x}}+S_{m_N}
\end{aligned} \tag{7}$$

Conclusion

In this exercise we calculated the posterior for the joint distribution of the parameters μ and Σ of the MVN. We saw that despite the NIW being a verbose distribution, it was straightfoward to calculate its posterior, given the conjugate prior and the likelihood. We used our old recipe of expressing the prior and the likelihood and combine them to achive the posterior. To update the parameters, we used a divide and conquer strategy, updating different parts of the expression and them combining them together. The final result is in accordance with the book result.