Bayesian Inference and Data Assimilation- Exercise Sheet 3

Group : Margherita Lea Corona, Yuvraj Dhepe, Nicola Horst, Aleksandra Sharapova

Exercise 1:

We assume that all the following random variables take values in $\mathbb{R}$.
Let X be a random variable such that $E[X] = m$ and $\text{Var}[X] = Q \implies E[X^2] = Q + m^2$, given that $\text{Var}[X] = E[X^2] - E[X]^2$. Then let W be a random variable with mean 0 and variance R, therefore its second moment is: R. Furthermore we assume $E[XW] = 0$ (given that $E[W] = 0$ this implies $\text{Cov}[X, W] = 0$, i.e the two variables are uncorrelated).
Let

$$Y = HX + W, \; H \in \mathbb{R} \setminus \{0\}.$$

We want to build an affine estimator of X

$$\hat{X} = K^*Y + b^*,$$

where $(K^*, b^*) = \arg\min_{K,b} E[(\hat{X} - X)^2]$.

1.

We first compute $E[(\hat{X} - X)^2]$ and then we derive this function of K, b w.r.t b (i.e we consider K fixed).

$$\begin{aligned}
E[(\hat{X} - X)^2] &= E[(KHX + KW + b - X)^2] = \\
&= E[(X(KH - 1) + (KW + b))^2]) = \\
&= E[(X^2(KH - 1)^2 + (KW + b)^2 + 2X(KH - 1)(KW + b)] = \\
&= E[X^2](KH - 1)^2 + K^2E[W^2] + b^2 + 2KbE[W] + \\
&+ 2K^2HE[XW] + 2E[X]KHb - 2KE[XW] - 2bE[X] = \\
&= E[X^2](KH - 1)^2 + K^2R + b^2 + 2mKHb - 2bm = \\
&= (Q + m^2)(KH - 1)^2 + K^2R + b^2 + 2mKHb - 2bm = \\
&= (Q + m^2)(KH - 1)^2 + K^2R + b^2 + 2mb(KH - 1),
\end{aligned}$$

where we have used the hypothesis above and the linearity of the expectation considering that the only random quantities are X and W.
$\implies \frac{\partial}{\partial b}E[(\hat{X} - X)^2] = 2b + 2mKH - 2m = 0 \iff b = m(1 - KH)$.
If we consider K fixed and $E[(\hat{X} - X)^2]$ as a function of b then $b^* = m(1 - KH)$ is a point of minimum (optimal b), since the second derivative is $2 > 0$ (i.e the function is convex).

2.

$$E[(\hat{X} - X)] = E[KY + b^* - X] = KHm + m(1 - KH) - m = KHm - mKH + m - m = 0,$$

where we have used that $E[X] = m$ and $E[KY + b^*] = KE[Y] + b^* = KE[HX + W] + b^* = KHm + 0 + m(1 - KH)$. This way we see that the estimator is unbiased ($E[\hat{X}] = E[X]$) regardless of K.

3.

We have already explicited $E[(\hat{X}-X)^2]$ in point **1.**, then we compute the derivative w.r.t K and we plug in the optimal b, b^*:

$$\begin{aligned}\left(\frac{\partial}{\partial K}E[(\hat{X}-X)^2]\right)\Big|_{b=b^*} &= (Q+m^2)2(KH-1)H+2KR+2mHb^* = \\ &= (Q+m^2)2(KH-1)H+2KR+2m^2H(1-KH) = \\ &= 2(Q+m^2)KH^2-2(Q+m^2)H+2KR+2m^2H-2m^2KH^2 = \\ &= 2QKH^2+2m^2KH^2-2QH-2m^2H+2m^2H+2KR-2m^2KH^2 = \\ &= 2K(QH^2+R)-2QH,\end{aligned}$$

therefore:

$$\left(\frac{\partial}{\partial K}E[(\hat{X}-X)^2]\right)\Big|_{b=b^*}=0 \iff K=\frac{HQ}{QH^2+R}.$$

Consequently we have $(K^*,b^*)=\left(\frac{HQ}{QH^2+R},m(1-K^*H)\right)=\left(\frac{HQ}{QH^2+R},m(\frac{R}{QH^2+R})\right)$.

We know that this is a point of minimum of the mean squared error considered as a function of K,b since the function is convex: the hessian matrix is positive definite (we see this by applying Sylvester criterium).

The optimal estimator is:

$$\hat{X}=\frac{HQ}{H^2Q+R}Y+m\frac{R}{H^2Q+R}.$$

Now we compute the mean squared error (minimal $E[(\hat{X}-X)]$ w.r.t K,b):

$$\begin{aligned}E[(\hat{X}-X)^2] &= (Q+m^2)(K^*H-1)^2+{K^*}^2R+{b^*}^2+2mb^*(K^*H-1) = \\ &= (Q+m^2)\left(\frac{-R}{QH^2+R}\right)^2+\left(\frac{HQ}{QH^2+R}\right)^2R+m^2\frac{R^2}{(QH^2+R)^2}-2m^2\left(\frac{R}{QH^2+R}\right)^2 = \\ &= \frac{RQ(R+H^2Q)}{(QH^2+R)^2}\end{aligned}$$

Exercise 2:

We consider a Gaussian random variable $(X_1,X_2)\in\mathbb{R}^2$ with distribution $n(z;\bar{z},\Sigma)$, where $z=\begin{pmatrix}x_1\\x_2\end{pmatrix}$. The mean is $\bar{z}=(\bar{x_1},\bar{x_2})^T$, while the covariance matrix is:

$$\Sigma=\begin{pmatrix}\sigma_{11}^2 & \sigma_{12}^2\\ \sigma_{21}^2 & \sigma_{22}^2\end{pmatrix}=\begin{pmatrix}\sigma_{11}^2 & \sigma_{12}^2\\ \sigma_{12}^2 & \sigma_{22}^2\end{pmatrix},$$

given that $\sigma_{12}^2=\mathrm{Cov}(X_1,X_2)=\mathrm{Cov}(X_2,X_1)=\sigma_{21}^2$ (the covariance is a symmetric operator).

The Gaussian PDF is:

$$n(z;\bar{z},\Sigma)=\frac{1}{2\pi\sqrt{|\Sigma|}}\exp(-\frac{1}{2}(z-\bar{z})^T\Sigma^{-1}(z-\bar{z})),$$

where $|\Sigma|:=det(\Sigma)$.

1.

First we find σ_c by observing that to get it it's sufficient to solve : $\sqrt{2\pi}\sigma_c\sqrt{2\pi}\sigma_{22} = 2\pi|\Sigma|^{-1/2}$. Given that $|\Sigma| = \sigma_{11}^2\sigma_{22}^2 - \sigma_{12}^2\sigma_{12}^2$ we have:

$$\sqrt{2\pi}\sigma_c\sqrt{2\pi}\sigma_{22} = 2\pi|\Sigma|^{1/2} \iff \sigma_c = \frac{|\Sigma|^{1/2}}{\sigma_{22}} \iff \sigma_c^2 = \sigma_{11}^2 - \frac{\sigma_{12}^4}{\sigma_{22}^2}.$$

Now we need to find $\bar{x}_c$, we start by rewriting the inverse of the covariance matrix in terms of σ_c^2:

$$\Sigma^{-1} = \frac{1}{|\Sigma|}\begin{pmatrix} \sigma_{22}^2 & -\sigma_{12}^2 \\ -\sigma_{12}^2 & \sigma_{11}^2 \end{pmatrix} = \frac{1}{\sigma_c^2\sigma_{22}^2}\begin{pmatrix} \sigma_{22}^2 & -\sigma_{12}^2 \\ -\sigma_{12}^2 & \sigma_{11}^2 \end{pmatrix} = \begin{pmatrix} \sigma_c^{-2} & -\frac{\sigma_{12}^2}{\sigma_c^2\sigma_{22}^2} \\ -\frac{\sigma_{12}^2}{\sigma_c^2\sigma_{22}^2} & \frac{\sigma_{11}^2}{\sigma_c^2\sigma_{22}^2} \end{pmatrix}.$$

To explicit $\bar{x}_c$ we focus only on the exponent of the exponential function, i.e:

$$-\frac{1}{2}((z-\bar{z})^T\Sigma^{-1}(z-\bar{z}))$$

and we solve te product using the Σ matrix with σ_c^2:

$$-\frac{1}{2}\left((z-\bar{z})^T\Sigma^{-1}(z-\bar{z})\right) = \tag{1}$$

$$= -\frac{1}{2}(x_1-\bar{x_1}, x_2-\bar{x_2})\begin{pmatrix} \sigma_c^{-2}(x_1-\bar{x_1}) - (\sigma_{12}^2\sigma_{22}^{-2}\sigma_c^{-2})(x_2-\bar{x_2}) \\ -(\sigma_{12}^2\sigma_{22}^{-2}\sigma_c^{-2})(x_1-\bar{x_1}) + \sigma_{11}^2\sigma_c^{-2}\sigma_{22}^{-2}(x_2-\bar{x_2}) \end{pmatrix} = \tag{2}$$

$$= -\frac{1}{2}\left[\sigma_c^{-2}(x_1-\bar{x_1})^2 - 2(\sigma_{12}^2\sigma_{22}^{-2}\sigma_c^{-2})(x_1-\bar{x_1})(x_2-\bar{x_2}) + \sigma_{11}^2\sigma_c^{-2}\sigma_{22}^{-2}(x_2-\bar{x_2})^2\right], \tag{3}$$

by collecting σ_c^{-2} and other algebraic manipulations we get:

$$-\frac{1}{2}\left[\sigma_c^{-2}(x_1-\bar{x_1})^2 - 2(\sigma_{12}^2\sigma_{22}^{-2}\sigma_c^{-2})(x_1-\bar{x_1})(x_2-\bar{x_2}) + \sigma_{11}^2\sigma_c^{-2}\sigma_{22}^{-2}(x_2-\bar{x_2})^2\right] = \tag{4}$$

$$= -\frac{1}{2\sigma_c^2}\left[(x_1-\bar{x_1})^2 - 2(\sigma_{12}^2\sigma_{22}^{-2})(x_1-\bar{x_1})(x_2-\bar{x_2}) + \frac{\sigma_{11}^2}{\sigma_{22}^2}(x_2-\bar{x_2})^2\right] = \tag{5}$$

$$= -\frac{1}{2\sigma_c^2}\left[x_1^2 - 2(x_1\bar{x_1} + x_1\sigma_{12}^2\sigma_{22}^{-2}(x_2-\bar{x_2})) + \bar{x_1}^2 + 2\bar{x_1}(\sigma_{12}^2\sigma_{22}^{-2})(x_2-\bar{x_2}) + \frac{\sigma_{11}^2}{\sigma_{22}^2}(x_2-\bar{x_2})^2\right], \tag{6}$$

now we notice that if we set $\bar{x_c} = \bar{x_1} + \sigma_{12}^2\sigma_{22}^{-2}(x_2-\bar{x_2})$ and we add to (6) $0 = \sigma_{12}^4\sigma_{22}^{-4}(x_2-\bar{x_2})^2 - \sigma_{12}^4\sigma_{22}^{-4}(x_2-\bar{x_2})^2$ we complete the square $(x_1-\bar{x_c})^2$:

$$-\frac{1}{2\sigma_c^2}\left[x_1^2 - 2(x_1\bar{x_1} + x_1\sigma_{12}^2\sigma_{22}^{-4}(x_2-\bar{x_2})) + \bar{x_1}^2 + 2\bar{x_1}(\sigma_{12}^2\sigma_{22}^{-2})(x_2-\bar{x_2}) + \frac{\sigma_{11}^2}{\sigma_{22}^2}(x_2-\bar{x_2})^2\right] = \tag{7}$$

$$= -\frac{1}{2\sigma_c^2}[x_1^2 - 2x_1(\bar{x_1} + \sigma_{12}^2\sigma_{22}^{-2}(x_2-\bar{x_2})) + (\bar{x_1}^2 + 2\bar{x_1}(\sigma_{12}^2\sigma_{22}^{-2})(x_2-\bar{x_2}) + \sigma_{12}^4\sigma_{22}^{-4}(x_2-\bar{x_2})^2) + \tag{8}$$

$$+\frac{\sigma_{11}^2}{\sigma_{22}^2}(x_2-\bar{x_2})^2 - \sigma_{12}^4\sigma_{22}^{-4}(x_2-\bar{x_2})^2] = -\frac{1}{2\sigma_c^2}\left(x_1^2 - 2x_1\bar{x}_c + \bar{x}_c^2 + \frac{\sigma_{11}^2}{\sigma_{22}^2}(x_2-\bar{x_2})^2 - \sigma_{12}^4\sigma_{22}^{-4}(x_2-\bar{x_2})^2\right) = \tag{9}$$

$$= -\frac{1}{2\sigma_c^2}\left[(x_1-\bar{x_c})^2+\frac{\sigma_{11}^2}{\sigma_{22}^2}(x_2-\bar{x_2})^2-\sigma_{12}^4\sigma_{22}^{-4}(x_2-\bar{x_2})^2\right]. \tag{10}$$

Since $\frac{\sigma_{11}^2}{\sigma_{22}^2}-\frac{\sigma_{12}^4}{\sigma_{22}^4}=\frac{\sigma_{11}^2\sigma_{22}^2-\sigma_{12}^2\sigma_{12}^2}{\sigma_{22}^4}=\frac{|\Sigma|}{\sigma_{22}^4}=\frac{\sigma_c^2\sigma_{22}^2}{\sigma_{22}^4}=\frac{\sigma_c^2}{\sigma_{22}^2}$, we get:

$$-\frac{1}{2\sigma_c^2}\left[(x_1-\bar{x_c})^2+\frac{\sigma_{11}^2}{\sigma_{22}^2}(x_2-\bar{x_2})^2-\sigma_{12}^4\sigma_{22}^{-4}(x_2-\bar{x_2})^2\right]=-\frac{1}{2\sigma_c^2}(x_1-\bar{x_c})^2-\frac{1}{2\sigma_{22}^2}(x_2-\bar{x_2})^2. \tag{11}$$

Finally, we merge all the above results togheter and, given that $\exp(a+b)=\exp(a)\exp(b)\forall\ a,b\in\mathbb{R}$ we obtain the final result:

$$n(z;\bar{z},\Sigma)=\frac{1}{2\pi\sqrt{|\Sigma|}}\exp\left(-\frac{1}{2}(z-\bar{z})^T\Sigma^{-1}(z-\bar{z})\right) \tag{12}$$

$$=\frac{1}{2\pi\sigma_c\sigma_{22}}\exp\left(-\frac{1}{2\sigma_c^2}(x_1-\bar{x_c})^2-\frac{1}{2\sigma_{22}^2}(x_2-\bar{x_2})^2\right)= \tag{13}$$

$$=\frac{1}{\sqrt{2\pi}\sigma_c}\exp\left(-\frac{1}{2\sigma_c^2}(x_1-\bar{x_c})^2\right)\frac{1}{\sqrt{2\pi}\sigma_{22}}\exp\left(-\frac{1}{2\sigma_{22}^2}(x_2-\bar{x_2})^2\right), \tag{14}$$

where: $\begin{cases}\sigma_c=\frac{|\Sigma|^{1/2}}{\sigma_{22}}\\ \bar{x_c}=\bar{x_1}+\sigma_{12}^2\sigma_{22}^{-2}(x_2-\bar{x_2})\end{cases}$

2.

With the formula obtained in **1.** we manage to separate the distribution of X_2 from the rest, therefore, since $\pi_{X_1|X_2}(x_1|x_2)=\frac{\pi_{(X_1,X_2)}(x_1,x_2)}{\pi_{X_2}(x_2)}$, we have:

$$\pi_{X_1|X_2}(x_1|x_2)=\frac{\pi_{(X_1,X_2)}(x_1,x_2)}{\pi_{X_2}(x_2)}=\frac{n(z;\bar{z},\Sigma)}{\pi_{X_2}(x_2)}=\frac{1}{\sqrt{2\pi}\sigma_c}\exp\left(-\frac{1}{2\sigma_c^2}(x_1-\bar{x_c})^2\right).$$

This is the density of a gaussian random variable of mean $\bar{x}_c$ and variance σ_c^2, which means $X_1|X_2=x_2\sim N(\bar{x_c},\sigma_c^2)$.

Exercise 3:

We observe that if X and W are independent then $\mathrm{Cov}[X,W]=0$, which implies, since $E[W]=0$, that $E[XW]=0$ (hypothesis in Ex 1).

1.

The mean of Z is $\bar{z}=\begin{pmatrix}E[X]\\E[Y]\end{pmatrix}=\begin{pmatrix}m\\mH\end{pmatrix}$, while the covariance matrix is:

$$\Sigma=\begin{pmatrix}\mathrm{Var}[X] & \mathrm{Cov}[X,Y]\\ \mathrm{Cov}[Y,X] & \mathrm{Var}[Y]\end{pmatrix}=\begin{pmatrix}Q & HQ\\ HQ & H^2Q+R\end{pmatrix},$$

given that $\mathrm{Var}[X]=Q$, $\mathrm{Var}[Y]=\mathrm{Var}[HX+W]=H^2\mathrm{Var}[Y]+\mathrm{Var}[W]=H^2Q+R$ (due to independece between X and W) and $\mathrm{Cov}[Y,X]=\mathrm{Cov}[X,Y]=E[XY]-E[X]E[Y]=E[HX^2+XW]-m^2H=H(Q+m^2)-m^2H=HQ$ (the covariance is a symmetric operator).

2.

$$\pi_{X|Y}(x|y)=\frac{1}{\sqrt{2\pi}\sigma_c}\exp\left(-\frac{1}{2\sigma_c^2}(x-\bar{x_c})^2\right),$$

where in this case:

$$\sigma_c^2 = Q - \frac{H^2Q^2}{H^2Q+R} = \frac{QR}{H^2Q+R},$$

$$\bar{x}_c = E[X] + \frac{HQ}{H^2Q+R}(y-Hm) = m + \frac{HQ}{H^2Q+R}y - \frac{H^2Q}{H^2Q+R}m = \frac{HQ}{H^2Q+R}y + m\frac{R}{H^2Q+R}$$

3. We see that the mean of $X|Y=y$ is $\hat{X}(y)$ and this makes sense since $\hat{X}$ is the optimal estimator w.r.t mean squared error minimization and we know that w.r.t this metric the best approximation of a variable by a constant (if $\{Y=y\}$ has happened, then $\hat{X}(y)$ is a constant) is its expected value.

Exercise 4:

First we find $\pi_{X_2}(x_2)$:

$$\begin{aligned}
\pi_{X_2}(x_2) &= \int_{\mathbb{R}} \pi_{(X_1,X_2)}(x_1,x_2)\,dx_1 = \\
&= \frac{1}{Z}e^{-x_2^2}\int_{\mathbb{R}} e^{-x_1^2(1+x_2^2)}\,dx_1 = \\
&= \frac{1}{Z}e^{-x_2^2}\frac{\sqrt{\pi}}{\sqrt{(1+x_2^2)}},
\end{aligned}$$

where the integral is solved using polar coordinates.

Now we can compute the conditional expectation:

$$\begin{aligned}
E[X_1^2X_2|X_2=a] &= aE[X_1^2|X_2=a] = \\
&= a\int_{\mathbb{R}} x_1^2\pi_{X_1|X_2=a}(x_1|a)\,dx_1 = \\
&= a\int_{\mathbb{R}} x_1^2\frac{\pi_{(X_1,X_2)}(x_1,a)}{\pi_{X_2}(a)}\,dx_1 = \\
&= a\int_{\mathbb{R}} x_1^2\frac{1}{Z}e^{-x_1^2-a^2-x_1^2a^2}\frac{Z\sqrt{1+a^2}}{\sqrt{\pi}}e^{a^2}\,dx_1 = \\
&= a\int_{\mathbb{R}} x_1^2e^{-x_1^2(1+a^2)}\frac{\sqrt{1+a^2}}{\sqrt{\pi}}\,dx_1 = \\
&= \frac{a}{2(1+a^2)}
\end{aligned}$$

where the first equality is due to the fact that given X, Y random variables and given $g(Y)$ a function of Y we have: $E[g(Y)X|Y=y] = g(y)E[X|Y=y]$. While the integral is computed first integrating by parts and then using again polar coordinates.