Bayesian inference and Data assimilation

Exercise 4

Jin W. Kim (jin.won.kim@uni-potsdam.de)

Problem 1a Since $X \sim f$, one can note that for any odd function $g(x)$,

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} \underbrace{g(x)f(x)}_{\text{odd function}} \,\mathrm{d}x = 0$$

In particular, any odd power of x is an odd function, and therefore $\mathbb{E}[X] = \mathbb{E}[X^3] = \ldots = 0$. Moreover, the second moment is equal to the variance because

$$\mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] = \sigma$$

Here we may assume that $\sigma > 0$, because otherwise X has to be zero almost surely, and it does not possess the pdf.[1]

In order to make X and $Z = aX^2 + bX + c$ uncorrelated, we want the covariance between X and Z be zero. Since

$$\begin{aligned}\mathrm{Cov}[X, Z] &= \mathbb{E}[XZ] - \mathbb{E}[X]\mathbb{E}[Z] \\ &= \mathbb{E}[X(aX^2 + bX + c)] - 0 \\ &= a\mathbb{E}[X^3] + b\mathbb{E}[X^2] + c \\ &= b\sigma\end{aligned}$$

we conclude that X and Z are uncorrelated if $b = 0$.

In order to make X and Z independent, we want the conditional distribution of Z given X be identical to the marginal distribution of Z. However, for any value of x, the conditional probability distribution becomes:

$$\mathbb{P}[Z = z \mid X = x] = \begin{cases} 1 & \text{if } z = ax^2 + bx + c \\ 0 & \text{otherwise} \end{cases}$$

Therefore the only case is such that the marginal distribution of Z is constant with probability 1, that is, $a = b = 0$ and $Z = c$ with probability 1.

[1] Sometimes one can consider the pdf of X is Dirac-delta ‘function’, but let us avoid unnecessary technicality.

Problem 1b

- Claim 1 is FALSE. A counterexample is from Problem 1a, where $b = 0$ but $a \neq 0$.
- Claim 2 is TRUE. Proof: Let $f_{XZ}(x, z)$ be the joint pdf, and $f_X(x)$ and $f_Z(z)$ are the marginal pdf's of X and Z, respectively. If X and Z are independent, $f_{XZ}(x, z) = f_X(x)f_Z(z)$. Therefore,

$$\begin{aligned}\mathbb{E}[XZ] &= \iint xz f_{XZ}(x, z)\,\mathrm{d}x\,\mathrm{d}z \\ &= \iint xz f_X(x) f_Z(z)\,\mathrm{d}x\,\mathrm{d}z \\ &= \Big(\int x f_X(x)\,\mathrm{d}x\Big)\Big(\int z f_Z(z)\,\mathrm{d}z\Big) \\ &= \mathbb{E}[X]\mathbb{E}[Z]\end{aligned}$$

  Therefore, $\mathrm{Cov}[X, Z] = \mathbb{E}[XZ] - \mathbb{E}[X]\mathbb{E}[Z] = 0$.

- claim 3 is also FALSE. Observe that

$$\begin{aligned}\mathrm{Var}[X + Z] &= \mathbb{E}\big[(X + Z - \mathbb{E}[X + Z])^2\big] \\ &= \mathbb{E}\big[((X - \mathbb{E}[X]) + (Z - \mathbb{E}[Z]))^2\big] \\ &= \mathbb{E}\big[(X - \mathbb{E}[X])^2 + 2(X - \mathbb{E}[X])(Z - \mathbb{E}[Z]) + (Z - \mathbb{E}[Z])^2\big] \\ &= \mathrm{Var}[X] + \mathrm{Var}[Z] + 2\,\mathrm{Cov}[X, Z]\end{aligned}$$

  Therefore, the claim is equivalent to $\mathrm{Cov}[X, Z] = 0$. However, we just have shown that X, Z may be uncorrelated but not independent in claim 1.

Problem 2 1. The joint distribution of X and Y are characterized by a 2×2 matrix of the form:

$$T := \begin{pmatrix} t_{00} & t_{01} \\ t_{10} & t_{11} \end{pmatrix}$$

where $t_{ij} = \mathbb{P}[X = i, Y = j]$. By the law of total probability,

$$\mathbb{P}[X = i] = \sum_{j=1}^{2} \mathbb{P}[X = i, Y = j] = \sum_{j=1}^{2} t_{ij}, \quad \mathbb{P}[Y = j] = \sum_{i=1}^{2} \mathbb{P}[X = i, Y = j] = \sum_{i=1}^{2} t_{ij}$$

Therefore we obtain the following:

$$t_{00} + t_{01} = \frac{1}{2}, \quad t_{10} + t_{11} = \frac{1}{2}, \quad t_{00} + t_{10} = \frac{1}{3}, \quad t_{01} + t_{11} = \frac{2}{3}$$

Parameterize $t_{00} = p$. Then it is straight forward that

$$t_{01} = \frac{1}{2} - p, \quad t_{10} = \frac{1}{3} - p, \quad t_{11} = \frac{1}{6} + p$$

Also, we want $0 \leq t_{ij} \leq 1$ for all pairs of (i, j), and therefore one obtains

$$p \in \big[0, \frac{1}{3}\big]$$

2. Since the variance of X and Y are given constant regardless of p, the correlation is proportional to the covariance $\operatorname{Cov}[X,Y]$. Observe that

$$\begin{aligned}\operatorname{Cov}[X,Y] &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\ &= \sum_{i,j} ij \cdot t_{ij} - \frac{1}{2}\cdot\frac{2}{3} \\ &= \frac{1}{6} + p - \frac{1}{3} \\ &= p - \frac{1}{6}\end{aligned}$$

It is minimized at $p = 0$ with $\operatorname{Cov}[X,Y] = -\frac{1}{6}$, and maximized at $p = \frac{1}{3}$ with $\operatorname{Cov}[X,Y] = \frac{1}{6}$. One can note that if there are more masses on either $(X,Y) = (0,0)$ and $(1,1)$ than $(0,1)$ or $(1,0)$, then X and Y has higher value of correlation, while they have negative correlation for the opposite case.

3. The two random variables become uncorrelated if $\operatorname{Cov}[X,Y] = 0$, which is at $p = \frac{1}{6}$.

4. Since $Z = 0$ with probability 1, the joint distribution does not have any meaning than the marginal on X. The correlation is always 0, and indeed X and Z are independent.

Problem 3

- Since X_1 and X_2 are Gaussian, the joint distribution is also Gaussian with

  $$\mathbb{E}\big[(X_1,X_2)\big] = (\bar{x}_1,\bar{x}_2), \qquad \operatorname{Var}\big[(X_1,X_2)\big] = \begin{pmatrix}\sigma_1 & c \\ c & \sigma_2\end{pmatrix}$$

  where c denotes the covariance:

  $$c = \operatorname{Cov}[X_1,X_2] = \mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2]$$

  Therefore, given the marginals, the covariance c determines the joint distribution. Now consider the square of the Wasserstein (2) distance for the sake of simplicity.

  $$\begin{aligned}\big(W_2(\pi_{X_1},\pi_{X_2})\big)^2 &= \inf_c\ \mathbb{E}\big[(X_1-X_2)^2\big] \\ &= \inf_c\ \mathbb{E}[X_1^2 - 2X_1X_2 + X_2^2] \\ &= \inf_c\ (\bar{x}_1^2+\sigma_1) + (\bar{x}_2+\sigma_2) - 2\big(c+\bar{x}_1\bar{x}_2\big) \\ &= \inf_c\ (\bar{x}_1-\bar{x}_2)^2 + \sigma_1+\sigma_2-2c\end{aligned}$$

  Thus it boils down to find the maximum eligible value of c. Recall the Cauchy-Schwarz inequality:

  $$\big(\operatorname{Cov}[X_1,X_2]\big)^2 \le \operatorname{Var}[X_1]\operatorname{Var}[X_2] \quad\Longrightarrow\quad c \le \sqrt{\sigma_1\sigma_2}$$

  Therefore by substitution,

  $$\begin{aligned}\big(W_2(\pi_{X_1},\pi_{X_2})\big)^2 &= (\bar{x}_1-\bar{x}_2)^2 + \sigma_1+\sigma_2 - 2\sqrt{\sigma_1\sigma_2} \\ &= (\bar{x}_1-\bar{x}_2)^2 + \big(\sqrt{\sigma_1}-\sqrt{\sigma_2}\big)^2\end{aligned}$$

- The KL divergence is given by

$$
\begin{aligned}
D_{KL}(\pi_{X_1} \| \pi_{X_2}) &= \int_{\mathbb{R}} \log \frac{\pi_{X_1}(x)}{\pi_{X_2}(x)} \pi_{X_1}(x)\, \mathrm{d}x \\
&= \int_{\mathbb{R}} \big[\log \pi_{X_1}(x) - \log \pi_{X_2}(x) \big] \pi_{X_1}(x)\, \mathrm{d}x \\
&= \int_{\mathbb{R}} \Big[-\log(\sqrt{2\pi\sigma_1}) - \frac{(x-\bar{x}_1)^2}{2\sigma_1} + \log(\sqrt{2\pi\sigma_2}) + \frac{(x-\bar{x}_2)^2}{2\sigma_2} \Big] \pi_{X_1}(x)\, \mathrm{d}x \\
&= \mathbb{E}\Big[\frac{1}{2} \log \frac{\sigma_2}{\sigma_1} - \frac{1}{2\sigma_1}(X_1 - \bar{x}_1)^2 + \frac{1}{2\sigma_2}(X_1 - \bar{x}_2)^2 \Big]
\end{aligned}
$$

Note that the first term is a constant, and the second term is the variance of X_1, that is,

$$
\frac{1}{2\sigma_1} \mathbb{E}\big[(X_1 - \bar{x}_1)^2 \big] = \frac{1}{2\sigma_1} \cdot \sigma_1 = \frac{1}{2}
$$

For the third term, observe that

$$
\begin{aligned}
\mathbb{E}\big[(X_1 - \bar{x}_2)^2 \big] &= \mathbb{E}[X_1^2 - 2\bar{x}_2 X_1 + \bar{x}_2^2] \\
&= (\sigma_1 + \bar{x}_1^2) - 2\bar{x}_1\bar{x}_2 + \bar{x}_2^2 \\
&= \sigma_1 + (\bar{x}_1 - \bar{x}_2)^2
\end{aligned}
$$

Collecting all three terms, we conclude that

$$
D_{KL}(\pi_{X_1} \| \pi_{X_2}) = \frac{1}{2} \Big[\log \frac{\sigma_2}{\sigma_1} + \big(\frac{\sigma_1}{\sigma_2} - 1 \big) + \frac{(\bar{x}_1 - \bar{x}_2)^2}{\sigma_2} \Big]
$$

Note that this value becomes zero if $\bar{x}_1 = \bar{x}_2$ and $\sigma_1 = \sigma_2$.