Inverse Problems — Example Sheet 3

Lucas Riedstra

December 1, 2020

Question 3. *Let $a \in \mathbb{R} \setminus \{0\}$. We consider the inverse problem $au + n = f_n$, where $u \in \mathbb{R}$ is the unknown parameter, $n \in \mathbb{R}$ is measurement noise, and $f_n \in \mathbb{R}$ is observed data. We assume that the noise and prior are Gaussian, $N \sim \mathrm{N}(0, \gamma^2)$ and $U \sim \mathrm{N}(m_0, \sigma_0^2)$, where $\gamma^2 > 0, \sigma_0^2 > 0$. Assume that the likelihood is given by*

$$L(f_n \mid u) := \frac{1}{\sqrt{2\pi\gamma^2}} \exp\left(-\frac{(au - f_n)^2}{2\gamma}\right).$$

(i) Compute the posterior measure $\mathbb{P}(U \in \cdot \mid aU + N = f_n)$.

Next, we assume that we take N independent observations of the data, i.e., we consider the likelihood

$$L\left(f_n^{(1:M)} \mid u\right) := \prod_{i=1}^{M} \frac{1}{\sqrt{2\pi\gamma^2}} \exp\left(-\frac{(au - f_n^{(i)})^2}{2\gamma^2}\right),$$

where $f_n^{(i)} \in \mathbb{R}$.

(ii) Compute the posterior measure $N(m_M, \sigma_M^2) := \mathbb{P}(U \in \cdot \mid aU + N = f_n^{(i)} (i = 1, \ldots, M))$.

(iii) Replace the data $f_n^{(1:M)}$ in the posterior by the random vector

$$F := \begin{pmatrix} au^\dagger \\ \vdots \\ au^\dagger \end{pmatrix} + \eta,$$

where $\eta \sim \mathrm{N}(0, \gamma^2 I)$ for some $u^\dagger \in \mathbb{R}$ and study the asymptotic behaviour of $\mathbb{E}[m_M], m_M, \sigma_M^2$ as $M \to \infty$. How do you explain your findings?

Proof. (i) We have, with $C = (2\pi\gamma^2)^{-1/2}$,

$$\int_{-\infty}^{\infty} L(f_n \mid v) \,\mathrm{d}v = C \int_{-\infty}^{\infty} \exp\left(-\frac{(av - f_n)^2}{2\gamma^2}\right) \mathrm{d}v = \frac{C}{a} \int_{-\infty}^{\infty} \exp\left(-\frac{(v - f_n)^2}{2\gamma^2}\right) \mathrm{d}v = \frac{1}{a}.$$

Thus we have

$$\frac{\mathrm{d}\mu_{\text{post}}}{\mathrm{d}\mu_0}(u) = \frac{a}{C} \exp\left(-\frac{(au - f_n)^2}{2\gamma^2}\right) = \frac{a}{C} \exp\left(-\frac{(u - f_n/a)^2}{2(\gamma/a)^2}\right),$$

which is the density of an $N(f_n/a, \gamma^2/a^2)$ distribution.

By question 1c, we have

$$\frac{\mathrm{d}\mu_{\text{post}}}{\mathrm{d}\lambda}(u) = \frac{\mathrm{d}\mu_{\text{post}}}{\mathrm{d}\mu_0}(u) \frac{\mathrm{d}\mu_0}{\mathrm{d}\lambda}(u),$$

which is (up to a constant) the product of an $N(f_n/a, \gamma^2/a^2)$ density with an $N(m_0, \sigma_0^2)$ density.

For this, we can use the following lemma:

Lemma 1. *Let $f_{\mu_1,\sigma_1}, f_{\mu_2,\sigma_2}$ be the density functions of* $\mathrm{N}(\mu_1,\sigma_1^2), \mathrm{N}(\mu_2,\sigma_2^2)$ *distributions respectively. Then the product $f_{\mu_1,\sigma_1} f_{\mu_2,\sigma_2}$ is proportional to an $f_{\mu_{\mathrm{prod}},\sigma_{\mathrm{prod}}}$ density, where*

$$\mu_{\mathrm{prod}} := \frac{\mu_1\sigma_2^2 + \mu_2\sigma_1^2}{\sigma_1^2+\sigma_2^2}, \quad \sigma_{\mathrm{prod}}^2 := \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}.$$

Proof. Since we are discussing proportionality, we only care about the exponents. We have

$$\begin{aligned}
\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(x-\mu_2)^2}{\sigma_2^2} &= \frac{(\sigma_1^2+\sigma_2^2)x^2 - 2(\mu_1\sigma_2^2+\mu_2\sigma_1^2)x + \mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2}{\sigma_1^2\sigma_2^2} \\
&= \frac{x^2 - 2\frac{\mu_1\sigma_2^2+\mu_2\sigma_1^2}{\sigma_1^2+\sigma_2^2}x + \frac{\mu_1^2\sigma_2^2+\mu_2^2\sigma_1^2}{\sigma_1^2+\sigma_2^2}}{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}} \\
&= \frac{\left(x - \frac{\mu_1\sigma_2^2+\mu_2\sigma_1^2}{\sigma_1^2+\sigma_2^2}\right)^2}{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}} + C,
\end{aligned}$$

where C is some constant independent of x. The claim follows. □

Plugging in our values we can compute the posterior density: it is an $\mathrm{N}(\mu_{\mathrm{prod}}, \sigma_{\mathrm{prod}}^2)$ density where

$$\begin{aligned}
\mu_{\mathrm{prod}} &= \frac{\frac{f_n\sigma_0^2}{a} + \frac{m_0\gamma^2}{a^2}}{\frac{\gamma^2}{a^2}+\sigma_0^2} = \frac{af_n\sigma_0^2 + m_0\gamma^2}{\gamma^2 + a^2\sigma_0^2}, \\
\sigma_{\mathrm{prod}}^2 &= \frac{\frac{\sigma_0^2\gamma^2}{a^2}}{\frac{\gamma^2}{a^2}+\sigma_0^2} = \frac{\sigma_0^2\gamma^2}{\gamma^2+a^2\sigma_0^2}.
\end{aligned}$$

(ii) We get similar computations as in the previous part, except that we have to compute the product of $N+1$ densities, namely $\mathrm{N}(f_n^{(1)}/a, \gamma^2/a^2), \ldots, \mathrm{N}(f_n^{(N)}/a, \gamma^2/a^2), \mathrm{N}(m_0, \sigma_0^2)$. Note that in the previous lemma, if $\sigma_1^2 = \sigma_2^2 = \sigma^2$, the formula gives

$$\mu_{\mathrm{prod}} = \frac{1}{2}(\mu_1+\mu_2), \quad \sigma_{\mathrm{prod}} = \frac{\sigma^2}{2},$$

and this generalises: for N observations we get

$$\mu_{\mathrm{prod}}^{(n)} = \frac{1}{n}(\mu_1 + \cdots + \mu_n) =: \bar{\mu}, \quad \sigma_{\mathrm{prod}}^{(n)} = \frac{\sigma^2}{n}.$$

This shows that the product of the $\mathrm{N}(f_n^{(1)}/a, \gamma^2/a^2), \ldots, \mathrm{N}(f_n^{(N)}/a, \gamma^2/a^2)$ densities is proportional to a $\mathrm{N}(\bar{f}_n/a, \gamma^2/(na^2))$ distribution, where $\bar{f}_n$ is the average of $f_n^{(1)}, \ldots, f_n^{(n)}$.

When we multiply this with prior density $\mathrm{N}(m_0, \sigma_0^2)$, we get

$$\begin{aligned}
\mu_M &= \frac{\frac{\bar{f}_n\sigma_0^2}{a} + \frac{m_0\gamma^2}{na^2}}{\frac{\gamma^2}{na^2}+\sigma_0^2} = \frac{na\bar{f}_n\sigma_0^2 + m_0\gamma^2}{\gamma^2+na^2\sigma_0^2}, \\
\sigma_M^2 &= \frac{\frac{\sigma_0^2\gamma^2}{na^2}}{\frac{\gamma^2}{na^2}+\sigma_0^2} = \frac{\sigma_0^2\gamma^2}{\gamma^2+na^2\sigma_0^2}.
\end{aligned}$$

(iii) Note that $\bar{F} = au^\dagger + \bar{\eta}$, and we know from elementary probability theory that if $\eta \sim N(0, \gamma^2 I)$, then for the average $\bar{\eta}$ we have $\bar{\eta} \sim N(0, \gamma^2/n)$. We get

$$\mathbb{E}[m_M] = \mathbb{E}\left[\frac{na(au^\dagger + \bar{\eta})\sigma_0^2 + m_0\gamma^2}{\gamma^2 + na^2\sigma_0^2}\right] = \frac{na^2u^\dagger\sigma_0^2 + m_0\gamma^2}{\gamma^2 + na^2\sigma_0^2} \overset{n\to\infty}{\to} u^\dagger,$$

so in the limit $n \to \infty$, we have $\mathbb{E}[m_M] \to u^\dagger$, which seems reasonable: the more observations we get, the less our prior assumptions are taken into account.

By the law of large numbers, we have

$$m_M = \frac{na^2\sigma_0^2u^\dagger + m_0\gamma^2}{na^2\sigma_0^2 + \gamma^2} + \frac{na\sigma_0^2\bar{\eta} + m_0\gamma^2}{na^2\sigma_0^2\gamma^2} \to u^\dagger,$$

since $\bar{\eta} \to 0$ as $n \to \infty$ by the law of large numbers.

Finally, since σ_M^2 is independent of the data (it depends only on the likelihood and the prior), we can simply let $n \to \infty$ in our expression for σ_M^2 and see $\sigma_M^2 \to 0$, which also makes sense: the more observations we get, the less variance we have.

□

Question 4. *Let $(\Omega, \mathcal{F})$ and let $\mathrm{Prob}(\Omega, \mathcal{F})$ be the space of probability measures on $(\Omega, \mathcal{F})$.*

(i) Show that $d_{\mathrm{TV}} \colon \mathrm{Prob}(\Omega, \mathcal{F})^2 \to [0, \infty) \colon (\mu, \nu) \mapsto \sup_{F\in\mathcal{F}} |\mu(F) - \nu(F)|$ is a metric on $\mathrm{Prob}(\Omega, \mathcal{F})$.

(ii) Let $\mu, \nu \in \mathrm{Prob}(\Omega, \mathcal{F})$ and ρ be a σ-finite measure with $\mu, \nu \ll \rho$. Show that

$$d_{\mathrm{TV}}(\mu, \nu) = \frac{1}{2}\int_\Omega \left|\frac{\mathrm{d}\mu}{\mathrm{d}\rho} - \frac{\mathrm{d}\nu}{\mathrm{d}\rho}\right| \mathrm{d}\rho\,.$$

(iii) Let $\mathcal{K} := \{h \colon (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}\mathbb{R}) \colon \sup_{\omega\in\Omega} |h(\omega)| \le 1\}$ and $\mu, \nu \in \mathrm{Prob}(\Omega, \mathcal{F})$. Show that

$$d_{\mathrm{TV}}(\mu, \nu) = \sup_{h\in\mathcal{K}} \frac{1}{2}\left|\int_\Omega h\,\mathrm{d}\mu - h\,\mathrm{d}\nu\right|.$$

(iv) Let Ω be a topological space and $(\Omega, \mathcal{F}) := (\Omega, \mathcal{B}\Omega)$. Let $(\mu_n)_{n\in\mathbb{N}} \in \mathrm{Prob}(\Omega, \mathcal{F})^{\mathbb{N}}$ and $\mu \in \mathrm{Prob}(\Omega, \mathcal{F})$. Show that

$$\lim_{n\to\infty} d_{\mathrm{TV}}(\mu_n, \mu) = 0 \implies \mu_n \to \mu \textit{ weakly, as } n \to \infty.$$

(v) Show that the converse of (iv) is in general not true.

Proof. (i) We check the metric definition. It is clear that d_{TV} is nonnegative and symmetric. Furthermore, we have

$$d_{\mathrm{TV}}(\mu, \nu) = 0 \implies \sup_{F\in\mathcal{F}} |\mu(F) - \nu(F)| = 0 \implies \mu(F) = \nu(F) \text{ for all } F \in \mathcal{F} \implies \mu = \nu.$$

Finally, if $\mu, \nu, \rho \in \mathrm{Prob}(\Omega, \mathcal{F})$, then by the "normal" triangle inequality we have

$$\begin{aligned}
d_{\mathrm{TV}}(\mu, \rho) &= \sup_{F\in\mathcal{F}} |\mu(F) - \rho(F)| = \sup_{F\in\mathcal{F}} |\mu(F) - \nu(F) + \nu(F) - \rho(F)| \\
&\le \sup_{F\in\mathcal{F}} (|\mu(F) - \nu(F)| + |\nu(F) - \rho(F)|) \\
&\le \sup_{F\in\mathcal{F}} |\mu(F) - \nu(F)| + \sup_{F\in\mathcal{F}} |\nu(F) - \rho(F)| \\
&= d_{\mathrm{TV}}(\mu, \nu) + d_{\mathrm{TV}}(\nu, \rho).
\end{aligned}$$

(ii) Write $A = \left\{x \in \Omega \mid \frac{\mathrm{d}\mu}{\mathrm{d}\rho}(x) > \frac{\mathrm{d}\nu}{\mathrm{d}\rho}(x)\right\}$ and $B = \left\{x \in \Omega \mid \frac{\mathrm{d}\mu}{\mathrm{d}\rho}(x) < \frac{\mathrm{d}\nu}{\mathrm{d}\rho}(x)\right\}$. Note that for any $X \subseteq A$ we have $\mu(X) \geq \nu(X)$ while for any $X \subseteq B$ we have $\mu(X) \leq \nu(X)$.

Also note that $\mu(A) - \nu(A) = \nu(B) - \mu(B)$. Therefore, we have

$$\begin{aligned}
d_{\mathrm{TV}}(\mu, \nu) &= \sup_{F \in \mathcal{F}} |\mu(F) - \nu(F)| \\
&= \sup_{F \in \mathcal{F}} |\mu(F \cap A) - \nu(F \cap A) + \mu(F \cap B) - \nu(F \cap B)| \\
&\leq \sup_{F \in \mathcal{F}} \max\{\mu(F \cap A) - \nu(F \cap A), \nu(F \cap B) - \mu(F \cap B)\} \\
&\leq \sup_{F \in \mathcal{F}} \max\{\mu(A) - \nu(A), \nu(B) - \mu(B)\} \\
&= \mu(A) - \nu(A).
\end{aligned}$$

Looking at the integral in the question, we see

$$\begin{aligned}
\frac{1}{2}\int_\Omega \left|\frac{\mathrm{d}\mu}{\mathrm{d}\rho} - \frac{\mathrm{d}\nu}{\mathrm{d}\rho}\right| \mathrm{d}\rho &= \frac{1}{2}\left(\int_A \left(\frac{\mathrm{d}\mu}{\mathrm{d}\rho} - \frac{\mathrm{d}\nu}{\mathrm{d}\rho}\right) \mathrm{d}\rho - \int_B \left(\frac{\mathrm{d}\mu}{\mathrm{d}\rho} - \frac{\mathrm{d}\nu}{\mathrm{d}\rho}\right) \mathrm{d}\rho\right) \\
&= \frac{1}{2}(\mu(A) - \nu(A) - \mu(B) + \nu(B)) \\
&= \mu(A) - \nu(A).
\end{aligned}$$

We conclude

$$d_{\mathrm{TV}}(\mu, \nu) = \mu(A) - \nu(A) = \frac{1}{2}\int_\Omega \left|\frac{\mathrm{d}\mu}{\mathrm{d}\rho} - \frac{\mathrm{d}\nu}{\mathrm{d}\rho}\right| \mathrm{d}\rho\,.$$

(iii) Let $\rho = \mu + \nu$, then we have $\mu \ll \rho$ and $\nu \ll \rho$, and ρ is (σ-)finite. Define A and B as in the solution to the previous exercise, then we have for $h \in \mathcal{K}$ that

$$\begin{aligned}
\frac{1}{2}\left|\int_\Omega h\,\mathrm{d}\mu - \int_\Omega h\,\mathrm{d}\nu\right| &= \frac{1}{2}\left|\int_\Omega h\left(\frac{\mathrm{d}\mu}{\mathrm{d}\rho} - \frac{\mathrm{d}\nu}{\mathrm{d}\rho}\right)\mathrm{d}\rho\right| \\
&\leq \frac{1}{2}\int_\Omega \left|\frac{\mathrm{d}\mu}{\mathrm{d}\rho} - \frac{\mathrm{d}\nu}{\mathrm{d}\rho}\right| \mathrm{d}\rho \\
&= \frac{1}{2}\int_A \left(\frac{\mathrm{d}\mu}{\mathrm{d}\rho} - \frac{\mathrm{d}\nu}{\mathrm{d}\rho}\right) \mathrm{d}\rho + \frac{1}{2}\int_B \left(\frac{\mathrm{d}\nu}{\mathrm{d}\rho} - \frac{\mathrm{d}\mu}{\mathrm{d}\rho}\right) \mathrm{d}\rho \\
&= \mu(A) - \nu(A) = d_{\mathrm{TV}}(\mu, \nu).
\end{aligned}$$

Furthermore, equality can be obtained by letting $h = \mathbb{1}_A - \mathbb{1}_B \in \mathcal{K}$, which concludes the proof.

(iv) Suppose that $d_{\mathrm{TV}}(\mu_n, \mu) \to 0$, and let $g\colon (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}\mathbb{R})$ be continuous and bounded. Since g is bounded, without loss of generality we can assume $\sup_{\omega\in\Omega} |g(\omega)| \leq 1$ (otherwise we divide by a constant). Now, by the previous exercise we have

$$d_{\mathrm{TV}}(\mu_n, \mu) \geq \frac{1}{2}\left|\int_\Omega g\,\mathrm{d}\mu - \int_\Omega g\,\mathrm{d}\mu_n\right|,$$

and since $d_{\mathrm{TV}}(\mu_n, \mu) \to 0$, we conclude that $\left|\int_\Omega g\,\mathrm{d}\mu - \int_\Omega g\,\mathrm{d}\mu_n\right| \to 0$, or equivalently that $\int_\Omega g\,\mathrm{d}\mu_n \to \int_\Omega g\,\mathrm{d}\mu$. Since g was arbitrarily chosen, we conclude that $\mu_n \to \mu$ weakly.

(v) Let μ_n be the measure on $(\mathbb{R}, \mathcal{B}\mathbb{R})$ corresponding to the uniform distribution on $[-\frac{1}{n}, \frac{1}{n}]$ with density function $f(x) = \frac{n}{2} \cdot \mathbb{1}_{[-\frac{1}{n}, \frac{1}{n}]}$, and let $\mu \coloneqq \delta_0$. We claim $\mu_n \to \mu$ weakly.

To prove this claim, let $g\colon (\mathbb{R}, \mathcal{B}\mathbb{R}) \to (\mathbb{R}, \mathcal{B}\mathbb{R})$ be continuous and bounded, and let $\varepsilon > 0$. Choose n large enough such that, on $[-1/n, 1/n]$, g takes values in $[g(0) - \varepsilon, g(0) + \varepsilon]$. Then we have

$$\int_{\mathbb{R}} g \,\mathrm{d}\mu_n = \frac{n}{2} \int_{-1/n}^{1/n} g(x) \,\mathrm{d}x \in [g(0) - \varepsilon, g(0) + \varepsilon],$$

and since ε was randomly chosen, we conclude $\int_{\mathbb{R}} g \,\mathrm{d}\mu_n \to g(0) = \int_{\mathbb{R}} g \,\mathrm{d}\delta_0$.

However, it is immediate that $d_{\mathrm{TV}}(\mu_n, \mu)$ does not converge to 0, since $\mu_n(\{0\}) = 0$ for all n while $\mu(\{0\}) = 1$.

$\square$