Inverse Problems — Example Sheet 3

Lucas Riedstra

December 1, 2020

Question 1. *Let $(\Omega, \mathcal{F})$ be a measurable space and μ, ν, ρ be σ-finite measures on $(\Omega, \mathcal{F})$. Prove the following statements.*

(a) Let $\nu \ll \mu$ and $a \geq 0$. Then, $a \cdot \nu \ll \mu$ and $\frac{\mathrm{d}a\cdot\nu}{\mathrm{d}\mu} = a\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ (μ-a.e.).

(b) Let $\nu \ll \mu$ and $\rho \ll \mu$. Then, $\nu + \rho \ll \mu$ and $\frac{\mathrm{d}\nu+\rho}{\mathrm{d}\mu} = \frac{\mathrm{d}\nu}{\mathrm{d}\mu} + \frac{\mathrm{d}\rho}{\mathrm{d}\mu}$ (μ-a.e.).

(c) Let $\rho \ll \nu$ and $\nu \ll \mu$. Then, $\rho \ll \mu$ and $\frac{\mathrm{d}\rho}{\mathrm{d}\mu} = \frac{\mathrm{d}\rho}{\mathrm{d}\nu} \cdot \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ (μ-a.e.).

Proof. (a) We have, for $F \in \mathcal{F}$, that $\mu(F) = 0 \implies \nu(F) = 0 \implies (a \cdot \nu)(F) = 0$. Furthermore, we have

$$(a \cdot \nu)(F) = a \cdot (\nu(F)) = a \cdot \int_F \frac{\mathrm{d}\nu}{\mathrm{d}\mu}\,\mathrm{d}\mu = \int_F \left(a\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right)\mathrm{d}\mu\,,$$

which proves the Radon-Nikodym derivative of $a \cdot \nu$ is $a\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$.

(b) We have, for $F \in \mathcal{F}$, that $\mu(F) = 0 \implies \nu(F) = 0$ and $\rho(F) = 0$ and therefore also $(\nu+\rho)(F) = 0$. Furthermore, we have

$$(\nu + \rho)(F) = \nu(F) + \rho(F) = \int_F \frac{\mathrm{d}\nu}{\mathrm{d}\mu}\,\mathrm{d}\mu + \int_F \frac{\mathrm{d}\rho}{\mathrm{d}\mu}\,\mathrm{d}\mu = \int_F \left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu} + \frac{\mathrm{d}\rho}{\mathrm{d}\mu}\right)\mathrm{d}\mu\,,$$

which proves the Radon-Nikodym derivative of $\nu + \rho$ is $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} + \frac{\mathrm{d}\rho}{\mathrm{d}\mu}$.

(c) We have, for $F \in \mathcal{F}$, that $\mu(F) = 0 \implies \nu(F) = 0 \implies \rho(F) = 0$. Furthermore, we have

$$\rho(F) = \int_F \frac{\mathrm{d}\rho}{\mathrm{d}\nu}\,\mathrm{d}\nu \stackrel{\star}{=} \int_F \frac{\mathrm{d}\rho}{\mathrm{d}\nu}\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\,\mathrm{d}\mu\,,$$

which proves the Radon-Nikodym derivative of ρ is $\frac{\mathrm{d}\rho}{\mathrm{d}\nu}\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$. Here, $\star$ follows from the fact that if a measure ν has μ-density g, then $\int_{\mathcal{X}} f\,\mathrm{d}\nu = \int_{\mathcal{X}} fg\,\mathrm{d}\mu$ for all $f \geq 0$ — this can be proved first for simple functions, and then extended to nonnegative integrable functions.

□

Question 2. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $U, U' \colon (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}\mathbb{R})$, $Y \colon (\Omega, \mathcal{F}) \to (\mathcal{Y}, \mathcal{B}\mathcal{Y})$ random variables. Moreover, let U and U' be integrable. Prove the following statements:*

(a) Let $c \in \mathbb{R}$ and $\mathbb{P}(U = c) = 1$. Then, $\mathbb{E}[U \mid Y = y] = c$ ($\mathbb{P}(Y \in \cdot)$-a.s.).

(b) Let $c \in \mathbb{R}$. Then $\mathbb{E}[cU \mid Y = y] = c\mathbb{E}[U \mid Y = y]$ ($\mathbb{P}(Y \in \cdot)$-a.s.).

(c) $\mathbb{E}[U + U' \mid Y = y] = \mathbb{E}[U \mid Y = y] + \mathbb{E}[U' \mid Y = y]$ ($\mathbb{P}(Y \in \cdot)$-a.s.).

Proof. Let $F \in \mathcal{B}\mathcal{Y}$.

(a) We have for all F

$$\int_{\{Y\in F\}} U\,\mathrm{d}\mathbb{P} = \int_{\{Y\in F\}} c\,\mathrm{d}\mathbb{P} = c\mathbb{P}(Y\in F) = \int_F c\mathbb{P}(Y\in \mathrm{d}y).$$

(b) We have for all F

$$\int_{\{Y\in F\}} cU\,\mathrm{d}\mathbb{P} = c\int_{\{Y\in F\}} U\,\mathrm{d}\mathbb{P} = c\int_F \mathbb{E}[U\mid Y=y]\mathbb{P}(Y\in \mathrm{d}y) = \int_F c\mathbb{E}[U\mid Y=y]\mathbb{P}(Y\in \mathrm{d}y).$$

(c) We have for all F

$$\begin{aligned}\int_{\{Y\in F\}} (U+U')\,\mathrm{d}\mathbb{P} &= \int_{\{Y\in F\}} U\,\mathrm{d}\mathbb{P} + \int_{\{Y\in F\}} U'\,\mathrm{d}\mathbb{P}\\ &= \int_F \mathbb{E}[U\mid Y=y]\mathbb{P}(Y\in \mathrm{d}y) + \int_F \mathbb{E}[U'\mid Y=y]\mathbb{P}(Y\in \mathrm{d}y)\\ &= \int_F (\mathbb{E}[U\mid Y=y] + \mathbb{E}[U'\mid Y=y])\mathbb{P}(Y\in \mathrm{d}y).\end{aligned}$$

□

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 $N(m_M,\sigma_M^2) := \mathbb{P}(U\in\cdot\mid aU+N=f_n^{(i)} (i=1,\dots,M))$.

(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 $f_{\mu_{\text{prod}},\sigma_{\text{prod}}}$ density, where*

$$\mu_{\text{prod}} \coloneqq \frac{\mu_1\sigma_2^2+\mu_2\sigma_1^2}{\sigma_1^2+\sigma_2^2}, \quad \sigma_{\text{prod}}^2 \coloneqq \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}.$$

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. $\square$

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

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

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

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

and this generalises: for N observations we get

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

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

$$\mu_M = \frac{\frac{\bar{f}_n \sigma_0^2}{a} + \frac{m_0\gamma^2}{na^2}}{\frac{\gamma^2}{na^2} + \sigma_0^2} = \frac{na\bar{f}_n\sigma_0^2 + m_0\gamma^2}{\gamma^2 + na^2\sigma_0^2},$$

$$\sigma_M^2 = \frac{\frac{\sigma_0^2\gamma^2}{na^2}}{\frac{\gamma^2}{na^2} + \sigma_0^2} = \frac{\sigma_0^2\gamma^2}{\gamma^2 + na^2\sigma_0^2}.$$

(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

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

(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 $\mu(X)\ge\nu(X)$ while for any $X\subseteq B$ we have $\mu(X)\le\nu(X)$.

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

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

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

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

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 $\mu := \delta_0$. 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 $\mu(\{0\}) = 1$.

$\square$

Question 5. *Let μ be a probability measure on $(\mathbb{R}, \mathcal{B}\mathbb{R})$ with cumulative distribution function (cdf) $F\colon \mathbb{R} \to [0,1]\colon x \mapsto \mu((-\infty, x])$, for $x \in \mathbb{R}$.*

(i) Let $Q\colon (0,1) \to \mathbb{R}$ be the quantile function of μ, i.e., $Q(y) := \inf\{x \in \mathbb{R} \mid F(x) \geq y\}$. Moreover, let $U \sim \mathrm{Unif}(0,1) := \lambda_1(\cdot \cap (0,1))$ be a uniformly distributed random variable on the interval $(0,1)$. Show that $\mathbb{P}(Q(U) \in \cdot) = \mu$. (Hint: you may use the fact that probability measures on $(\mathbb{R}, \mathcal{B}\mathbb{R})$ are uniquely determined by their CDF).

(ii) Derive the quantile function for the exponential distribution, i.e., the distribution with cdf $F(x) := 1 - \exp(-\lambda x)$ for some $\lambda > 0$.

(iii) Use the idea from (i) and your quantile function from (ii) to generate independent samples (i.e., realisations of random variables) with $\lambda = 1$.

(a) Plot the cdf of the exponential distribution along with the empirical cdf of

$$M \in \{10, 100, 1000, 10000\}$$

of your samples. (Hint: you can use the `ecdf` command to obtain a representation of your empirical cdf). What do you observe?

(b) Compute the sample mean $\bar{X}_M$ of your exponentially distributed samples $X_1, \ldots, X_m$; for $M = \{2^n \mid n = 1, \ldots, 20\}$. What do you observe?

(iv) Repeat (iii)(b) using the quantile function $Q'(y) := \tan(\pi(y - 0.5))$ $(y \in \mathbb{R})$. What do you observe?

(v) Let $X_1, \ldots, X_M \sim \mu$ be i.i.d., and assume that $\mathrm{Var}(X_1)$ is finite. Show that

$$\mathbb{E}\Big[\big(\mathbb{E}[X_1] - \bar{X}_M\big)^2\Big] = \frac{\mathrm{Var}(X_1)}{M}.$$

(vi) Can you recover the rate from (v) in your experiments in (iii)? What could be the issue in (iv)?

Proof. (i) Since probability measures on $(\mathbb{R}, \mathcal{B}\mathbb{R})$ are uniquely determined by their CDF, we must show that the CDF of μ agrees with the CDF of $\mathbb{P}(Q(U) \in \cdot)$, i.e., for all $x \in \mathbb{R}$

$$F(x) = \mu((-\infty, x]) = \mathbb{P}(Q(U) \in (-\infty, x]).$$

For this, note that

$$\mathbb{P}(Q(U) \in (-\infty, x]) = \mathbb{P}(Q(U) \leq x) \stackrel{\star}{=} \mathbb{P}(U \leq F(x)) = F(x),$$

(ii) For the exponential distribution, the cdf F is an invertible function between $(0, \infty)$ and $(0, 1)$, so the quantile function is just the inverse. We have

$$y = 1 - e^{-\lambda x} \iff e^{-\lambda x} = 1 - y \iff -\lambda x = \ln(1 - y) \iff x = -\frac{\ln(1-y)}{\lambda},$$

so the quantile function is given by $Q(y) = -\frac{\ln(1-y)}{\lambda}$.

(iii) (a) From approximately $M = 1000$, the empirical CDF aligns almost exactly with the true cdf.

(b) Keeps getting closer to 1.

(iv) The means are all over the place and don't seem to converge to anything.

(v) We have

$$\begin{aligned}
\mathbb{E}\Big[\mathbb{E}[X_1]^2 - 2\mathbb{E}[X_1]\bar{X}_M + \bar{X_M}^2\Big] &= \mathbb{E}[X_1]^2 - 2\mathbb{E}[X_1]\mathbb{E}[\bar{X}_M] + \mathbb{E}[\bar{X}_M^2] \\
&= \mathbb{E}[X_1]^2 - 2\mathbb{E}[X_1]^2 + \frac{1}{M^2}\sum_{ij}\mathbb{E}[X_i X_j] \\
&= -\mathbb{E}[X_1]^2 + \frac{1}{M}\mathbb{E}[X_1^2] + \frac{M-1}{M}\mathbb{E}[X_1]^2 \\
&= \frac{\mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2}{M} = \frac{1}{M}\operatorname{Var}(X_1).
\end{aligned}$$

(vi) In (iv), the issue is that the distribution does not have finite mean or variance.

□