Exercise Sheet 4

These exercises relate to the material covered in the lecture of Week 4, and possibly previous weeks' lectures and exercises. Please submit your solutions to these exercises at the beginning of the lecture of Week 6, i.e. 12:00 on 19 November 2015. Environmentally-friendly submissions by e-mail in PDF form are welcomed! The numbers in the margin indicate approximately how many points are available for each part.

Throughout, $\mathcal{M}_1(\mathcal{X}, \mathscr{F})$ denotes the space of probability measures on a measurable space $(\mathcal{X}, \mathscr{F})$.

Exercise 4.1. Let $\mu_0 = \mathcal{N}(m_0, C_0)$ and $\mu_1 = \mathcal{N}(m_1, C_1)$ be non-degenerate Gaussian probability measures on $\mathbb{R}^n$. Show that **[6]**

$$D_{\mathrm{KL}}(\mu_0 \| \mu_1) = \frac{1}{2}\left(\log \frac{\det C_1}{\det C_0} - n + \operatorname{tr}(C_1^{-1} C_0) + \left\| m_0 - m_1 \right\|_{C_1^{-1}}^2 \right). \tag{4.1}$$

You may use the chain rule for Radon–Nikodým derivatives: if λ, μ, and ν are arbitrary σ-finite measures on $(\mathcal{X}, \mathscr{F})$ and $\lambda \ll \mu \ll \nu$, then

$$\frac{\mathrm{d}\lambda}{\mathrm{d}\nu}(x) = \frac{\mathrm{d}\lambda}{\mathrm{d}\mu}(x) \frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x) \quad \text{for } \nu\text{-a.e. } x \in \mathcal{X}.$$

You may also use the fact that, if $X \sim \mathcal{N}(m, C)$ is an $\mathbb{R}^n$-valued Gaussian random vector and $A \in \mathbb{R}^{n \times n}$ is symmetric, then

$$\mathbb{E}[X \cdot AX] = \operatorname{tr}(AC) + m \cdot Am.$$

Optional extra exercise, good for developing intuition: use formula (4.1) to *informally* explain why the covariance operator of a Gaussian measure on an infinite-dimensional Banach space must have summable eigenvalues (i.e. must be trace class), and why Gaussian measures on infinite-dimensional spaces are either equivalent or mutually singular (cf. the Feldman–Hájek and Cameron–Martin theorems).

Exercise 4.2. Let $f \colon [0, \infty] \to \mathbb{R} \cup \{+\infty\}$ be convex with $f(0) = 1$, and let

$$D_f(\mu \| \nu) := \begin{cases} \displaystyle\int_{\mathcal{X}} f\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x) \right) \mathrm{d}\nu(x), & \text{if } \mu \ll \nu \text{ and } \frac{\mathrm{d}\mu}{\mathrm{d}\nu} \notin L^1(\mathcal{X}, \nu), \\ +\infty, & \text{otherwise,} \end{cases}$$

be the induced f-divergence on $\mathcal{M}_1(\mathcal{X}, \mathscr{F})$.

(a) Show that $(x, y) \mapsto y f(x/y)$ is a convex function from $(0, \infty) \times (0, \infty)$ to $\mathbb{R} \cup \{+\infty\}$. **[6]**

(b) Hence show that $D_f(\,\cdot\,\|\,\cdot\,)$ is jointly convex in its two arguments, i.e. for all μ_0, μ_1, ν_0, and $\nu_1 \in \mathcal{M}_1(\mathcal{X}, \mathscr{F})$ and $t \in [0, 1]$, **[6]**

$$D_f\big((1-t)\mu_0 + t\mu_1 \big\| (1-t)\nu_0 + t\nu_1\big) \leq (1-t) D_f(\mu_0 \| \nu_0) + t D_f(\mu_1 \| \nu_1).$$

Exercise 4.3. The following result is a useful one that frequently allows statements about f-divergences to be reduced to the case of a finite or countable sample space. Let $\mu \in \mathcal{M}_1(\mathcal{X}, \mathscr{F})$, and let $f \colon [0, \infty] \to [0, \infty]$ be convex. Given a partition $\mathcal{A} = \{A_n \mid n \in \mathbb{N}\}$ of $\mathcal{X}$ into countably many pairwise disjoint measurable sets, define $\mu_{\mathcal{A}} \in \mathcal{M}_1(\mathbb{N}, 2^{\mathbb{N}})$ by $\mu_{\mathcal{A}}(n) := \mu(A_n)$.

(a) Suppose that $\mu(A_n) > 0$ and that $\mu \ll \nu$, where μ and ν are σ-finite measures on $(\mathcal{X}, \mathscr{F})$ but not necessarily probability measures. Show that, for each $n \in \mathbb{N}$, **[6]**

$$\frac{1}{\nu(A_n)} \int_{A_n} f\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu} \right) \mathrm{d}\nu \geq f\left(\frac{\mu(A_n)}{\nu(A_n)} \right).$$

(b) Hence prove the following result, known as the *partition inequality*: for any $\mu, \nu \in \mathcal{M}_1(\mathcal{X}, \mathscr{F})$ with $\mu \ll \nu$,

$$D_f(\mu \| \nu) \geq D_f(\mu_{\mathcal{A}} \| \nu_{\mathcal{A}}). \tag{4.2}$$

Show also that, for strictly convex f, equality holds in (4.2) if and only if $\mu(A_n) = \nu(A_n)$ for each $n \in \mathbb{N}$. **[10]**

Exercise 4.4. Show that Pinsker's inequality

$$2 d_{\mathrm{TV}}(\mu, \nu) \leq D_{\mathrm{KL}}(\mu \| \nu)$$

cannot be reversed. In particular, give an example of a measurable space $(\mathcal{X}, \mathscr{F})$ such that, for any $\varepsilon > 0$, there exist $\mu, \nu \in \mathcal{M}_1(\mathcal{X}, \mathscr{F})$ with $d_{\mathrm{TV}}(\mu, \nu) \leq \varepsilon$ but $D_{\mathrm{KL}}(\mu \| \nu) = +\infty$. Hint: consider a 'small' perturbation to the cumulative distribution function of a probability measure on $\mathbb{R}$. **[6]**

Exercise 4.5. Let $(\mathcal{V}, \| \cdot \|)$ be a Banach space, and suppose that $f \colon \mathcal{X} \to \mathcal{V}$ has finite second moment with respect to $\mu, \nu \in \mathcal{M}_1(\mathcal{X})$. Show that **[10]**

$$\big\| \mathbb{E}_\mu[f] - \mathbb{E}_\nu[f] \big\| \leq 2 \sqrt{\mathbb{E}_\mu[\|f\|^2] + \mathbb{E}_\nu[\|f\|^2]} \, d_{\mathrm{H}}(\mu, \nu),$$

where d_{H} denotes the Hellinger distance

$$\begin{aligned} d_{\mathrm{H}}(\mu, \nu)^2 &= \int_{\mathcal{X}} \left| \sqrt{\frac{\mathrm{d}\mu}{\mathrm{d}\nu}} - 1 \right|^2 \mathrm{d}\nu \\ &= \int_{\mathcal{X}} \left| \sqrt{\frac{\mathrm{d}\mu}{\mathrm{d}\rho}} - \sqrt{\frac{\mathrm{d}\nu}{\mathrm{d}\rho}} \right|^2 \mathrm{d}\rho \end{aligned}$$

for any reference measure ρ, with the usual convention that $d_{\mathrm{H}}(\mu, \nu) = +\infty$ when the necessary densities do not exist or are not integrable.