Chapter 12

Contrast Functions on Statistical Models

This chapter deals with some important examples of contrast functions on a space of density functions, such as: Bregman divergence, Kullback–Leibler relative entropy, f-divergence, Hellinger distance, Chernoff information, Jefferey distance, Kagan divergence, and exponential contrast function. The relation with the skewness tensor and α-connection is made. The goal of this chapter is to produce hands-on examples for the theoretical concepts introduced in Chap. 11.

12.1 A First Example

We start with a suggestive example of Bregman divergence. We show that the Kullback–Leibler relative entropy on a statistical model is a particular example of Bregman divergence.

Let $\mathcal{S} = \mathcal{P}(\mathcal{X})$, where $\mathcal{X} = \{x^1, \dots, x^{n+1}\}$ and consider the global chart $\phi : \mathcal{S} \to \mathbb{E} \subset \mathbb{R}^n$

$$\phi(p) = (\ln p_1, \dots, \ln p_n) = (\xi^1, \dots, \xi^n),$$

with the parameter space

$$\mathbb{E} = \{(\xi^1, \dots, \xi^n); \xi^k > 0, \sum_{k=1}^{n} \xi^k < 1\}.$$

O. Calin and C. Udrişte, *Geometric Modeling in Probability and Statistics*,
DOI 10.1007/978-3-319-07779-6_12,

The contrast function on $\mathcal{S}$ is then given by

$$\begin{aligned} D_{\mathcal{S}}(p||q) &= D_{\mathcal{S}}\big(\phi^{-1}(p)||\phi^{-1}(q)\big) \\ &= D(\xi_1||\xi_2), \end{aligned}$$

where $D(\cdot\,||\,\cdot)$ is the Bregman divergence on $\mathbb{E}$ induced by the convex function $\varphi(\xi) = \sum_{i=1}^{n} e^{\xi_i}$, i.e.,

$$D(\xi_1||\xi_2) = \varphi(\xi_2) - \varphi(\xi_1) - \sum_{i=1}^{n} \partial_i\varphi(\xi_1)(\xi_2^i - \xi_1^i).$$

Therefore

$$\begin{aligned} D_{\mathcal{S}}(p||q) &= D(\xi_1||\xi_2) \\ &= \sum_i e^{\xi_2^i} - \sum_i e^{\xi_1^i} - \sum_i e^{\xi_1^i}(\xi_2^i - \xi_1^i) \\ &= \sum_i p_i - \sum_i q_i - \sum_i p_i \ln\frac{q_i}{p_i} \\ &= \sum_i p_i \ln\frac{p_i}{q_i} = D_{KL}(p||q). \end{aligned}$$

Hence, the induced contrast function $D_{\mathcal{S}}$ on $\mathcal{P}(\mathcal{X})$ in this case is the Kullback–Leibler relative entropy.

12.2 f-Divergence

An important class of contrast functions on statistical models was introduced by Csiszár [31, 32]. Let $f : (0, \infty) \to \mathbb{R}$ be a function satisfying the following conditions

(a) f is convex;

(b) $f(1) = 0$;

(c) $f''(1) = 1$.

For each probability distributions p, q, consider

$$D_f(p||q) = E_p\Big[f\Big(\frac{q(x)}{p(x)}\Big)\Big] = \int_{\mathcal{X}} p(x) f\Big(\frac{q(x)}{p(x)}\Big)\,dx. \tag{12.2.1}$$

We shall assume that the previous integral converges and we can differentiate under the integral sign.

Proposition 12.2.1 *The operator $D_f(\cdot\,||\,\cdot)$ is a contrast function on the statistical model $\mathcal{S} = \{p_\xi\}$.*

Proof: We check the properties of a contrast function.

(i) positive: Jensen's inequality applied to the convex function f provides

$$\begin{aligned} D_f(p||q) &= E_p\Big[f\Big(\frac{q(x)}{p(x)}\Big)\Big] \geq f\Big(E_p\Big[\frac{q(x)}{p(x)}\Big]\Big) \\ &= f\Big(\int_{\mathcal{X}} p(x)\frac{q(x)}{p(x)}\,dx\Big) = f(1) = 0. \end{aligned}$$

(ii) non-degenerate: Let $p \neq q$. Since f is strictly convex at 1, then

$$D_f(p||q) = E_p\Big[f\Big(\frac{q(x)}{p(x)}\Big)\Big] > f\Big(E_p\Big[\frac{q(x)}{p(x)}\Big]\Big) = f(1) = 0,$$

and hence $D(p||q) \neq 0$, which implies the non-degenerateness.

(iii) The vanishing property of the first variation along the diagonal $\{\xi_1 = \xi_2\}$ is a consequence of (i) and (ii).

(iv) Let $p = p_{\xi_0}$ and $q = p_\xi$. We shall compute the Hessian of

$$D_f(p_{\xi_0}||p_\xi) = \int_{\mathcal{X}} p_{\xi_0}(x) f\Big(\frac{p_\xi(x)}{p_{\xi_0}(x)}\Big)\,dx \qquad (12.2.2)$$

along the diagonal $\xi^0 = \xi$. Differentiating we have

$$\begin{aligned} \partial_{\xi^j} f\Big(\frac{p_\xi}{p_{\xi_0}}\Big) &= f'\Big(\frac{p_\xi}{p_{\xi_0}}\Big)\frac{1}{p_{\xi_0}}\partial_{\xi^j} p_\xi \\ \partial_{\xi^i}\partial_{\xi^j} f\Big(\frac{p_\xi}{p_{\xi_0}}\Big) &= f''\Big(\frac{p_\xi}{p_{\xi_0}}\Big)\Big(\frac{p_\xi}{p_{\xi_0}}\Big)^2 \partial_{\xi^i}(\ln p_\xi)\partial_{\xi^j}(\ln p_\xi) \\ &\quad + f'\Big(\frac{p_\xi}{p_{\xi_0}}\Big)\frac{1}{p_\xi}\partial_{\xi^i}\partial_{\xi^j} p_\xi . \end{aligned}$$

Differentiating under the integral we get

$$\begin{aligned}
\partial_{\xi^i}\partial_{\xi^j} D_f(p_{\xi_0}||p_\xi)_{|\xi=\xi_0} &= f''(1)\int p_{\xi_0}\partial_{\xi^i}\ln p_\xi\,\partial_{\xi^j}\ln p_\xi\,dx_{|\xi=\xi_0} \\
&\quad + f'(1)\partial_{\xi^i}\partial_{\xi^j}\int p_\xi(x)\,dx \\
&= f''(1)E_\xi[\partial_{\xi^i}\ell(\xi)\partial_{\xi^j}\ell(\xi)] \\
&= E_\xi[(\partial_{\xi^i}\ell)(\partial_{\xi^j}\ell)] = g_{ij}(\xi),
\end{aligned}$$

which is strictly positive definite, since it is the Fisher–Riemann information matrix. Hence $D_f(\cdot\,||\,\cdot)$ is a contrast function.

■

Theorem 12.2.2 *The Riemannian metric induced by the contrast function $D_f(\cdot\,||\,\cdot)$ on the statistical model $\mathcal{S} = \{p_\xi\}$ is the Fisher–Riemann information matrix*

$$g_{ij}(\xi) = \partial_{\xi^i}\partial_{\xi^j} D_f(p_{\xi_0}||p_\xi)_{|\xi=\xi_0}.$$

Proof: It follows from the calculation performed in the part (iv) above. ■

Let $f^*(u) = uf\Big(\frac{1}{u}\Big)$. Since

$$\begin{aligned}
f^*(1) &= f(1) = 0 \\
f^{*\prime\prime}(u) &= \frac{1}{u^3}f''\Big(\frac{1}{u}\Big) \geq 0 \\
f^{*\prime\prime}(1) &= f''(1) = 1,
\end{aligned}$$

then f^* satisfies properties (a)–(c), and hence $D_{f^*}(\cdot\,||\,\cdot)$ is a contrast function, which defines the same Riemannian metric as $D_f(\cdot\,||\,\cdot)$.

Proposition 12.2.3 *The contrast function $D_{f^*}(\cdot\,||\,\cdot)$ is the dual of $D_f(\cdot\,||\,\cdot)$.*

Proof: Consider the dual $D_f^*(p||q) = D_f(q||p)$. Then we have

$$\begin{aligned}
D_{f^*}(p||q) &= \int_{\mathcal{X}} p(x)f^*\Big(\frac{q(x)}{p(x)}\Big)\,dx \\
&= \int_{\mathcal{X}} p(x)\frac{q(x)}{p(x)}f\Big(\frac{p(x)}{q(x)}\Big)\,dx \\
&= \int_{\mathcal{X}} q(x)f\Big(\frac{p(x)}{q(x)}\Big)\,dx \\
&= D_f(q||p) = D_f^*(p||q), \qquad \forall p,q\in\mathcal{S}.
\end{aligned}$$

Therefore $D_{f^*} = D_f^*$. ■

In the following we shall find the induced connections. Let $\nabla^{(f)}$ be the linear connection induced by the contrast function $D_f(\cdot\,||\,\cdot)$, and denote by $\Gamma^{(f)}_{ij,k}$ its components on a local basis.

Proposition 12.2.4 *We have*

$$\Gamma^{(f)}_{ij,k}(\xi) = E_\xi\Big[\big(\partial_i\partial_j\ell - (f'''(1)+1)\partial_i\partial_j\ell\big)\partial_k\ell\Big]. \tag{12.2.3}$$

Proof: From formula (11.5.18) we find

$$\Gamma^{(f)}_{ij,k}(\xi) = -\partial_{\xi_0^i}\partial_{\xi_0^j}\partial_{\xi^k}D_f(p_{\xi_0}||p_\xi)_{|\xi=\xi_0}. \tag{12.2.4}$$

We shall compute the derivatives on the right side. Differentiating in (12.2.2) yields

$$\partial_{\xi^k}D_f(p_{\xi_0}||p_\xi) = \int_{\mathcal{X}} f'\Big(\frac{p_\xi}{p_{\xi_0}}\Big)p_\xi\partial_{\xi^k}\ell(\xi)\,dx. \tag{12.2.5}$$

Before continuing the computation we note that

$$\begin{aligned}
\partial_{\xi_0^j}f'\Big(\frac{p_\xi}{p_{\xi_0}}\Big) &= f''\Big(\frac{p_\xi}{p_{\xi_0}}\Big)\Big(\frac{-p_\xi}{p_{\xi_0}}\Big)\partial_{\xi_0^j}\ell(\xi_0)\\
\partial_{\xi_0^i}\partial_{\xi_0^j}f'\Big(\frac{p_\xi}{p_{\xi_0}}\Big) &= f'''\Big(\frac{p_\xi}{p_{\xi_0}}\Big)\frac{p_\xi^2}{p_{\xi_0}^2}\partial_{\xi_0^i}\ell(\xi_0)\partial_{\xi_0^j}\ell(\xi_0)\\
&\quad +f''\Big(\frac{p_\xi}{p_{\xi_0}}\Big)\frac{p_\xi}{p_{\xi_0}}\partial_{\xi_0^i}\ell(\xi_0)\partial_{\xi_0^j}\ell(\xi_0)\\
&\quad -f''\Big(\frac{p_\xi}{p_{\xi_0}}\Big)\frac{p_\xi}{p_{\xi_0}}\partial_{\xi_0^i}\partial_{\xi_0^j}\ell(\xi_0).
\end{aligned}$$

Applying now $\partial_{\xi_0^i}\partial_{\xi_0^j}$ to (12.2.5), using the foregoing formulas, and taking $\xi_0 = \xi$, yields

$$\begin{aligned}
\partial_{\xi_0^i}\partial_{\xi_0^j}\partial_{\xi^k}D_f(p_{\xi_0}||p_\xi)_{|\xi=\xi_0} &= \int_{\mathcal{X}}\Big[(f'''(1)+f''(1))p_\xi(\partial_{\xi^i}\ell)(\partial_{\xi^j}\ell)(\partial_{\xi^k}\ell)\\
&\quad -f''(1)(\partial_{\xi^i}\partial_{\xi^j}\ell)(\partial_{\xi^k}\ell)\Big]\,dx\\
&= E_\xi\Big[\big(\partial_{\xi^i}\partial_{\xi^j}\ell - (f'''(1)+1)\partial_{\xi^i}\partial_{\xi^j}\ell\big)\partial_{\xi^k}\ell\Big].
\end{aligned}$$

Applying (12.2.4) we arrive at (12.2.3). ∎

The relation with the geometry of α-connections is given below.

Theorem 12.2.5 *The connection induced by $D_f(\cdot\,||\,\cdot)$ is an α-connection*

$$\nabla^{(f)} = \nabla^{(\alpha)},$$

with $\alpha = 2f'''(1) + 3$.

Proof: It suffices to show the identity in local coordinates. Recall first the components of the α-connection given by (1.11.34)

$$\Gamma^{(\alpha)}_{ij,k} = E_\xi\Big[\Big(\partial_i\partial_j\ell + \frac{1-\alpha}{2}\partial_i\ell\partial_j\ell\Big)\partial_k\ell\Big]. \tag{12.2.6}$$

Comparing with (12.2.3) we see that $\Gamma^{(f)}_{ij,k} = \Gamma^{(\alpha)}_{ij,k}$ if and only if $\alpha = 2f'''(1) + 3$. ■

We make the remark that $\nabla^{(f^*)} = \nabla^{(-\alpha)}$, which follows from the properties of dual connections induced by contrast functions. We shall show shortly that for any α there is a function f satisfying (a)–(c) and solving the equation $\alpha = 2f'''(1) + 3$.

Proposition 12.2.6 *The skewness tensor induced by the contrast function $D_f(\cdot\,||\,\cdot)$ is given in local coordinates by*

$$T^{(f)}_{ijk} = (2f'''(1) + 3)E_\xi[(\partial_i\ell)(\partial_j\ell)(\partial_k\ell)].$$

Proof: Using Theorem 12.2.5, formula (12.2.6) and the aforementioned remarks, we have

$$\begin{aligned} T^{(f)}_{ijk} &= \Gamma^{(f^*)}_{ijk} - \Gamma^{(f)}_{ijk} = \Gamma^{(-\alpha)}_{ijk} - \Gamma^{(\alpha)}_{ijk} \\ &= E_\xi\Big[\Big(\partial_i\partial_j\ell + \frac{1+\alpha}{2}\partial_i\ell\partial_j\ell\Big)\partial_k\ell\Big] \\ &\quad -E_\xi\Big[\Big(\partial_i\partial_j\ell + \frac{1-\alpha}{2}\partial_i\ell\partial_j\ell\Big)\partial_k\ell\Big] \\ &= \alpha E_\xi[(\partial_i\ell)(\partial_j\ell)(\partial_k\ell)] \\ &= (2f'''(1) + 3)E_\xi[(\partial_i\ell)(\partial_j\ell)(\partial_k\ell)]. \end{aligned}$$

■

12.3 Particular Cases

This section presents a few classical examples of contrast functions as particular examples of $D_f(\cdot\,||\cdot\,)$. These are constructed by choosing several examples of functions f that satisfy conditions (a)–(c) and verify the equation $\alpha = 2f'''(1) + 3$. We make the remark that if f is such a function, then $f_c(u) = f(u) + c(u-1)$, $c \in \mathbb{R}$, is also a function that induces the same contrast function, $D_{f_c} = D_f$. Therefore, the correspondence between functions f and contrast functions is not one-to-one.

12.3.1 Hellinger Distance

Consider $f(u) = 4(1-\sqrt{u})$ and the associated contrast function

$$\begin{aligned} D_f(p||q) &= 4\int_{\mathcal{X}} p(x)\Big(1-\sqrt{\frac{q(x)}{p(x)}}\Big)dx = 4\Big(1-\int_{\mathcal{X}}\sqrt{p(x)q(x)}\,dx\Big) \\ &= 2\Big(2-\int_{\mathcal{X}} 2\sqrt{p(x)q(x)}\,dx\Big) \\ &= 2\int_{\mathcal{X}}\Big(p(x)-2\sqrt{p(x)q(x)}+q(x)\Big)\,dx \\ &= 2\int_{\mathcal{X}}\big(\sqrt{p(x)}-\sqrt{q(x)}\big)^2\,dx \\ &= H^2(p,q). \end{aligned}$$

$H(p,q)$ is called the *Hellinger distance*, and is a true distance on the statistical model $\mathcal{S} = \{p_\xi\}$. Since in this case $\alpha = 2f'''(1) + 3 = 0$, the linear connection induced by $H^2(p,q)$ is exactly the Levi–Civita connection, $\nabla^{(0)}$, on the Riemannian manifold $(\mathcal{S}, g)$.

Example 12.3.1 Consider two exponential distributions, $p(x) = \alpha e^{-\alpha x}$ and $q(x) = \beta e^{-\beta x}$, $x \geq 0$, $\alpha, \beta > 0$. Then

$$\begin{aligned} H^2(p,q) &= 4 - 4\int_0^\infty \sqrt{p(x)q(x)}\,dx \\ &= 4 - 4\sqrt{\alpha\beta}\int_0^\infty e^{-\frac{\alpha+\beta}{2}x}\,dx \\ &= 4 - \frac{8\sqrt{\alpha\beta}}{\alpha+\beta}, \end{aligned}$$

hence the Hellinger distance is $H(p,q) = 2\sqrt{1-\dfrac{2\sqrt{\alpha\beta}}{\alpha+\beta}}$.

The Hellinger distance can also be defined between two discrete distributions $p = (p_k)$ and $q = (q_k)$, replacing the integral by a sum

$$H(p,q) = 2\Big(1-\sum_{k\geq 0}\sqrt{p_k q_k}\Big)^{1/2} = \Big(2\sum_{k\geq 0}\big(\sqrt{p_k}-\sqrt{q_k}\big)^2\Big)^{1/2}.$$

Example 12.3.2 Consider two Poisson distributions, $p_k = \dfrac{\alpha^k}{k!}e^{-\alpha}$ and $q_k = \dfrac{\beta^k}{k!}e^{-\beta}$, $k \geq 0$. Then

$$\begin{aligned}\sum_{k\geq 0}\sqrt{p_k q_k} &= \sum_{k\geq 0}\frac{(\sqrt{\alpha\beta})^k}{k!}e^{-\frac{\alpha+\beta}{2}}\\ &= e^{-\frac{\alpha+\beta}{2}}e^{\sqrt{\alpha\beta}}\sum_{k\geq 0}\frac{(\sqrt{\alpha\beta})^k}{k!}e^{-\sqrt{\alpha\beta}}\\ &= e^{\sqrt{\alpha\beta}-\frac{\alpha+\beta}{2}}.\end{aligned}$$

Hence, the Hellinger distance becomes

$$H(p,q) = 2\Big(1 - \sum_{k\geq 0}\sqrt{p_k q_k}\Big)^{1/2} = 2\sqrt{1 - e^{\sqrt{\alpha\beta}-\frac{\alpha+\beta}{2}}}.$$

12.3.2 Kullback–Leibler Relative Entropy

The contrast function associated with function $f(u) = -\ln u$ is given by

$$D_f(p||q) \;\; = \;\; \int_{\mathcal{X}} p(x)\ln\frac{p(x)}{q(x)}\,dx = D_{KL}(p||q),$$

which is the *Kullback–Leibler information* or the *relative entropy*. In this case $\alpha = 2f'''(1)+3 = -1$, so the associated connection is $\nabla^{(-1)}$.

It is worthy to note that the convex function $f(u) = u\ln u$ induces the contrast function

$$D_f(p||q) \;\; = \;\; \int_{\mathcal{X}} q(x)\ln\frac{q(x)}{p(x)}\,dx = D_{KL}(q||p) = D^*_{KL}(p||q),$$

which is the dual of the *Kullback–Leibler information*, see [51, 53]. Since $\alpha = 2f'''(1) + 3 = 1$, the induced connection is $\nabla^{(1)}$.

12.3.3 Chernoff Information of Order α

The convex function

$$f^{(\alpha)} = \frac{1}{1-\alpha^2}(1 - u^{\frac{1+\alpha}{2}}), \qquad \alpha \neq \pm 1$$

induces the contrast function

$$D^{(\alpha)}(p||q) = \frac{4}{1-\alpha^2}\Big\{1 - \int_{\mathcal{X}} p(x)^{\frac{1-\alpha}{2}} q(x)^{\frac{1+\alpha}{2}}\, dx\Big\},$$

see Chernoff [27]. For the computation of $D^{(\alpha)}$ in the case of exponential, normal and Poisson distributions, see Problems 12.9., 12.10. and 12.11. We note that for $\alpha = 0$ we retrieve the squared Hellinger distance, $D^{(0)}(p||q) = H^2(p,q)$.

12.3.4 Jeffrey Distance

The function $f(u) = \frac{1}{2}(u-1)\ln u$ induces the contrast function

$$\begin{aligned} J(p,q) &= D_f(p||q) = \frac{1}{2}\int_{\mathcal{X}} p(x)\Big(1 - \frac{q(x)}{p(x)}\Big)\ln\frac{p(x)}{q(x)}\, dx \\ &= \frac{1}{2}\int_{\mathcal{X}} \big(p(x) - q(x)\big)\big(\ln p(x) - \ln q(x)\big)\, dx, \end{aligned}$$

see Jeffrey [47]. A computation shows that $\alpha = 0$, so the induced connection is the Levi–Civita connection $\nabla^{(0)}$. In fact, the Jeffrey contrast function is the same as the symmetric Kullback–Leibler relative entropy

$$\begin{aligned} J(p,q) &= \frac{1}{2}\int_{\mathcal{X}} p(x)\Big(1 - \frac{q(x)}{p(x)}\Big)\ln\frac{p(x)}{q(x)}\, dx \\ &= \frac{1}{2}\int_{\mathcal{X}} p(x)\ln\frac{p(x)}{q(x)} + \frac{1}{2}\int_{\mathcal{X}} q(x)\ln\frac{q(x)}{p(x)}\, dx \\ &= \frac{1}{2}\Big(D_{KL}(p||q) + D_{KL}(q||p)\Big). \end{aligned}$$

12.3.5 Kagan Divergence

Choosing $f(u) = \frac{1}{2}(1-u)^2$ yields

$$\begin{aligned} D_{\chi^2}(p||q) &= D_f(p||q) = \frac{1}{2}\int_{\mathcal{X}} p(x)\Big(1 - \frac{q(x)}{p(x)}\Big)\, dx \\ &= \frac{1}{2}\int_{\mathcal{X}} \frac{(p(x) - q(x))^2}{q(x)}\, dx, \end{aligned}$$

called the *Kagan contrast function*, see Kagan [48]. In this case $\alpha = 2f'''(1) + 3 = 3$, and therefore the induced connection is $\nabla^{(3)}$. It is worth noting the relation with the minimum chi-squared estimation in the discrete case, see Kass and Vos [49], p.243. In this case the Kagan divergence becomes

$$D_{\chi^2}(p,q) = \frac{1}{2}\sum_{i=1}^{n} \frac{(p_i - q_i)^2}{q_i}.$$

12.3.6 Exponential Contrast Function

The contrast function associated with the convex function $f(u) = \frac{1}{2}(\ln u)^2$ is

$$\mathcal{E}(p||q) = D_f(p||q) = \frac{1}{2}\int_{\mathcal{X}} p(x)\big(\ln p(x) - \ln q(x)\big)^2\,dx.$$

The induced connection in this case is $\nabla^{(-3)}$.

We note that all function candidates of the form $f(u) = K(\ln u)^{2k}$ are convex, but the condition $f''(1) = 1$ is verified only for $k = 1, 2$ (with appropriate constants K).

12.3.7 Product Contrast Function with (α, β)-Index

The following 2-parameter family of contrast functions is introduced and studied in Eguchi [40]

$$D_{\alpha,\beta}(p||q) = \frac{2}{(1-\alpha)(1-\beta)}\int \Big\{1 - \Big(\frac{p(x)}{q(x)}\Big)^{\frac{1-\alpha}{2}}\Big\}\Big\{1 - \Big(\frac{p(x)}{q(x)}\Big)^{\frac{1-\beta}{2}}\Big\}\,dx,$$

and is induced by the function

$$f_{\alpha,\beta}(u) = \frac{2}{(1-\alpha)(1-\beta)}(1 - u^{\frac{1-\alpha}{2}})(1 - u^{\frac{1-\beta}{2}}).$$

This connects to the previous contrast functions, see Problem 12.3.

It is worthy to note that the contrast function $D_{\alpha,\beta}(\cdot\,||\,\cdot)$ can be written as the following convex combination of Chernoff informations, see Problem 12.3, part (e).

We end this section with a few suggestive examples. The computations are left as exercises to the reader.

Example 12.3.1 Consider the statistical model $\mathcal{S} = \{p_\mu; \mu \in \mathbb{R}^k\}$, where

$$p_\mu(x) = (2\pi)^{-k/2} e^{-\frac{\|x-\mu\|^2}{2}}, \qquad x \in \mathbb{R}^k$$

is a k-dimensional Gaussian density with $\sigma = 1$. Problem 12.4 provides exact formulas for the aforementioned contrast functions in terms of the Euclidean distance $\|\cdot\|$.

Example 12.3.2 (Exponential Model) Let $\mathcal{S} = \{p_\xi\}$, where

$$p_\xi = \xi e^{-\xi x}, \qquad \xi > 0, x > 0.$$

A computation shows

$$\begin{aligned}
D_{KL}(p_\xi||p_{\xi'}) &= \frac{\xi'}{\xi} - \ln\frac{\xi'}{\xi} - 1 \\
J(p_\xi, p_{\xi'}) &= \frac{(\xi' - \xi)^2}{2\xi\xi'} \\
H^2(p_\xi, p_{\xi'}) &= \frac{4(\sqrt{\xi} - \sqrt{\xi'})^2}{\xi + \xi'} \\
D^{(\alpha)}(p_\xi||p_{\xi'}) &= \frac{4}{1-\alpha^2}\left\{1 - \frac{\xi^{\frac{1-\alpha}{2}}\xi'^{\frac{1+\alpha}{2}}}{\frac{1+\alpha}{2}\xi' + \frac{1-\alpha}{2}\xi}\right\} \\
D_{\chi^2}(p_\xi||p_{\xi'}) &= \frac{1}{2}\left[\frac{1}{\left(2 - \frac{\xi}{\xi'}\right)\frac{\xi}{\xi'}} - 1\right] \\
\mathcal{E}(p_\xi||p_{\xi'}) &= \frac{1}{2}\left\{\frac{\xi'}{\xi} - \ln\frac{\xi'}{\xi} - 1\right\}.
\end{aligned}$$

It is worthy to note that all these contrast functions provide the same Riemannian metric on $\mathcal{S}$ given by $g_{11} = \frac{1}{\xi^2}$, which is the Fisher information. The induced distance between p_ξ and $p_{\xi'}$ is a hyperbolic distance, i.e., $dist(p_\xi, p_{\xi'}) = |\ln\frac{\xi}{\xi'}|$.

12.4 Problems

12.1. Consider the exponential family

$$p(x;\xi) = e^{C(x) + \xi^i F_i(x) - \psi(\xi)}, \quad i = 1, \cdots, n,$$

with $\psi(\xi)$ convex function, and define

$$D(\xi_0||\xi) = \psi(\xi) - \psi(\xi_0) - \langle \partial\psi(\xi_0), \xi - \xi_0\rangle.$$

(a) Prove that $D(\cdot||\cdot)$ is a contrast function;

(b) Find the dual contrast function $D^*(\cdot||\cdot)$;

(c) Prove that the Riemann metric induced by the contrast function $D(\cdot||\cdot)$ is the Fisher–Riemann metric of the exponential family. Find a formula for it using the function $\psi(\xi)$;

(d) Find the components of the dual connections $\nabla^{(D)}$ and $\nabla^{(D^*)}$ induced by the contrast function $D(\cdot||\cdot)$;

(e) Show that the skewness tensor induced by the contrast function $D(\cdot||\cdot)$ is $T_{ijk}(\xi) = \partial_i \partial_j \partial_k \psi(\xi)$.

12.2. Prove that the Hellinger distance

$$H(p,q) = \sqrt{2\int_{\mathcal{X}} (\sqrt{p(x)} - \sqrt{q(x)})^2 \, dx}$$

satisfies the distance axioms.

12.3. Consider the Eguchi contrast function

$$D_{\alpha,\beta}(p||q) = \frac{2}{(1-\alpha)(1-\beta)} \int \Big\{1 - \Big(\frac{p}{q}\Big)^{\frac{1-\alpha}{2}}\Big\}\Big\{1 - \Big(\frac{p}{q}\Big)^{\frac{1-\beta}{2}}\Big\} \, dx.$$

Let $H(\cdot,\cdot)$, $D^{(\alpha)}(\cdot||\cdot)$, $J(\cdot,\cdot)$, $\mathcal{E}(\cdot||\cdot)$ be the Hellinger distance, the Chernoff information of order α, the Jefferey distance, and the exponential contrast function, respectively. Prove the following relations:

(a) $D_{0,0}(p||q) = H^2(p,q)$

(b) $D_{-\alpha,\alpha}(p||q) = \frac{1}{2}\big(D^{(\alpha)}(p||q) + D^{(-\alpha)}(p||q)\big)$

(c) $\lim\limits_{\alpha\to 1} D_{-\alpha,\alpha}(p||q) = J(p,q)$

(d) $\lim\limits_{\alpha\to -1} D_{\alpha,\alpha}(p||q) = \mathcal{E}(p||q)$

(e) $D_{\alpha,\beta}(p||q) = \lambda_1 D^{(-\alpha)} + \lambda_2 D^{(-\beta)} + \lambda_3 D^{(\frac{1-\alpha-\beta}{2})}$,

where

$$\lambda_1 = \frac{1+\alpha}{2(1-\beta)}, \quad \lambda_2 = \frac{1+\beta}{2(1-\alpha)}, \quad \lambda_3 = -\frac{(\alpha+\beta)(2-\alpha-\beta)}{2(1-\alpha)(1-\beta)},$$

and show that $\lambda_1 + \lambda_2 + \lambda_3 = 1$.

12.4. Consider the statistical model defined by the k-dimensional Gaussian family, $\mathcal{S} = \{p_\mu; \mu \in \mathbb{R}^k\}$,

$$p_\mu(x) = (2\pi)^{-k/2} e^{-\frac{\|x-\mu\|^2}{2}}, \qquad x \in \mathbb{R}^k.$$

Prove the following relations:

$$(a)\quad D_{KL}(p_\mu||p_{\mu'}) = \frac{1}{2}\|\mu-\mu'\|^2$$

$$(b)\quad J(p_\mu, p_{\mu'}) = \frac{1}{2}\|\mu-\mu'\|^2$$

$$(c)\quad H^2(p_\mu, p_{\mu'}) = 4\Big[1 - e^{-\frac{\|\mu-\mu'\|^2}{8}}\Big]$$

$$(d)\quad D^{(\alpha)}(p_\mu||p_{\mu'}) = \frac{4}{1-\alpha^2}\Big[1 - e^{-\frac{1-\alpha^2}{8}\|\mu-\mu'\|^2}\Big]$$

$$(e)\quad \mathcal{E}(p_\mu||p_{\mu'}) = \frac{1}{2}\|\mu-\mu'\|^2\Big[1+\frac{1}{4}\|\mu-\mu'\|^2\Big],$$

where $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^k$.

12.5. Let $D_f(\cdot||\cdot)$ be the f-divergence. Prove the following convexity property

$$D_f\Big(\lambda p_1+(1-\lambda)p_2||\lambda q_1+(1-\lambda)q_2\Big) \;\le\; \lambda D_f(p_1||q_1) + (1-\lambda)D_f(p_2||q_2),$$

$\forall\lambda \in [0,1]$ and p_1, p_2, q_1, q_2 distribution functions.

12.6. Prove the formulas for the contrast function in the case of the exponential distribution presented by Example 12.3.2.

12.7. Consider the normal distributions $p(x) = \frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}$ and $q(x) = \frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}$.

(a) Show that

$$\int_{-\infty}^{\infty}\sqrt{p(x)q(x)}\,dx = \sqrt{\frac{2\sigma_1\sigma_2}{\sigma_1^2+\sigma_2^2}e^{A-B}},$$

where

$$A = \frac{\Big(\frac{\mu_1}{2\sigma_1^2}+\frac{\mu_2}{2\sigma_2^2}\Big)^2}{\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}}, \quad B = \frac{\mu_1^2}{4\sigma_1^2}+\frac{\mu_2^2}{4\sigma_2^2}.$$

(b) Find the Hellinger distance $H(p,q)$.

12.8. Find the Hellinger distance between two gamma distributions.

12.9. Consider two exponential distributions, $p(x) = ae^{-ax}$ and $q(x) = be^{-bx}$, $x \geq 0$. Show that the Chernoff information of order α is

$$D^{\alpha}(p||q) = \frac{4}{1-\alpha^2}\left\{1 - \frac{2a^{\frac{1-\alpha}{2}}b^{\frac{1+\alpha}{2}}}{a(1-\alpha)+b(1+\alpha)}\right\}, \quad \alpha \neq \pm 1.$$

12.10. Consider the normal distributions $p(x) = \frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}$ and $q(x) = \frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}$. Show that the Chernoff information of order α is

$$D^{\alpha}(p||q) = \frac{4}{1-\alpha^2}\left\{1 - A\sqrt{\frac{\pi}{a}}e^{\frac{b^2}{4a}-c}\right\}, \qquad |\alpha| < 1,$$

where

$$a = \frac{1-\alpha}{4\sigma_1^2} + \frac{1+\alpha}{4\sigma_2^2}$$
$$b = \frac{\mu_1(1-\alpha)}{2\sigma_1^2} + \frac{\mu_2(1+\alpha)}{2\sigma_2^2}$$
$$c = \frac{\mu_1^2(1-\alpha)}{4\sigma_1^2} + \frac{\mu_2^2(1+\alpha)}{4\sigma_2^2}.$$

12.11. The Chernoff information of order α for discrete distributions (p_n) and (q_n) is given by

$$D^{(\alpha)}(p||q) = \frac{4}{1-\alpha^2}\left\{1 - \sum_{n\geq 0} p_n^{\frac{1-\alpha}{2}} q_n^{\frac{1+\alpha}{2}}\right\}.$$

Let $p_n = \frac{\lambda_1^n}{n!}e^{-\lambda_1}$ and $q_n = \frac{\lambda_2^n}{n!}e^{-\lambda_2}$ be two Poisson distributions.

(a) Show that

$$D^{(\alpha)}(p||q) = \frac{4}{1-\alpha^2}\left\{1 - e^{\lambda_1^{(1-\alpha)/2}\lambda_2^{(1+\alpha)/2} - \lambda_1(1-\alpha)/2 - \lambda_2(1+\alpha)/2}\right\}.$$

(b) Show that the square of the Hellinger distance is given by

$$H^2(p,q) = 4\{1 - e^{\sqrt{\lambda_1\lambda_2} - \frac{\lambda_1+\lambda_2}{2}}\}.$$