Appendix A

Calculations

Distribution comparing measures

Overview. At this point, distribution comparing measures shall be reviewed and characterized. Depending on the syntax used within the community in question, these measures are designed to determine either the *proximity*, *similarity*, *distance* or *divergence* of the two considered probability distributions f_1 and f_2.

Following the characterization (c.f. [143, 171]) they can be grouped into measures related to the KULLBACK–LEIBLER distance, as a representative of the ALI-SILVEY class (c.f. [9, 145, 146]), which is also known as *relative entrophy*,

$$\mathcal{D}_{\mathrm{KL}}(f_1||f_2) = \int f_2(x) \log \frac{f_2(x)}{f_1(x)} \, \mathrm{d}x, \tag{A.1}$$

and into ones related to the BHATTACHARYYA measure [30]

$$\mathcal{D}_{\mathrm{B}}(f_1||f_2) = \int \sqrt{f_1(x)} \sqrt{f_2(x)} \, \mathrm{d}x. \tag{A.2}$$

Bhattacharyya related and beyond. At least, two other measures should mentioned that are closely related to the BHATTACHARYYA one. The first one is the MATUSITA metric (c.f. [169])

$$\mathcal{D}_{\mathrm{M}}(f_1||f_2) = \int \left(\sqrt{f_1(x)} - \sqrt{f_2(x)} \right)^2 \mathrm{d}x, \tag{A.3}$$

which is related to the (A.2) by

$$\begin{aligned} \int \left(\sqrt{f_1(x)} - \sqrt{f_2(x)} \right)^2 \mathrm{d}x &= 2 - 2 \int \sqrt{f_1(x)} \sqrt{f_2(x)} \, \mathrm{d}x \\ \mathcal{D}_{\mathrm{M}}(f_1||f_2) &= 2 - 2\mathcal{D}_{\mathrm{B}}(f_1||f_2) \end{aligned}$$

and the second one is the HELLINGER distance

$$\mathcal{D}_{\mathrm{H}}(f_1||f_2) = \sqrt{\int \left(\sqrt{f_1(x)} - \sqrt{f_2(x)}\right)^2 \mathrm{d}x}, \tag{A.4}$$

which is related to the (A.2) and (A.3) by

$$\mathcal{D}_{\mathrm{H}}(f_1||f_2) = \sqrt{\mathcal{D}_{\mathrm{M}}(f_1||f_2)} = \sqrt{2 - 2\mathcal{D}_{\mathrm{B}}(f_1||f_2)}.$$

The BHATTACHARYYA measure can be interpreted as a special case of the CHERNOFF distance (c.f. [56])

$$\mathcal{D}_{\mathrm{C}}(f_1||f_2) = \max_{0\leq t\leq 1} \left\{-\log\mu(t)\right\}, \quad \mu(t) = \int \left[f_1(x)\right]^{1-t} \left[f_2(x)\right]^t \mathrm{d}x, \tag{A.5}$$

which coincides for $t = 1/2$ with the BHATTACHARYYA measure.

Properties. Both families of measure, the (A.1) as well as (A.5) are additive in the sense that the measure of two joint distribution of statistically independent, identically distributed random variables can be represented as the sum of the marginal measures. If the random variables are not identically distributed, the CHERNOFF measure in (A.5) for $t \neq 1/2$ is not additive, whereas the KULLBACK–LEIBLER and BHATTACHARYYA remain additive (c.f. [129]).

When fitting a set of data to a distribution from a family of distributions, the one with the minimal KULLBACK–LEIBLER distance is the maximum likelihood estimation, assuming that the *true* distribution exists among the family. For comparing two normal distribution, the $t-$test as well as KULLBACK–LEIBLER distance are equal. Further, the $\chi^2 - -$function is the first term of the Taylor expension of the KULLBACK–LEIBLER distance (c.f. [64]).

Applications. The BHATTACHARYYA measure and its relatives are successfully been used for example in quantum mechanics (c.f. [219]), in objection recognition and feature tracking (c.f. [147, 229]) as well as in automatic self generation in neural networks (c.f. [228]). There, model selection is performed by comparing data and model prediction. In the latter case, Poisson distributions, representing frequency distributed data, are compared with each other (c.f. [3]). Since Poisson distributions have equal mean and variance, it could be shown that the BHATTACHARYYA measure compensate the bias of the $\chi^2 - -$statistics. However, for arbitary distributions, the BHATTACHARYYA measure does not eliminate the bias within the $\chi^2 - -$model–fitting.

Bhattacharyya for Gaussians. The BHATTACHARYYA measure for two Gaussian distributions f_1 and f_2 with $f_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $f_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$

$$f_1(x) = \frac{1}{2\pi\sigma_1^2} \exp\left\{-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right\} \tag{A.6}$$

$$f_2(x) = \frac{1}{2\pi\sigma_2^2} \exp\left\{-\frac{(x-\mu_2)^2}{2\sigma_2^2}\right\} \tag{A.7}$$

The integral

$$\int_{-\infty}^{\infty} (f_1(x))^n \ (f_2(x))^n \ \mathrm{d}x = \frac{\sqrt{2\pi\sigma_1^2\sigma_2^2/n}}{(2\pi\sigma_1^2\sigma_2^2)^n \sqrt{\sigma_1^2+\sigma_2^2}} \exp\left\{-\frac{n\,(\mu_1-\mu_2)^2}{2\,(\sigma_1^2+\sigma_2^2)}\right\} \tag{A.8}$$

For $n = 1/2$, this results in

$$\int_{-\infty}^{\infty} \sqrt{f_1(x)\ f_2(x)}\ \mathrm{d}x = \frac{\sqrt{2\sigma_1\sigma_2}}{\sqrt{\sigma_1^2+\sigma_2^2}} \exp\left\{-\frac{(\mu_1-\mu_2)^2}{4\,(\sigma_1^2+\sigma_2^2)}\right\}. \tag{A.9}$$

Linear overlap

In the following section, the linear overlap functional $\mathcal{F}_{\mathcal{L}}$ shall be derived. Let therefore the parameter distribution be normally distributed as well as the imposed model variability distribution, as the parameter and model variability distributions are linear transforms of each other. In order to interpret the overlap according to (4.1), one has to enforce $\|\mathcal{M}\|_2 = \|\mathcal{D}\|_2 = 1$.

For that reason the normalization factors have to be calculated. For the reason, the integral of two normally distributed densities f_1 and f_2 are calculated. With the notation of (A.6) and (A.7) as well as the integration of (A.8), one gets

$$\langle f_1, f_2 \rangle = \frac{1}{\sqrt{2\pi\,(\sigma_1^2+\sigma_2^2)}} \ \exp\left\{-\frac{(\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right\}. \tag{A.10}$$

Therefore the normalization factor for the variabilities has to be chosen according to

$$\langle f, f \rangle = \frac{1}{2\sigma\sqrt{\pi}} \tag{A.11}$$

resulting in the data and model variabilities

$$\mathcal{D} = \frac{1}{\sqrt[4]{\pi}\,\sqrt{\sigma_D}}\ e^{-\frac{(x-\mu_D)^2}{2\sigma_D^2}} \tag{A.12}$$

$$\mathcal{M} = \frac{1}{\sqrt[4]{\pi}\,\sqrt{\sigma_M}}\ e^{-\frac{(x-\mu_M)^2}{2\sigma_M^2}}. \tag{A.13}$$

by dividing the densities (A.6) and (A.7) through the corresponding square root of (A.11). Even more, the dividing (A.10) through the factors results in the linear overlap functional

$$\langle \mathcal{D}, \mathcal{M} \rangle_2 = \sqrt{\frac{2\,\sigma_D\,\sigma_M}{\sigma_D^2 + \sigma_M^2}}\; e^{-\frac{(\mu_D-\mu_M)^2}{2(\sigma_D^2+\sigma_M^2)}}. \tag{A.14}$$

In comparison to the BHATTACHARYYA measure in (A.9), the mean deviation is divided only by factor 2 instead of 4. Therefore a larger mean deviation is "punished" more in the overlap setting.

Maximal Overlap

Equation (A.14) can be used to derive a necessary condition for a maximal overlap in the linear case of section 4.4 for a constant mean deviation and constant data variability. Therefore, the overlap is considered to be a function of the variance of the data σ_M only. By choosing

$$u(\sigma_M) = \sqrt{\frac{2\,\sigma_D\,\sigma_M}{\sigma_D^2 + \sigma_M^2}} \quad \text{and} \quad v(\sigma_M) = e^{-\frac{(\mu_D-\mu_M)^2}{2(\sigma_D^2+\sigma_M^2)}}$$

and applying the product rule, considering

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\,\sigma_M} u(\sigma_M) &= \frac{1}{u(\sigma_M)} \frac{\sigma_D(\sigma_D^2 - \sigma_M^2)}{(\sigma_D^2 + \sigma_M^2)^2} \quad \text{and} \\ \frac{\mathrm{d}}{\mathrm{d}\,\sigma_M} v(\sigma_M) &= \sigma_M\, \frac{(\mu_D - \mu_M)^2}{(\sigma_D^2 + \sigma_M^2)^2}\; v(\sigma_M), \end{aligned}$$

the first derivative of the overlap

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\,\sigma_M} \langle \mathcal{D}, \mathcal{M} \rangle &= \frac{\mathrm{d}}{\mathrm{d}\,\sigma_M} u(\sigma_M)\, v(\sigma_M) \\ &= \frac{u(\sigma_M)\, v(\sigma_M)}{2\sigma_M(\sigma_D^2 + \sigma_D^2)} \left(\sigma_M^4 - 2(\mu_D - \mu_M)^2 \sigma_M^2 - \sigma_D^4\right) \\ &= e^{-\frac{(\mu_D-\mu_M)^2}{2(\sigma_D^2+\sigma_M^2)}} \sqrt{\frac{\sigma_D}{2\sigma_M\left(\sigma_D^2 + \sigma_M\right)^3}}\; \left(\sigma_M^4 - 2(\mu_D - \mu_M)^2 \sigma_M^2 - \sigma_D^4\right). \end{aligned}$$

Since the functions u and v as well as σ_M are supposed to be positive, the necessary condition for σ_M maximizing the overlap is

$$\sigma_M = \sqrt{(\mu_D - \mu_M)^2 + \sqrt{(\mu_D - \mu_M)^4 + \sigma_D^4}} \;\geq\; \sigma_D. \tag{A.15}$$

Since the overlap is continuous and vanishing for values close to zero and infinity, equation (4.11) is also sufficient.