X MATHEMATICAL SUPPLEMENT

X.1 *Relative information entropies*

The Shannon-entropy S (or, equivalently, the Gibbs-entropy as it is called in the statistical physics community rather than in the information theory community)

$$S = -\sum_i p_i \ln p_i \tag{X.553}$$

can be extended to measuring the relative entropy between two discrete distributions p_i and q_i, yielding the Kullback-Leibler divergence ΔS,

$$\Delta S = -\sum_i p_i \ln\left(\frac{p_i}{q_i}\right) = \left\langle \ln\left(\frac{p_i}{q_i}\right)\right\rangle \tag{X.554}$$

which really plays its strength when generalised to continuous distributions $p(x)\mathrm{d}x$ and $q(x)\mathrm{d}x$,

$$\Delta S = -\int \mathrm{d}x\, p(x) \ln\left(\frac{p(x)}{q(x)}\right) = -\left\langle \ln\left(\frac{p(x)}{q(x)}\right)\right\rangle \tag{X.555}$$

The relative entropy comes with a large advantage as it is invariant under transformations of the random variable in the continuous case (the problem does not arise in the discrete case, anyways). The transformation law is commonly written as $p(x)\mathrm{d}x = p(z)\mathrm{d}z$ and results from integration by substitution:

$$\int \mathrm{d}x\, p(x) = \int \mathrm{d}z\, p(x(z))\frac{\mathrm{d}x}{\mathrm{d}z} \tag{X.556}$$

with a transformation Jacobian $J = \mathrm{d}x/\mathrm{d}z$. In contrast to the straightforward entropy $S = -\int \mathrm{d}x\, p(x) \ln p(x)$ which transforms to $-\int \mathrm{d}z\, p(z)[\ln(p(z)) + \ln J]$ and picks up an additional term depending on the transformation, this additional term cancels in the ratio $p(z)/q(z) = p(x)/q(x)$ of the relative entropy. That effectively means, that the continuum limit of the Shannon-entropy can not be defined in an invariant way:

$$S = -\int \mathrm{d}x\, p(x) \ln p(x) = -\langle \ln p(x)\rangle . \tag{X.557}$$

Related entropy measures, that are likewise (i) positive definite and bounded by 0, (ii) additive for independent random processes and (iii) growing with the number of possible outcomes are Rényi-entropies S_α

$$S_\alpha = -\frac{1}{\alpha-1} \ln \int \mathrm{d}x\, p(x)p(x)^{\alpha-1} = -\frac{1}{\alpha-1} \ln \left\langle p(x)^{\alpha-1}\right\rangle \tag{X.558}$$

for any constant $0 < \alpha \neq 1$. There are corresponding definitions of relative entropies ΔS_α

$$\Delta S_\alpha = -\frac{1}{\alpha-1} \ln \int \mathrm{d}x\, p(x)\left(\frac{p(x)}{q(x)}\right)^{\alpha-1} = -\frac{1}{\alpha-1} \ln \left\langle \left(\frac{p(x)}{q(x)}\right)^{\alpha-1}\right\rangle \tag{X.559}$$

One often runs into problems with Rényi-entropies when dealing with conditional and joint probabilities, which miraculously works with Kullback-Leibler divergences: Joint probabilities $p(x,z)$ can be generated in in a two-step random process as

$$p(x,z) = p(x|z)p(z) = p(z|x)p(x) \tag{X.560}$$

with conditional probabilities, which are obviously connected through Bayes' law. The conditional entropy $\mathrm{S}(z|x)$ of $p(x,z)$ relative to $p(x)$ is given by

$$\begin{aligned}
\mathrm{S}(z|x) = -\int \mathrm{d}x \int \mathrm{d}z\; p(x,z) \ln\left(\frac{p(x,z)}{p(x)}\right) = \\
-\int \mathrm{d}x \int \mathrm{d}z\; p(x,z) \ln p(x,z) + \int \mathrm{d}x \int \mathrm{d}z\; p(x,z) \ln p(x) = \\
-\int \mathrm{d}x \int \mathrm{d}z\; p(x,z) \ln p(x,z) + \int \mathrm{d}x\; \ln p(x) \int \mathrm{d}z\; p(x,z) = \\
-\int \mathrm{d}x \int \mathrm{d}z\; p(x,z) \ln p(x,z) + \int \mathrm{d}x\; p(x) \ln p(x) = \mathrm{S}(x,z) - \mathrm{S}(x)
\end{aligned} \tag{X.561}$$

because of the marginalisation $\int \mathrm{d}z\; p(x,z) = p(x)$ in the second term, such that we can write down the entropy-version of Bayes' law, making use of the symmetry of $\mathrm{S}(x,z)$:

$$\mathrm{S}(z|x) + \mathrm{S}(x) = \mathrm{S}(z,x) = \mathrm{S}(x,z) = \mathrm{S}(x|z) + \mathrm{S}(z) \tag{X.562}$$

which is impossible to formulate in terms of Rényi-entropies due to the logarithm acting on an integral.