14 The Data Processing Theorem

The Data Processing Theorem compares the Kullback-Leibler distance (relative entropy) between the probability distributions of two inputs to a system to the distance between the corresponding output distributions. The Data Processing Theorem says the ratio of the output distance to the input distance must be less than or equal to one. Suppose X_n is a sequence of statistically independent Gaussian random variables having some mean and variance σ_X^2. The "system" simply adds N_n, white Gaussian noise statistically independent of the input having zero mean and variance σ_N^2, to produce the output: $Y_n = X_n + N_n$. Here we explore the Data Processing Theorem when the input distributions differ only in their means.

(a) If X_n has mean m_0 or m_1, what is the ratio of the output to the input Kullback-Leibler distances?

(b) The Data Processing Theorem does not suggest whether feedback can increase or decrease the ratio of distances. Suppose

$$Y_n = aY_{n-1} + X_n + N_n \ ,$$

indicating the previous output modifies the value of each current output. For Y_n to be stationary, the coefficient must satisfy $|a| < 1$. What is the ratio of Kullback-Leibler distances between the output and input values at some time index n_0? Can the ratio be greater than one?

(c) More generally, we can study $\mathbf{Y} = \mathbf{A}(\mathbf{X}+\mathbf{N})$, which represents a more general expression for feedback. Assume that $\mathbf{A}$ is an invertible matrix and that $\mathbf{X}$ and $\mathbf{N}$ are Gaussian random vectors, statistically independent of each other with covariance matrices $\sigma_X^2\mathbf{I}$ and $\sigma_N^2\mathbf{I}$, respectively. As before, $\mathbf{N}$ has zero mean and the mean of $\mathbf{X}$ assumes two possible values. What is the ratio of Kullback-Leibler distances between the output and input vectors? Comment on your answer in light of your result for part (b).