Continuity of mutual entropy in the large signal-to-noise ratio limit

Mark Kelbert[1)] **and Yuri Suhov**[2)]

Abstract. This note addresses the issue of the proof of the entropy power inequality (EPI), an important tool in the analysis of Gaussian channels of information transmission, proposed by Shannon. We analyse continuity properties of the mutual entropy of the input and output signals in an additive memoryless channel and show how this can be used for a correct proof of the entropy-power inequality.

1. Introduction

The impetus for composing this short note was provided by the (even shorter) note [VG] suggesting an elegant way of deriving the so-called entropy-power inequality (EPI) playing a crucial role in the analysis of channels with additive noise. The way of reasoning in [VG] is often referred to as a direct probabilistic method, as opposite to the so-called analytic method; see [R, L, CT]. For the history of the question, see [CT]; for reader's convenience, the statement of the EPI is given at the end of this section. Unfortunately, we were not able to find a rigorous foundation for a key statement, Eqn (5) from [VG], which was exacerbated by the fact that Lemma 7 from a longer paper [GSV], on which the proof of Eqn (5) from [VG] was based turned to be incorrect. Also, Lemma 6 from [GSV] is stated for discrete signals (for which it is holds true) but had been used in [GSV] for continuous signals (where it is incorrect). Another unsatisfactory aspect of papers [GSV] and [VG] was that a number of important auxiliary assertions in these papers were stated (and proved) in a manner leaving open the question under precisely what conditions they remain valid, on the input signal and the additive noise distributions.

Looking at these matters, we realised that a number of lemmas on the limiting behaviour of the mutual and conditional entropies emerge, where the signal-to-noise ratio tends to zero or $+\infty$. We collect these lemmas in Section 2. In Section 3 we reproduce a corrected way of reasoning where, by following the scheme proposed in [VG], one is able to establish the EPI rigorously.

We were inspired by an early paper [D] where a number of important (and elegant) results have been proven, about limiting behaviour of various entropies.

To introduce the entropy power inequality (EPI), consider two independent random variables (RVs) X_1 and X_2 taking values in $\mathbb{R}^d$, with probability density functions (PDFs) $f_{X_1}(x)$ and $f_{X_2}(x)$, respectively, where $x \in \mathbb{R}^d$. Let $h(X_i)$, $i = 1, 2$ stand for the differential entropies

$$h(X_i) = -\int_{\mathbb{R}^d} f_{X_i}(x) \ln\, f_{X_i}(x) \mathrm{d}x := -\mathbb{E} \ln\, f_{X_i}(X_i),$$

and assume that $-\infty < h(X_1), h(X_2) < +\infty$. The EPI states that

$$e^{\frac{2}{d}h(X_1+X_2)} \geq e^{\frac{2}{d}h(X_1)} + e^{\frac{2}{d}h(X_2)}, \tag{1}$$

(1) Mathematics Department, Swansea University, Singleton Park, Swansea, SA2 8PP, UK; International Institute of Earthquake Prediction Theory and Mathematical Geophysics, Profsoyuznaya str. 84/32, Moscow 117997, Russia

2) Statistical Laboratory, DPMMS, University of Cambridge, Wilberforce Road, Cambridge, CB 3 0WB; Institute for Information Transmission Problems, RAS, 19, Bol'shoi Karetnyi Per, Moscow 101447, Russia

or, equivalently,

$$h(X_1 + X_2) \geq h(Y_1 + Y_2), \tag{2}$$

where Y_1 and Y_2 are *any* independent normal RVs with $h(Y_1) = h(X_1)$ and $h(X_2) = h(Y_2)$. This inequality was first proposed by Shannon [SW]; as was mentioned earlier, it is used in the analysis of (memoryless) Gaussian channels of signal transmission. A rigorous proof of (1), (2) remains a subject of a growing amount of literature; see, e.g., references cited above. In particular, the question under what conditions upon PDFs f_{X_i} Eqns (1), (2) hold true remains largely open.

Moreover, we would like to note that the EPI remains true when one or both of RVs X_1, X_2 have atoms in their distributions, i.e., admit values with positive probabilities. In this case the corresponding differential entropies $h(X_1)$ and $h(X_2)$ are replaced with 'general' entropies:

$$\begin{aligned} h(X_i) &= -\sum_x p_{X_i}(x) \ln\, p_{X_i}(x) - \int f_{X_i}(x) \ln\, f_{X_i}(x) \mathrm{d}x \\ &= -\int g_{X_i}(x) \ln\, g_{X_i}(x) m\big(\mathrm{d}x\big) := -\mathbb{E} \ln\, g_{X_i}(X_i). \end{aligned}$$

Here $\sum\limits_x$ represents summation over a finite or countable set $\mathbb{D}(= \mathbb{D}(X_I))$ of points $x \in \mathbb{R}^d$. Further, given an RV X, $p_X(x)$ stands for the (positive) probability assigned: $p_X(x) = \mathbb{P}(X = x) > 0$, with the total sum $\eta(X) := \sum\limits_x p_X(x) \leq 1$. Next, f_X, as before, denotes the PDF for values forming an absolutely continuous part of the distribution of X (with $\displaystyle\int f_X(x)\mathrm{d}x = 1 - \eta(X)$, so when $\eta(X) = 1$, the RV X has a discrete distribution, and $h(X) = -\sum\limits_x p_X(x) \ln\, p_X(x)$). Further, $m(= m_X)$ is a reference measure (a linear combination of the counting measure on the discrete part and the Lebesgue measure on the absolutely continuous part of the distribution of X) and g_X the respective Radon-Nikodym derivative:

$$g(x) = p_X(x)\mathbf{1}(x \in \mathbb{D}) + f_X(x), \quad \text{with} \quad \int g_X(x) m(\mathrm{d}x) = 1.$$

We will refer to g_X as a probabiliy mass function (PMF) of RV X (with a slight abuse of traditional termonology). It is also possible to incorporate an (exotic) case where a RV X_i has a singular continuous component in its distribution, but we will not bother about this case in the present work. The scheme of proving the EPI for a discrete (or a mixed) case remains intact but continuity results look slightly different; see Section 2.

2. Continuity of the mutual entropy

Throughout the paper, all random variables are taking values in $\mathbb{R}^d$ (i.e., are d-dimensional real random vectors). If Y is such an RV then the notation $h(Y)$, $f_Y(x)$, $p_Y(x)$, $g_Y(x)$ and $m(\mathrm{d}x)$ have the same meaning as in Section 1 (it will be clear from the local context which particular form of the entropy $h(Y)$ we refer to).

Similarly, $f_{X,Y}(x, y)$ and, more generally, $g_{X,Y}(x, y)$, $x, y \in \mathbb{R}^d$, stand for the joint PDF and joint PMF of two RVs X, Y (relative to a suitable reference measure $m(\mathrm{d}x \times \mathrm{d}y)\big(= m_{X,Y}(\mathrm{d}x \times \mathrm{d}y)\big)$ on $\mathbb{R}^d \times \mathbb{R}^d$). Correspondingly, $h(X, Y)$ denotes the joint entropy of X and Y and $i(X : Y)$ their mutual entropy:

$$h(X, Y) = -\int g_{X,Y}(x, y) \ln\, g_{X,Y}(x, y) m\big(\mathrm{d}x \times \mathrm{d}y\big), \quad i(X : Y) = h(X) + h(Y) - h(X, Y).$$

We will use representations involving conditional entropies:

$$i(X : Y) = h(X) - h(X|Y) = h(Y) - h(Y|X),$$

where

$$h(X|Y) = h(X,Y) - h(Y), \; h(Y|X) = h(X,Y) - h(X).$$

In this section we deal with various entropy-continuity properties related to so-called additive channels where a RV X (a signal) is transformed into the sum $X+U$, with RV U representing 'noise' in the channel. In fact, we will adopt a slightly more general scheme where X is compared with $\gamma X + U$, $\gamma > 0$ being a parameter, and study limits where $\gamma \to +\infty$ or $\gamma \to 0+$. We will assume that RVs X and U are independent (though this assumption may be relaxed), and that the 'noise' U has a PDF $f_U(x)$ with $\displaystyle\int f_U(x)\mathrm{d}x = 1$. However, the signal X may have a general distribution including a discrete and an absolutely continuous part.

We begin with the analysis of behaviour of the mutual entropy $I(X : X\sqrt{\gamma}+U)$ when $\gamma \to +\infty$: this analysis will be used in Section 3, in the course of proving the EPI. First, we consider the case where X has a PDF $f_X(x)$ with $\displaystyle\int f_X(x)\mathrm{d}x = 1$.

Lemma 2.1. *Let X, U be independent RVs with PDFs f_X and f_U where $\displaystyle\int f_X(x)\mathrm{d}x = \int f_U(x)\mathrm{d}x = 1$. Suppose that $\displaystyle\int \big(f_X(x)|\ln\, f_X(x)| + f_U(x)|\ln\, f_U(x)|\big)\mathrm{d}x < +\infty$. Also assume that f_X is continuous and bounded: $f_X \in C^0(\mathbb{R}^d)$ and* $\sup\;[f_X(x):\; x \in \mathbb{R}] = b < +\infty$. *Then*

$$h(X) = \lim_{\gamma\to\infty} \big[I\big(X : X\sqrt{\gamma}+U\big) + h\big(U/\sqrt{\gamma}\big)\big]. \tag{5}$$

Proof of Lemma 2.1. Set: $Y := X\sqrt{\gamma} + U$. The problem is, obviously, equivalent to proving that $\big[h(X|Y) - h\big(U/\sqrt{\gamma}\big)\big] \to 0$. Writing $h\big(U/\sqrt{\gamma}\big) = -\ln\sqrt{\gamma} - \displaystyle\int f_U(u)\ln\, f_U(u)\mathrm{d}u$, we obtain

$$\begin{aligned}
&h(X|Y) - h(U/\sqrt{\gamma})\\
&\quad= -\int \mathrm{d}x\mathbf{1}\big(f_X(x)>0\big)f_X(x)\int f_U\big(y - x\sqrt{\gamma}\big)\ln\, \frac{f_X(x)f_U\big(y - x\sqrt{\gamma}\big)}{\displaystyle\int \mathrm{d}u f_X(u)f_U\big(y - u\sqrt{\gamma}\big)}\mathrm{d}y\\
&\qquad\quad + \ln\sqrt{\gamma} + \int\; f_U(u)\ln\; f_U(u)\mathrm{d}u\\
&\quad= \int \mathrm{d}x\mathbf{1}\big(f_X(x)>0\big)f_X(x)\int f_U(y)\ln\left[\frac{\sqrt{\gamma}\displaystyle\int \mathrm{d}u f_X(u)f_U\big(y+(x-u)\sqrt{\gamma}\big)}{f_X(x)}\right]\mathrm{d}y\\
&\quad= \int \mathrm{d}x\mathbf{1}\big(f_X(x)>0\big)f_X(x)\int f_U(y)\ln\left[\frac{\displaystyle\int f_X\left(x+\frac{y-v}{\sqrt{\gamma}}\right)f_U(v)\mathrm{d}v}{f_X(x)}\right]\mathrm{d}y.
\end{aligned} \tag{6}$$

Due to continuity and boudedness of f_X, the ratio under the logarithm converges to 1 as $\gamma \to +\infty$, $\forall$ $x, y \in \mathbb{R}$ (Lebesgue's dominated convergence theorem is helpful here). To finish the proof, we employ Lebesgue's dominated convergence theorem once more.

To this end, we write the logarithm as the difference of the logarithms of the numerator and the denominator. The logarithm of the numerator is assessed from above by b; this yields that

$$\ln\int f_X\left(x+\frac{y-v}{\sqrt{\gamma}}\right)f_U(v)\mathrm{d}v - \ln\; f_X(x) \le b + |\ln\; f_X(x)|.$$

To assess the logarithm from below, consider first a special (scalar) case where $X \sim U \sim$U$[0,1]$; here all integrals will be over $\mathbb{R}$. Take $\gamma > 4$. In this case we have that the integral of interest,

$$I = \int \mathrm{d}x \mathbf{1}\big(f_X(x) > 0\big) f_X(x) \int \mathrm{d}y f_U(y) \left| \ln \left[\frac{\int f_X\left(x + \frac{y-v}{\sqrt{\gamma}}\right) f_U(v)\mathrm{d}v}{f_X(x)} \right] \right|,$$

is equal to

$$\int_0^1 \mathrm{d}x \int_0^1 \mathrm{d}y \left| \ln \int_{(y+(x-1)\sqrt{\gamma})\vee 0}^{(y+x\sqrt{\gamma})\wedge 1} \mathrm{d}v \right| = I(0) + I(1)$$

where

$$I(0) = \left(\int_0^{1/\sqrt{\gamma}} \mathrm{d}x \int_{1-x\sqrt{\gamma}}^{1} \mathrm{d}y + \int_{1/\sqrt{\gamma}}^{1-1/\sqrt{\gamma}} \mathrm{d}x \int_0^1 \mathrm{d}y + \int_{1-1/\sqrt{\gamma}}^{1} \mathrm{d}x \int_0^{(1-x)\sqrt{\gamma}} \mathrm{d}y \right) \left| \ln \int_0^1 \mathrm{d}v \right| = 0,$$

and $I(1) = I_-(1) + I_+(1)$, with

$$I_-(1) = \int_0^{1/\sqrt{\gamma}} \mathrm{d}x \int_0^{1-x\sqrt{\gamma}} \mathrm{d}y \left| \ln \int_0^{y+x\sqrt{\gamma}} \mathrm{d}v \right|, \quad I_+(1) = \int_{1-1/\sqrt{\gamma}}^{1} \mathrm{d}x \int_{(1-x)\sqrt{\gamma}}^{1} \mathrm{d}y \left| \ln \int_{y+(x-1)\sqrt{\gamma}}^{1} \mathrm{d}v \right|.$$

Note that the integrals under the logarithms in $I_-(1)$ and $I_+(1)$ are between 0 and 1.

Thus,

$$I_-(1) = -\int_0^{1/\sqrt{\gamma}} \mathrm{d}x \int_0^{1-x\sqrt{\gamma}} \mathrm{d}y \ln \left(y + x\sqrt{\gamma}\right) = \int_0^{1/\sqrt{\gamma}} \mathrm{d}x \Big[1 - x\sqrt{\gamma} + x\sqrt{\gamma} \ln \left(x\sqrt{\gamma}\right)\Big] = \frac{1}{4\sqrt{\gamma}}.$$

Similarly, $I_+(1) = 1/\big(4\sqrt{\gamma}\big)$ and so $I = I(1) = 1/\big(2\sqrt{\gamma}\big)$. The assertion of Lemma 2.1 then follows.

In the multidimensional case, when $X \sim U \sim$U$([0,1]^d)$, a similar argument gives that $I = \Big[1 + 1/\big(2\sqrt{\gamma}\big)\Big]^d - 1$, and the assertion of Lemma 2.1 again follows.

The next case to analyse is (still scalar) where $X \sim$U$[a,b]$ and $= U \sim$U$[A,B]$. Take $\gamma > 4(B-A)^2/(b-a)^2$ and adopt the same scheme as before:

$$I = \int_a^b \frac{\mathrm{d}x}{b-a} \int_A^B \frac{\mathrm{d}y}{B-A} \left| \ln \int_{A\vee[y+(x-b)\sqrt{\gamma}]}^{B\wedge[y+(x-a)\sqrt{\gamma}]} \frac{\mathrm{d}v}{B-A} \right| = I(0) + I(1),$$

where

$$\begin{aligned} I(0) \; &= \left(\int_a^{a+(B-A)/\sqrt{\gamma}} \frac{\mathrm{d}x}{b-a} \int_{B-(x-a)\sqrt{\gamma}}^{B} \frac{\mathrm{d}y}{B-A} + \int_{a+(B-A)/\sqrt{\gamma}}^{b-(B-A)/\sqrt{\gamma}} \frac{\mathrm{d}x}{b-a} \int_A^B \frac{\mathrm{d}y}{B-A} \right. \\ &\left. + \int_{b-(B-A)/\sqrt{\gamma}}^{b} \frac{\mathrm{d}x}{b-a} \int_A^{A+(b-x)\sqrt{\gamma}} \frac{\mathrm{d}y}{B-A} \right) \left| \ln \int_A^B \frac{\mathrm{d}v}{B-A} \right| = 0, \end{aligned}$$

and $I(1) = I_-(1) + I_+(1)$, with

$$I_-(1) = \int_a^{a+(B-A)/\sqrt{\gamma}} \frac{\mathrm{d}x}{b-a} \int_A^{B-(x-a)\sqrt{\gamma}} \frac{\mathrm{d}y}{B-A} \left| \ln \int_A^{y+(x-a)\sqrt{\gamma}} \frac{\mathrm{d}v}{B-A} \right|,$$

and

$$I_+(1) = \int\limits_{b-(B-A)/\sqrt{\gamma}}^{b} \frac{\mathrm{d}x}{b-a} \int\limits_{A+(b-x)\sqrt{\gamma}}^{B} \frac{\mathrm{d}y}{B-A} \left| \ln \int\limits_{y+(x-b)\sqrt{\gamma}}^{B} \frac{\mathrm{d}v}{B-A} \right|.$$

The analysis of integrals $I_\pm(1)$ proceeds along the same lines as before. Viz., setting $\Gamma = \dfrac{(b-a)^2}{(B-A)^2}\gamma$, with $\sqrt{\Gamma} > 2$, we have that

$$\begin{aligned} I_-(1) &= -\int\limits_{a}^{a+(B-A)/\sqrt{\gamma}} \frac{\mathrm{d}x}{b-a} \int\limits_{A}^{B-(x-a)\sqrt{\gamma}} \frac{\mathrm{d}y}{B-A} \ln \frac{y+(x-a)\sqrt{\gamma}-A}{B-A} \\ &= -\int\limits_{0}^{1/\sqrt{\Gamma}} \mathrm{d}x \int\limits_{0}^{1-x\sqrt{\Gamma}} \mathrm{d}y \ln\left(y+x\sqrt{\Gamma}\right) \end{aligned}$$

which equals $1/\left(4\sqrt{\gamma}\right)$ by the above argument. Similarly, $I_+(1) = 1/\left(4\sqrt{\gamma}\right)$. Thus, in the case under consideration,

$$I = I(1) = \frac{B-A}{2(b-a)\sqrt{\gamma}}.$$

This goes to 0 as $\gamma \to +\infty$ which again proves the assertion of Lemma 2.1.

In the multi-dimensional case, with $X \sim \mathrm{U}\left(\underset{1\le j\le d}{\times}[a_j, b_j]\right)$ and $U \sim \mathrm{U}\left(\underset{1\le j\le d}{\times}[A_j, B_j]\right)$, we obtain, in a similar fashion, that

$$I = \prod_{1\le j\le d}\left[1 + \frac{B_j - A_j}{2(b_j - a_j)\sqrt{\gamma}}\right] - 1,$$

and Lemma 2.1 again follows.

The above argument remains pretty much intact in a more general case where PDF f_X admits finitely many values and PDF f_U has a compact support. More preciseley, consider first a scalar case where $f_X(x) = \sum\limits_i \alpha_i \mathbf{1}(a_i < x < b_i)$ and intervals $(a_i, b_i) \subset \mathbb{R}$ are pairwise disjoint with $\sum\limits_i \alpha_i(b_i - a_i) = 1$, while $f_U(x) = 0$ for $x \in \mathbb{R}\setminus(A, B)$. Take $\gamma > 4(B-A)^2/\left[\min\,(b_i - a_i)^2\right]$. Then

$$\begin{aligned} I &= \int \mathrm{d}x \mathbf{1}\left(f_X(x) > 0\right) f_X(x) \int \mathrm{d}y f_U(y) \left| \ln \left[\int f_X\left(x + \frac{y-v}{\sqrt{\gamma}}\right) f_U(v)\mathrm{d}v \Big/ f_X(x) \right] \right| \\ &= \sum_i \alpha_i \int\limits_{a_i}^{b_i} \mathrm{d}x \int\limits_{A}^{B} \mathrm{d}y f_U(y) \left| \ln \left[\sum_k \alpha_k \int\limits_{A\vee(y+(x-b_k)\sqrt{\gamma})}^{B\wedge(y+(x-a_k)\sqrt{\gamma})} f_U(v)\mathrm{d}v \Bigg/ \alpha_i \right] \right| \;=\; I(0) + I(1). \end{aligned}$$

Here $I(0) = I_-(0) + I_0(0) + I_+(0)$ and $I(1) = I_-(1) + I_+(1)$ where

$$I_-(0) = \sum_i \alpha_i \int\limits_{a_i}^{a_i+(B-A)/\sqrt{\gamma}} \mathrm{d}x \int\limits_{B-(x-a_i)\sqrt{\gamma}}^{B} \mathrm{d}y f_U(y) \ln \left[\frac{\alpha_i}{\alpha_i} \int\limits_{A}^{B} \mathrm{d}v f_U(v) \;+\; 0 \right] = 0,$$

$$I_0(0) = \sum_i \alpha_i \int\limits_{a_i+(B-A)/\sqrt{\gamma}}^{b_i-(B-A)/\sqrt{\gamma}} \mathrm{d}x \int\limits_{A}^{B} \mathrm{d}y f_U(y) \ln \left[\frac{\alpha_i}{\alpha_i} \int\limits_{A}^{B} \mathrm{d}v f_U(v) \;+\; 0 \right] = 0,$$

$$I_+(0) = \sum_i \alpha_i \int\limits_{b_i-(B-A)/\sqrt{\gamma}}^{b_i} \mathrm{d}x \int\limits_{A}^{A+(b_i-x)\sqrt{\gamma}} \mathrm{d}y f_U(y) \ln \left[\frac{\alpha_i}{\alpha_i} \int\limits_{A}^{B} \mathrm{d}v f_U(v) \;+\; 0 \right] = 0,$$

and

$$I_+(1) = \sum_i \alpha_i \int\limits_{a_i}^{a_i+(B-A)/\sqrt{\gamma}} \mathrm{d}x \int\limits_{A}^{B-(x-a_i)\sqrt{\gamma}} \mathrm{d}y f_U(y) \ln \left[\int\limits_{A}^{y+(x-a_i)\sqrt{\gamma}} \mathrm{d}v f_U(v) \; + \dots \right],$$

$$I_+(1) = \sum_i \alpha_i \int\limits_{b_i-(B-A)/\sqrt{\gamma}}^{b_i} \mathrm{d}x \int\limits_{A+(b_i-x)\sqrt{\gamma}}^{B} \mathrm{d}y f_U(y) \ln \left[\int\limits_{y+(x-b_i)\sqrt{\gamma}}^{B} \mathrm{d}v f_U(v) \; + \dots \right].$$

■

Remark. An assertion of the type of Lemma 2.1 is crucial for deriving the EPI by a direct probabilistic method, and the fact that it was not provided in [VG] made the proof of the EPI given in [VG] incomplete (the same is true of other papers on this subject).

In the discrete case where signal X takes finitely or countably many values, one has the following

Lemma 2.2. *Let X and U be independent RVs. Assume that X admits discrete values x_1, x_2, … with probabilities $p_X(x_1)$, $p_X(x_2)$, …, and has $h(X) = -\sum\limits_{x_i} p_X(x_i) \ln\ p_X(x) < +\infty$. Next, assume that U has a bounded PDF $f_U(x)$ with $\displaystyle\int f_U(x)\mathrm{d}x = 1$ and $\sup\,[f_U(x) :\ x \in \mathbb{R}^d] = a < +\infty$, and*

$$\lim_{\alpha\to\pm\infty} f_U(x + \alpha x_0) = 0, \;\; \forall\, x, x_0 \in \mathbb{R}^d \;\; \textit{with } x_0 \neq 0.$$

Finally, suppose that $\displaystyle\int f_U(x)\big|\ln\ f_U(x)\big|\mathrm{d}x < +\infty$. Then

$$h(X) = \lim_{\gamma\to\infty} I(X : \sqrt{\gamma}X + U). \tag{3}$$

Proof of Lemma 2.2. Setting as before, $Y = \sqrt{\gamma}X + U$, we again reduce our task to proving that $h(X|Y) \to 0$.

Now write

$$\begin{aligned} h(X|Y) &= -\sum_{i\geq 1} p_X(x_i) \int f_U(y - x_i\sqrt{\gamma}) \ln\ \frac{p_X(x_i) f_U(y - x_i\sqrt{\gamma})}{\sum\limits_{j\geq 1} p_X(x_j) f_U(y - x_j\sqrt{\gamma})} \mathrm{d}y \\ &= \sum_{i\geq 1} p_X(x_i) \int f_U(y) \ln\ \left[1 + \sum_{j:j\neq i} p_X(x_j) p_X(x_i)^{-1} f_U\big(y + (x_j - x_i)\sqrt{\gamma}\big) f_U(y)^{-1} \right] \mathrm{d}y. \end{aligned} \tag{4}$$

The expression under the logarithm clearly converges to 1 as $\gamma \to +\infty$, $\forall\, i \geq 1$ and $y \in \mathbb{R}^d$. Thus, the whole integrand

$$f_U(y) \ln\ \left[1 + \sum_{j:j\neq i} p_X(x_j) p_X(x_i)^{-1} f_U\big(y + (x_j - x_i)\sqrt{\gamma}\big) f_U(y)^{-1} \right] \to 0, \;\; \forall\, i \geq 1,\; y \in \mathbb{R}^d.$$

To guarantee the convergence of the integral we set $q_i = \sum\limits_{j:j\neq i} p_X(x_j) p_X(x_i)^{-1} = p_X(i)^{-1} - 1$ and

$\psi(y) = \ln\ f_U(y)$ and use the bound

$$\ln\ \left[1 + \sum_{j:j\neq i} p_X(x_j)p_X(x_i)^{-1} f_U(y - x_j\sqrt{\gamma}) f_U(y)^{-1}\right] \leq \ln\ \left(1 + aq_i e^{-\psi(y)}\right)$$
$$\leq \mathbf{1}\big(aq_i e^{-\psi(y)} > 1\big) \ln\ \left(2aq_i e^{-\psi(y)}\right) + \mathbf{1}\big(aq_i e^{-\psi(y)} \leq 1\big) \ln\ 2$$
$$\leq 2\ln\ 2 + \ln\ a + \ln\ (q_i + 1) + |\psi(y)|.$$

We then again apply Lebesgue's dominated convergence theorem and deduce that $\lim\limits_{\gamma\to+\infty} h(X|Y) = 0$. ■

In the general case, the arguments developed lead to the following continuity property:

Lemma 2.3. *Let RVs X and U be independent. Assume a general case where X may have discrete and absolutely continuous parts on its distribution while U has a PDF f_U with $\int f_U(x)\mathrm{d}x = 1$. Suppose the PDF f_X, with $\int f_X(x)\mathrm{d}x := 1 - \eta(X) \leq 1$, is continuous and bounded. Next, suppose that the PDF f_U is bounded and*

$$\lim_{\alpha\to\pm\infty} f_U(u + u_0\alpha) = 0,\ \ \forall\, u, u_0 \in \mathbb{R}^d,\ \ \textit{with}\ \ u_0 \neq 0.$$

Finally, assume that $\int g_X(u)\big|\ln\ g_X(u)\big| m_X(\mathrm{d}u) + \int f_U(u)\big|\ln\ f_U(u)\big|\mathrm{d}u < +\infty$. Then

$$h(X) = \lim_{\gamma\to\infty} \Big[I(X : X\sqrt{\gamma} + U) + [1 - \eta(X)] h\left(U/\sqrt{\gamma}\right)\Big].$$

The proof of the EPI in Section 3 requires an analysis of the behaviour of $I(X : X\sqrt{\gamma} + N)$ also when $\gamma \to 0$. Here we are able to cover a general case for RV X in a single assertion:

Lemma 2.4. *Let X, U be independent RVs. Assume that U has a bounded and continuous PDF $f_U \in C^0(\mathbb{R}^d)$, with $\int f_U(x)\mathrm{d}x = 1$ and $\sup\,[f_U(x)\ x \in \mathbb{R}^d] = a < +\infty$ whereas the distribution of X may have discrete and continuous parts. Next, assume, as in Lemma 2.3, that*

$$\int g_X(u)\big|\ln\ g_X(u)\big| m_X(\mathrm{d}u) + \int f_U(u)\big|\ln\ f_U(u)\big|\mathrm{d}u < +\infty.$$

Then

$$\lim_{\gamma\to 0} I\big(X : X\sqrt{\gamma} + U\big) = 0. \tag{7}$$

Proof of Lemma 2.4. Setting again $Y = X\sqrt{\gamma} + U$, we now reduce the the task to proving that $h(X|Y) \to h(X)$. Here we write

$$\begin{aligned} h(X|Y) &= -\int g_X(x) \int f_U\big(y - x\sqrt{\gamma}\big) \ln\ \frac{g_X(x) f_U\big(y - x\sqrt{\gamma}\big)}{\int g_X(u) f_U\big(y - u\sqrt{\gamma}\big) m_X(\mathrm{d}u)} \mathrm{d}y m_X(\mathrm{d}x) \\ &= \int g_X(x) \int f_U(y) \ln \left[\frac{\int g_X(u) f_U\big(y + (x-u)\sqrt{\gamma}\big) m_X(\mathrm{d}u)}{g_X(x) f_U(y)}\right] \mathrm{d}y m_X(\mathrm{d}x). \end{aligned} \tag{8}$$

Due to continuity of f_U, the ratio under the logarithm converges to $(g_X(x))^{-1}$ as $\gamma \to 0$, $\forall\ x, y \in \mathbb{R}$. Hence, the integral in (10) converges to $h(X)$ as $\gamma \to 0$. Again, the proof is completed with the help of the Lebesgue dominated convergence theorem. ■

Remark. Lemma 2.4 is another example of a missing step in proposed direct probabilistic proofs of the EPI.

3. The entropy-power inequality

In this section we show how to deduce the EPI from the lemmas established in Section 2. We begin with a convenient representation of the mutual entropy $I(X : X\sqrt{\gamma}+U)$ in the case where U is a d-variate normal RV.

Let ϕ_Σ (or, briefly, ϕ) stand for the standard d-variate normal PDF with mean vector 0 and a $d \times d$ covariance matrix Σ:

$$\phi_\Sigma(x) = \frac{1}{(2\pi)^{d/2}\text{det }\Sigma^{1/2}} \exp\left[-\frac{1}{2}\left\langle x, \Sigma^{-1}x\right\rangle\right],\ x \in \mathbb{R}^d.$$

Here and below, $\left\langle\,\cdot\,,\,\cdot\,\right\rangle$ stands for the Euclidean scalar product in $\mathbb{R}^d$, and we assume that Σ is strictly positive definite. The fact that an RV N is multivariate normal is written shortly as $N \sim$N$(0,\Sigma)$.

Lemma 3.1. *Let X and N be two independent RV, where $N \sim N(0,\Sigma)$ while the distribution of X may have a discrete and a continuous part. Suppose that $\displaystyle\int g_X(x)||x||^2 m_X(\mathrm{d}x) < +\infty$. Given $\gamma > 0$, write the mutual entropy between X and $X\sqrt{\gamma}+N$:*

$$\begin{aligned} I(X : X\sqrt{\gamma}+N) \quad &= -\int g_X(x)\phi(u-x\sqrt{\gamma})\ln\ \left[g_X(x)\phi(u-x\sqrt{\gamma})\right]\mathrm{d}u m_X(\mathrm{d}x) \\ &+\int g_X(x)\ln\ g_X(x)m_X(\mathrm{d}x) \\ &+\int f_{\sqrt{\gamma}X+N}(u)\ln\ f_{\sqrt{\gamma}X+N}(u)\mathrm{d}u \end{aligned}$$

where

$$f_{\sqrt{\gamma}X+N}(u) = \int g_X(x)\phi(u-x\sqrt{\gamma})m_X(\mathrm{d}x).$$

Then

$$\frac{\mathrm{d}}{\mathrm{d}\gamma}\Big[I(X : X\sqrt{\gamma}+N) + h\big(N/\sqrt{\gamma}\big)\Big] = \frac{1}{2}M(X;\gamma) - \frac{1}{2\gamma} \tag{10}$$

where

$$M(X;\gamma) = \mathbb{E}\Big[\big\|X - \mathbb{E}(X|X+\sqrt{\gamma}N)\big\|^2\Big]$$

Where is dependence on Σ in the RHS of (10)??

An analog of Lemma 3.1 has been established in [VG] and [GSV] in a scalar case(??) and under a (tacit) assumption that RV X has a PDF f_X, with $\displaystyle\int f_X(x)\mathrm{d}x = 1$. See [VG], Lemma ??. We present here an elementary proof under the assumptions stated above.

Proof of Lemma 3.1. Differentiate expression for $I(X : X\sqrt{\gamma}+N)$ given in (3), and observe that the derivative of the joint entropy $h(X, X\sqrt{\gamma}+N)$ vanishes, as $h(X, X\sqrt{\gamma}+N)$ does not change with $\gamma > 0$:

$$\begin{aligned} &h(X, X\sqrt{\gamma}+N) \\ &\quad = -\int g_X(x)\phi(u-x\sqrt{\gamma})\big[\ln\ g_X(x) + \ln\ \phi(u-x\sqrt{\gamma})\big]m_X(\mathrm{d}x)\mathrm{d}u \\ &\qquad = h(X) + h(N). \end{aligned}$$

The derivative of the marginal entropy $h(X\sqrt{\gamma}+N)$ requires some calculations:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\gamma}h(\sqrt{\gamma}X+N) &= -\frac{\mathrm{d}}{\mathrm{d}\gamma}\int f_{\sqrt{\gamma}X+N}(u)\ln\ f_{\sqrt{\gamma}X+N}(u)\mathrm{d}u\\ &= \int\frac{1}{2\sqrt{\gamma}}\int g_X(y)\phi(u-\sqrt{\gamma}y)\left\langle(u-\sqrt{\gamma}y),\Sigma^{-1}y\right\rangle m_X(\mathrm{d}y)\\ &\qquad\times\ln\int g_X(z)\phi(u-\sqrt{\gamma}z)m_X(\mathrm{d}z)\mathrm{d}u\\ &\quad+\int\frac{1}{2\sqrt{\gamma}}\int g_X(y)\phi(u-\sqrt{\gamma}y)m_X(\mathrm{d}y)\\ &\quad\times\frac{\displaystyle\int g_X(w)\phi(u-\sqrt{\gamma}w)\left\langle(u-\sqrt{\gamma}w),\Sigma^{-1}w\right\rangle m_X(\mathrm{d}w)}{\displaystyle\int g_X(z)\phi(u-\sqrt{\gamma}z)m_X(\mathrm{d}z)}\mathrm{d}u.\end{aligned}\tag{11}$$

The second summand vanishes, as (i) the integrals

$$\int g_X(y)\phi(u-\sqrt{\gamma}y)m_X(\mathrm{d}y)\ \text{ and }\ \int g_X(z)\phi(u-\sqrt{\gamma}z)m_X(\mathrm{d}z)$$

cancel each other and (ii) the remaining integration can be taken first in du??, which yields 0 for $\forall\ w$. The first integral we integrate by parts. This leads to the representation

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}\gamma}I(X:\sqrt{\gamma}X+N)=\\ &=\frac{1}{2\sqrt{\gamma}}\int\int g_X(y)\phi(u-\sqrt{\gamma}y)\\ &\times\frac{\displaystyle\int g_X(x)\phi(u-\sqrt{\gamma}x)\left\langle(u-\sqrt{\gamma}x),\Sigma^{-1}y\right\rangle m_X(\mathrm{d}x)}{\displaystyle\int g_X(z)\phi(u-\sqrt{\gamma}z)m_X(\mathrm{d}z)}m_X(\mathrm{d}y)\mathrm{d}u\\ &=\frac{1}{2\sqrt{\gamma}}\int\int g_X(y)\phi(u-\sqrt{\gamma}y)\\ &\times\frac{\displaystyle\int g_X(x)\phi(u-\sqrt{\gamma}x)\Big[\left\langle(u-\sqrt{\gamma}y,\Sigma^{-1}y\right\rangle+\sqrt{\gamma}\left\langle(y-x),\Sigma^{-1}y\right\rangle\Big]m_X(\mathrm{d}x)}{\displaystyle\int g_X(z)\phi(u-\sqrt{\gamma}z)m_X(\mathrm{d}z)}m_X(\mathrm{d}y)\mathrm{d}u.\end{aligned}$$

The integral arising from the summand $\left\langle(u-\sqrt{\gamma}y,\Sigma^{-1}y\right\rangle$ vanishes, because the mean vector in PDF ϕ is zero. The remaining contributions, from $\left\langle y,\Sigma^{-1}y\right\rangle-\left\langle x,\Sigma^{-1}y\right\rangle$, is equal to

$$\frac{1}{2}\mathbb{E}\Big[\ \|X-\mathbb{E}(X|\sqrt{\gamma}X+N)\|^2\ \Big].$$

On the other hand, the first term in RHS of (11) equals

$$\begin{aligned}&\int\left\|x-\frac{\displaystyle\int g_X(y)\phi(u-\sqrt{\gamma}y)\Sigma??y m_X(\mathrm{d}y)}{\displaystyle\int g_X(z)\phi(u-\sqrt{\gamma}z)m_X(\mathrm{d}z)}\right\|^2 g_X(x)\phi(u-\sqrt{\gamma}x)m_X(\mathrm{d}x)\mathrm{d}u\\ &\quad=\mathbb{E}\Big[\|X-\mathbb{E}(X|\sqrt{\gamma}X+N)\|^2\Big]\equiv M(X;\gamma).\qquad\blacksquare\end{aligned}$$

We are now going to derive the EPI (1), (2). We follow the line of argument proposed in [VG] and based on Lemma 3.1. First, suppose that X is a RV with a PDF f_X where $\displaystyle\int f_X(x)\mathrm{d}x=1$. Then we assume that $f_X(x)$ satisfies the assumptions stated in Lemma 4 and Lemma 2 and use these lemmas with $U = N \sim \mathrm{N}(0, \Sigma)$. Consequently, $\forall\ \epsilon > 0$,

$$\begin{aligned} h(X) &= \lim_{\gamma\to+\infty}\Big[I\big(X : X\sqrt{\gamma}+N\big)+h\big(N/\sqrt{\gamma}\big)\Big] \\ &= \int\limits_{\epsilon}^{+\infty}\frac{\mathrm{d}}{\mathrm{d}\gamma}\left[I\big(X : X\sqrt{\gamma}+N\big)+h\big(N/\sqrt{\gamma}\big)\right]\mathrm{d}\gamma + I\big(X : X\sqrt{\epsilon}+N\big)+h\big(N/\sqrt{\epsilon}\big) \\ &= \frac{1}{2}\int_{\epsilon}^{+\infty}\Big[M(X;\gamma)-\frac{1}{\gamma}\mathbf{1}(\gamma>1)\Big]\mathrm{d}\gamma + h(N) + I\big(X : X\sqrt{\epsilon}+N\big). \end{aligned} \tag{12}$$

Here we use the identity $\int\limits_{\epsilon}^{1}(1/\gamma)\mathrm{d}\gamma = \ln\epsilon$. By Lemma 3 the last term in (12) tends to 0 as $\epsilon \to 0$. Hence, for an RV X with PDF $f_X \in C^0$ we obtain

$$h(X) = h(N) + \frac{1}{2}\int_0^{\infty}\Big[M(X;\gamma) - \mathbf{1}(\gamma > 1)\frac{1}{\gamma}\Big]\mathrm{d}\gamma. \tag{13}$$

In the case where X attains discrete values in $\mathbb{R}^d$, Eqn (13) is replaced by

$$h(X) = h(N)+) + \frac{1}{2}\int_0^{\infty} M(X;\gamma)\mathrm{d}\gamma \tag{14}$$

and is established under a simple assumption that $h(X) < +\infty$, by virtue of Lemmas 1 and 3–4.

The proof of EPI is based on Eqn (13) and the following result from [L].

Lemma 5. ([L], Theorem 6) *Let $\mathcal{X}$ be a given class of probability distributions on $\mathbb{R}^d$. The inequality*

$$h(X_1\cos\theta + X_2\sin\psi) \geq h(X_1)\cos^2\theta + h(X_2)\sin^2\theta, \tag{15}$$

for any $\theta \in [0, 2\pi]$ and any pair of independent RVs X_1, X_2 with distributions from $\mathcal{X}$, holds true iff the Entropy Power inequality is valid for any pair of RVs X_1, X_2 with distributions from $\mathcal{X}$.

Theorem 1. *Let X_1, X_2 be RVs with values in $\mathbb{R}^d$ and with continuous and bounded PDFs $f_{X_1}(x)$, $f_{X_2}(x)$, $x \in \mathbb{R}^d$. Assume that the differential entropies $h(X_1)$ and $h(X_2)$ satisfy $-\infty < h(X_1), h(X_2) < +\infty$. Then the EPI (see Eqns (1)–(2)) holds true.*

Proof According to Lemma 5, it suffices to check bound (14) $\forall\ \theta \in (0, 2\pi)$ and $\forall$ pair of RVs X_1, X_2 with continuous and bounded PDFs $f_{X_i}(x)$, $i = 1, 2$. Take any such pair and let N be $\mathrm{N}(0, \mathbf{I})$ where $\mathbf{I}$ is the $d \times d$ unit matrix. Following the argument developed in [GV], we apply formula (13) for the RV $X_1\cos\phi + X_2\sin\phi$:

$$h(X_1\cos\phi + X_2\sin\phi) = h(N) + \frac{1}{2}\int_0^{\infty}\Big[M(X_1\cos\phi + X_2\sin\phi;\gamma) - \mathbf{1}(\gamma > 1)\frac{1}{\gamma}\Big]\mathrm{d}\gamma.$$

To verify Eqn (14) we need to check that

$$M(X_1\cos\phi + X_2\sin\phi;\gamma) \geq \cos\phi^2 M(X_1;\gamma) + \sin\phi^2 M(X_2;\gamma). \tag{16}$$

To this end, we take two independent RVs $N_1, N_2 \sim \mathrm{N}(0, \mathbf{I})$ and set

$$Z_1 = \sqrt{\gamma}X_1 + N_1, Z_2 = \sqrt{\gamma}X_2 + N_2, \quad \text{and} \quad Z = Z_1\cos\phi + Z_2\sin\phi.$$

Then inequality (15) holds true because

$$\begin{aligned}
&\mathbb{E}\Big[\,\|X-\mathbb{E}(X|Z)\|^2\Big] \geq \mathbb{E}\Big[\,\|X-\mathbb{E}(X|Z_1,Z_2)\|^2\Big] \\
&\quad= \mathbb{E}\Big[\,\|X_1-\mathbb{E}(X_1|Z_1)\|^2\Big]\cos\phi^2 + \mathbb{E}\Big[\,\|X_2-\mathbb{E}(X_2|Z_2)\|^2\Big]\sin\phi^2. \qquad \blacksquare
\end{aligned}$$

For discrete RVs, the assertion is:

Theorem 2. *Let X_1, X_2 be RVs with discrete values in $\mathbb{R}^d$. Assume that the entropies $h(X_1), h(X_2 < +\infty$. Then the EPI holds true.*

The proof of Theorem 2 follows the same scheme as that of Theorem 1. Finally, in the same fashion one obtains the more general

Theorem 3. *Let X_1, X_2 be RVs with values in $\mathbb{R}^d$ and general distributions including discrete and absolutely continuous parts. Assume that the entropies $h(X_1), h(X_2)$ satisfy $-\infty < h(X_1), h(X_2) < +\infty$ and that the PDFs f_{X_i} are are bounded and continuous. Then the EPI holds true.*

References

[D] Dobrushin R.L., Passage to the limit under the information and entropy signs. *Theory Probab. Appl.*, **5** (1959), 1, 25-32

[CT] Cover, T.M., and Thomas, J.A.. Elements of Information Theory. N.Y.: Wiley, 2006

[GSV] Guo D., Shamai, S., and Verdu, S. Mutual information and minimum mean-square error in Gaussian channels. *IEEE Trans. on Inform. Theory*, **51** (2005) 4, 1261-1288

[L] Lieb E.H., Proof of entropy conjecture of Wehrl. *Commun. Math. Phys.*, **62** (1978), 1, 35-41

[R] Rioul, O. A simple proof of the entropy-power inequality via properties of mutual information. In: Proceeding of the 2007 IEEE Inter. Symp. on Information Theory, Nice, France, June 2007

[SW] Shannon, C., and Weaver, W. *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, IL, 1949

[VG] Verdu, S., and Guo D. A simple proof of the entropy–power inequality. *IEEE Trans. Inform. Theory*, **52** (2006) 5, 2165-2166