Weighted entropy: basic inequalities

Mark Kelbert[a,*]**, Izabella Stuhl**[b]**, Yuri Suhov**[c,d]

[a]*Higher School of Economics, Moscow, RF*
[b]*Math Dept, Penn State University, PA, USA*
[c]*Cambridge University, UK*
[d]*Penn State University, PA, USA*

mkelbert@hse.ru (M. Kelbert), stuhlizabella@googlemail.com (I. Stuhl),
yms@statslab.cam.ac.uk (Y. Suhov)

Abstract   This paper represents an extended version of an earlier note [10]. The concept of weighted entropy takes into account values of different outcomes, i.e., makes entropy context-dependent, through the weight function. We analyse analogs of the Fisher information inequality and entropy power inequality for the weighted entropy and discuss connections with weighted Lieb's splitting inequality. The concepts of rates of the weighted entropy and information are also discussed.

Keywords   Weighted entropy, Gibbs inequality, Ky-Fan inequality, Fisher information inequality, entropy power inequality, Lieb's splitting inequality, rates of weighted entropy and information

2010 MSC   94A17

1   Introduction

This paper represents an extended version of an earlier note [10].[1] We also follow earlier publications discussing related topics: [20, 21, 19, 18]. The Shannon entropy

---

*Corresponding author.
[1]AMCTA-2017, Talks at Conference "Analytical and Computational Methods in Probability Theory and its Applications"

VTeX

(SE) of a probability distribution **p** or the Shannon differential entropy (SDE) of a probability density function (PDF) f

$$h(\mathbf{p}) = -\sum_i p(x_i) \log p(x_i), \qquad h(f) = -\int f(x) \log f(x) dx \qquad (1.1)$$

is context-free, i.e., does not depend on the nature of outcomes x_i or x, but only upon probabilities $p(x_i)$ or values $f(x)$. It gives the notion of entropy a great flexibility which explains its successful applications. However, in many situations it seems insufficient, and the context-free property appears as a drawback. Viz., suppose you learn a news about severe weather conditions in an area far away from your place. Such conditions usually do not happen; an event like this has a small probability $p \ll 1$ and conveys a high information $-\log p$. At the same time you hear that a tree near your parking lot in the town has fallen and damaged a number of cars. The probability of this event is also low, so the amount of information is again high. However, the value of this information for you is higher than in the first event. Considerations of this character can motivate a study of weighted information and entropy, making them context-dependent.

Definition 1.1. Let us define the *weighted entropy* (WE) as

$$h_\varphi^{\mathrm{w}}(\mathbf{p}) = -\sum_i \varphi(x_i) p(x_i) \log p(x_i). \qquad (1.2)$$

Here a non-negative *weight function* (WF) $x_i \mapsto \varphi(x_i)$ is introduced, representing a value/utility of an outcomes x_i. A similar approach can be used for the differential entropy of a probability density function (PDF) f. Define the *weighted differential entropy* (WDE) as

$$h_\varphi^{\mathrm{w}}(f) = -\int \varphi(\mathbf{x}) f(\mathbf{x}) \log f(\mathbf{x}) d\mathbf{x}. \qquad (1.3)$$

An initial example of a WF φ may be $\varphi(\mathbf{x}) = \mathbf{1}(\mathbf{x} \in A)$ where A is a particular subset of outcomes (an event). A heuristic use of the WE with such a WF was demonstrated in [4, 5]. Another example repeatedly used below is $f(\mathbf{x}) = f_C^{\mathrm{No}}(\mathbf{x})$, a d-dimensional Gaussian PDF with mean $\mathbf{0}$ and covariance matrix C. Here

$$h_\varphi^{\mathrm{w}}(f_C^{\mathrm{No}}) = \frac{\alpha_\varphi(C)}{2} \log[(2\pi)^d \det(C)] + \frac{\log e}{2} \mathrm{tr}[C^{-1} \Phi_{C,\varphi}] \quad \text{where}$$

$$\alpha_\varphi(C) = \int_{\mathbb{R}^d} \varphi(\mathbf{x}) f_C^{\mathrm{No}}(\mathbf{x}) d\mathbf{x}, \quad \Phi_{C,\varphi} = \int_{\mathbb{R}^d} \mathbf{x}\mathbf{x}^{\mathrm{T}} \varphi(\mathbf{x}) f_C^{\mathrm{No}}(\mathbf{x}) d\mathbf{x}. \qquad (1.4)$$

For $\varphi(\mathbf{x}) = 1$ we get the normal SDE $h(f_C^{\mathrm{No}}) = \frac{1}{2} \log[(2\pi e)^d \det C]$.

In this note we give a brief introduction into the concept of the weighted entropy. We do not always give proofs, referring the reader to the quoted original papers. Some basic properties of WE and WDE have been presented in [20]; see also references therein to early works on the subject. Applications of the WE and WDE to the security quantification of information systems are discussed in [15]. Other domains range from the stock market to the image processing, see, e.g., [6, 9, 12, 14, 23, 26].

Throughout this note we assume that the series and integrals in (1.2)–(1.3) and the subsequent equations converge absolutely, without stressing it every time again. To unify the presentation, we will often use integrals $\int_{\mathcal{X}} d\mu$ relative to a reference σ-finite measure μ on a Polish space $\mathcal{X}$ with a Borel σ-algebra $\mathfrak{X}$. In this regard, the acronym PM/DF (probability mass/density function) will be employed. Usual measurability assumptions will also be in place for the rest of the presentation. We also assume that the WF $\varphi > 0$ on an open set in $\mathcal{X}$.

In some parts of the presentation, the sums and integrals comprising a PM/DF will be written as expectations: this will make it easier to explain/use assumptions and properties involved. Viz., Eqns (1.2)–(1.3) can be given as $h_\varphi^{\mathrm{w}}(\mathbf{p}) = -\mathbb{E}\varphi(\mathbf{X}) \log \mathbf{p}(\mathbf{X})$ and $h_\varphi^{\mathrm{w}}(f) = -\mathbb{E}\varphi(\mathbf{X}) \log f(\mathbf{X})$ where $\mathbf{X}$ is a random variable (RV) with the PM/DF $\mathbf{p}$ or f. Similarly, in (1.4), $\alpha_\varphi(C) = \mathbb{E}\varphi(\mathbf{X})$ and $\boldsymbol{\Phi}_{C,\varphi} = \mathbb{E}\varphi(\mathbf{X})\mathbf{X}\mathbf{X}^{\mathrm{T}}$ where $\mathbf{X} \sim \mathrm{N}(\mathbf{0}, C)$.

2 The weighted Gibbs inequality

Given two non-negative functions f, g (typically, PM/DFs), define the *weighted Kullback-Leibler divergence* (or the relative WE, briefly RWE) as

$$D_\varphi^{\mathrm{w}}(f\|g) = \int_{\mathcal{X}} \varphi(\mathbf{x}) f(\mathbf{x}) \log \frac{f(\mathbf{x})}{g(\mathbf{x})} d\mu(\mathbf{x}). \qquad (2.1)$$

Theorem 1.3 from [20] states:

Theorem 2.1. *Suppose that*

$$\int_{\mathcal{X}} \varphi(\mathbf{x})\big[f(\mathbf{x}) - g(\mathbf{x})\big] d\mu(\mathbf{x}) \geq 0. \qquad (2.2)$$

Then $D_\varphi^{\mathrm{w}}(f\|g) \geq 0$. Moreover, $D_\varphi^{\mathrm{w}}(f\|g) = 0$ iff $\varphi(\mathbf{x})[\frac{g(\mathbf{x})}{f(\mathbf{x})} - 1] = 0$ f-a.e.

Example 2.2. For an exponential family in the canonical form

$$f_{\underline{\theta}}(\mathbf{x}) = h(\mathbf{x}) \exp\big(\langle\underline{\theta}, T(\mathbf{x})\rangle - A(\underline{\theta})\big), \quad \mathbf{x} \in \mathbb{R}^d, \ \underline{\theta} \in \mathbb{R}^m, \qquad (2.3)$$

with the sufficient statistics $T(\mathbf{x})$ we have

$$D_\varphi^{\mathrm{w}}(f_{\underline{\theta}_1}\|f_{\underline{\theta}_2}) = e^{A_\varphi(\underline{\theta}_1) - A(\underline{\theta}_1)}\big(A(\underline{\theta}_2) - A(\underline{\theta}_1) - \langle\nabla A_\varphi(\underline{\theta}_1), \underline{\theta}_2 - \underline{\theta}_1\rangle\big), \qquad (2.4)$$

where ∇ stands for the gradient w.r.t. to the parameter vector $\underline{\theta}$, and

$$A_\varphi(\underline{\theta}) = \log \int \varphi(\mathbf{x}) h(\mathbf{x}) \exp\big(\langle\underline{\theta}, T(\mathbf{x})\rangle\big) d\mathbf{x}. \qquad (2.5)$$

3 Concavity/convexity of the weighted entropy

Theorems 2.1 and 2.2 from [20] offer the following assertion:

Theorem 3.1. (a) *The WE/WDE functional $f \mapsto h_\varphi^W(f)$ is concave in argument f. Namely, for any PM/DFs $f_1(x)$, $f_2(x)$ and $\lambda_1, \lambda_2 \in [0, 1]$ such that $\lambda_1 + \lambda_2 = 1$,*

$$h_\varphi^W(\lambda_1 f_1 + \lambda_2 f_2) \geq \lambda_1 h_\varphi^W(f_1) + \lambda_2 h_\varphi^W(f_2). \tag{3.1}$$

The equality iff $\varphi(x)[f_1(x) - f_2(x)] = 0$ holds for $(\lambda_1 f_1 + \lambda_2 f_2)$-a.a. x.

(b) *However, the RWE functional $(f, g) \mapsto D_\varphi^W(f \| g)$ is convex: given two pairs of PDFs (f_1, f_2) and (g_1, g_2),*

$$\lambda_1 D_\varphi^W(f_1 \| g_1) + \lambda_2 D_\varphi^W(f_2 \| g_2) \geq D_\varphi^W(\lambda_1 f_1 + \lambda_2 f_2 \| \lambda_1 g_1 + \lambda_2 g_2), \tag{3.2}$$

with equality iff $\lambda_1 \lambda_2 = 0$ or $\varphi(x)[f_1(x) - f_2(x)] = \varphi(x)[g_1(x) - g_2(x)] = 0$ μ-a.e.

4 Weighted Ky-Fan and Hadamard inequalities

The map $C \mapsto \delta(C) := \log \det(C)$ gives a concave function of a (strictly) positive-definite $(d \times d)$ matrix C: $\delta(C) - \lambda_1 \delta(C_1) - \lambda_2 \delta(C_2) \geq 0$, where $C = \lambda_1 C_1 + \lambda_2 C_2$, $\lambda_1 + \lambda_2 = 1$ and $\lambda_{1,2} \geq 0$. This is the well-known Ky-Fan inequality. It terms of differential entropies it is equivalent to the bound

$$h(f_C^{No}) - \lambda_1 h(f_{C_1}^{No}) - \lambda_2 h(f_{C_2}^{No}) \geq 0 \tag{4.1}$$

and is closely related to a maximising property of the Gaussian differential entropy $h(f_C^{No})$.

Theorem 4.1 below presents one of new bounds of Ky-Fan type, in its most explicit form, for the WF $\varphi(\mathbf{x}) = \exp(\mathbf{x}^T \mathbf{t})$, $\mathbf{t} \in \mathbb{R}^d$. Cf. Theorem 3.5 from [20]. In this case the identity $h_\varphi^W(f^{No}) = \exp(\frac{1}{2}\mathbf{t}^T C \mathbf{t}) h(f^{No})$ holds true. Introduce a set

$$\mathcal{S} = \{\mathbf{t} \in \mathbb{R}^d : F^{(1)}(\mathbf{t}) \geq 0, \ F^{(2)}(\mathbf{t}) \leq 0\}. \tag{4.2}$$

Here functions $F^{(1)}$ and $F^{(2)}$ incorporate parameters C_i and λ_i:

$$F^{(1)}(\mathbf{t}) = \sum_{i=1}^{2} \lambda_i \exp\left(\frac{1}{2}\mathbf{t}^T C_i \mathbf{t}\right) - \exp\left(\frac{1}{2}\mathbf{t}^T C \mathbf{t}\right), \quad \mathbf{t} \in \mathbb{R}^d,$$

$$F^{(2)}(\mathbf{t}) = \left[\sum_{i=1}^{2} \lambda_i \exp\left(\frac{1}{2}\mathbf{t}^T C_i \mathbf{t}\right) - \exp\left(\frac{1}{2}\mathbf{t}^T C \mathbf{t}\right)\right] \log[(2\pi)^d \det(C)]$$

$$+ \sum_{i=1}^{2} \lambda_i \exp\left(\frac{1}{2}\mathbf{t}^T C_i \mathbf{t}\right) \mathrm{tr}[C^{-1} C_i] - d \exp\left(\frac{1}{2}\mathbf{t}^T C \mathbf{t}\right), \quad \mathbf{t} \in \mathbb{R}^d. \tag{4.3}$$

Theorem 4.1. *Given positive-definite matrices C_1, C_2 and $\lambda_1, \lambda_2 \in [0, 1]$ with $\lambda_1 + \lambda_2 = 1$, set $C = \lambda_1 C_1 + \lambda_2 C_2$. Assume $\mathbf{t} \in \mathcal{S}$. Then*

$$h(f_C^{No}) \exp\left(\frac{1}{2}\mathbf{t}^T C \mathbf{t}\right) - h(f_{C_1}^{No}) \exp\left(\frac{1}{2}\mathbf{t}^T C_1 \mathbf{t}\right) - h(f_{C_2}^{No}) \exp\left(\frac{1}{2}\mathbf{t}^T C_2 \mathbf{t}\right) \geq 0, \tag{4.4}$$

with equality iff $\lambda_1 \lambda_2 = 0$ or $C_1 = C_2$.

For $\mathbf{t} = \mathbf{0}$ we obtain $\varphi \equiv 1$, and (4.4) coincides with (4.1). Theorem 4.1 is related to the maximisation property of the weighted Gaussian entropy which takes the form of Theorem 4.2. Cf. Example 3.2 in [20].

Theorem 4.2. *Let $f(\mathbf{x})$ be a PDF on $\mathbb{R}^d$ with mean $\mathbf{0}$ and $(d \times d)$ covariance matrix C. Let $f^{\mathrm{No}}(\mathbf{x})$ stand for the Gaussian PDF, again with the mean $\mathbf{0}$ and covariance matrix C. Define $(d \times d)$ matrices*

$$\boldsymbol{\Phi} = \int_{\mathbb{R}^d} \varphi(\mathbf{x}) \mathbf{x} \mathbf{x}^{\mathrm{T}} f(\mathbf{x}) d\mathbf{x}, \qquad \boldsymbol{\Phi}_C^{\mathrm{No}} = \int_{\mathbb{R}^d} \varphi(\mathbf{x}) \mathbf{x} \mathbf{x}^{\mathrm{T}} f_C^{\mathrm{No}}(\mathbf{x}) d\mathbf{x}. \qquad (4.5)$$

Cf. (1.4). Assume that

$$\int_{\mathbb{R}^d} \varphi(\mathbf{x}) \big[f(\mathbf{x}) - f_C^{\mathrm{No}}(\mathbf{x}) \big] d\mathbf{x} \geq 0$$

and

$$\log \big[(2\pi)^d (\det C) \big] \int_{\mathbb{R}^d} \varphi(\mathbf{x}) \big[f(\mathbf{x}) - f_C^{\mathrm{No}}(\mathbf{x}) \big] d\mathbf{x} + \mathrm{tr} \big[C^{-1} \big(\boldsymbol{\Phi}_C^{\mathrm{No}} - \boldsymbol{\Phi} \big) \big] \leq 0.$$

Then $h_\varphi^{\mathrm{w}}(f) \leq h_\varphi^{\mathrm{w}}(f_C^{\mathrm{No}})$, with equality iff $\varphi(\mathbf{x})[f(\mathbf{x}) - f_C^{\mathrm{No}}(\mathbf{x})] = 0$ a.e.

Theorems 4.1 and 4.2 are a part of a series of the so-called weighted *determinantal* inequalities. See [20, 22]. Here we will focus on a weighted version of *Hadamard* inequality asserting that for a $(d \times d)$ positive-definite matrix $C = (C_{ij})$, $\det C \leq \prod_{j=1}^d C_{jj}$ or $\delta(C) \leq \sum_{j=1}^d \log C_{jj}$. Cf. [20], Theorem 3.7. Let $f_{C_{jj}}^{\mathrm{No}}$ stand for the Gaussian PDF on $\mathbb{R}$ with the zero mean and the variance C_{jj}. Set:

$$\alpha = \int_{\mathbb{R}^d} \varphi(\mathbf{x}) f_C^{\mathrm{No}}(\mathbf{x}) d\mathbf{x} \quad (\text{cf. (1.4)}).$$

Theorem 4.3. *Assume that*

$$\int_{\mathbb{R}^d} \varphi(\mathbf{x}) \bigg[f_C^{\mathrm{No}}(\mathbf{x}) - \prod_{j=1}^d f_{C_{jj}}^{\mathrm{No}}(x_j) \bigg] d\mathbf{x} \geq 0.$$

Then, with the matrix $\boldsymbol{\Phi} = (\Phi_{ij})$ as in (4.5),

$$\alpha \log \prod_{j=1}^d (2\pi C_{jj}) + (\log e) \sum_{j=1}^d C_{jj}^{-1} \Phi_{jj} - \alpha \log \big[(2\pi)^d (\det C) \big] - (\log e) \mathrm{tr}\, C^{-1} \boldsymbol{\Phi} \geq 0.$$

5   A weighted Fisher information matrix

Let $\mathbf{X} = (X_1, \dots, X_d)$ be a random $(1 \times d)$ vector with PDF $f_{\underline{\theta}}(\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}, \underline{\theta})$ where $\underline{\theta} = (\theta_1, \dots, \theta_m) \in \mathbb{R}^m$. Suppose that $\underline{\theta} \to f_{\underline{\theta}}$ is C^1. Define a score vector $S(\mathbf{X}, \underline{\theta}) = \mathbf{1}(f_{\underline{\theta}}(\mathbf{x}) > 0)(\frac{\partial}{\partial \theta_i} \log f_{\underline{\theta}}(\mathbf{x}), i = 1, \dots, m)$. The $m \times m$ weighted Fisher information matrix (WFIM) is defined as

$$J_\varphi^{\mathrm{w}}(f_{\underline{\theta}}) = J_\varphi^{\mathrm{w}}(\mathbf{X}, \underline{\theta}) = \mathbb{E}\big[\varphi(\mathbf{X}) S(\mathbf{X}; \underline{\theta}) S^T(\mathbf{X}; \underline{\theta}) \big]. \qquad (5.1)$$

Theorem 5.1 (Connection between WFIM and weighted KL-divergence measures).
For smooth families $\{f_\theta, \theta \in \Theta \in \mathbb{R}^1\}$ and a given WF φ, we get

$$
D_\varphi^w(f_{\theta_1} \| f_{\theta_2}) = \frac{1}{2} J_\varphi^w(X, \theta_1)(\theta_2 - \theta_1)^2 + \mathbb{E}_{\theta_1}\big[\varphi(X) D_\theta \log f_{\theta_1}(X)\big](\theta_1 - \theta_2)
$$
$$
- \frac{1}{2} \mathbb{E}_{\theta_1}\left[\varphi(X) \frac{D_\theta^2 f_{\theta_1}(X)}{f_{\theta_1}(X)}\right](\theta_2 - \theta_1)^2 + o\big(|\theta_1 - \theta_2|^2\big) \mathbb{E}_{\theta_1}\big[\varphi(X)\big]
$$
(5.2)

where D_θ stands for $\frac{\partial}{\partial\theta}$.

Proof. By virtue of a Taylor expansion of $\log f_{\theta_2}$ around θ_1, we obtain

$$
\log f_{\theta_2} = \log f_{\theta_1} + D_\theta \log f_{\theta_1}(\theta_2 - \theta_1) + \frac{1}{2} D_\theta^2 \log f_{\theta_1}(\theta_2 - \theta_1)^2 + O_x\big(|\underline{\theta_2} - \underline{\theta_1}|^3\big).
$$
(5.3)

Here $O_x(|\theta_2 - \theta_1|^3)$ denotes the reminder term which has a hidden dependence on x.
Multiply both sides of (5.3) by φ and take expectations assuming that we can interchange differentiation and expectation appropriately. Next, observe that

$$
D_\theta^2 \log f_{\theta_1} = \frac{D_\theta^2 f_{\theta_1}}{f_{\theta_1}} - \frac{(D_\theta f_{\theta_1})^2}{f_{\theta_1}^2}.
$$
(5.4)

Hence

$$
\mathbb{E}_{\theta_1}\big[\varphi D_\theta^2 \log f_{\theta_1}\big] = \mathbb{E}_{\theta_1}\left[\varphi \frac{D_\theta^2 f_{\theta_1}}{f_{\theta_1}}\right] - J_\varphi^w(f_{\theta_1}).
$$
(5.5)

Therefore the claimed result, i.e., (5.2), is achieved. $\qquad\square$

6 Weighted entropy power inequality

Let $\mathbf{X}_1, \mathbf{X}_2$ be independent RVs with PDFs f_1, f_2 and $\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2$. The famous
Shannon entropy power inequality (EPI) states that

$$
h(\mathbf{X}_1 + \mathbf{X}_2) \geq h(\mathbf{N}_1 + \mathbf{N}_2),
$$
(6.1)

where $\mathbf{N}_1, \mathbf{N}_2$ are Gaussian $N(\mathbf{0}, \sigma^2 \mathbf{I}_d)$ RVs such that $h(\mathbf{X}_i) = h(\mathbf{N}_i), i = 1, 2$.
Equivalently,

$$
e^{\frac{2}{d} h(\mathbf{X}_1 + \mathbf{X}_2)} \geq e^{\frac{2}{d} h(\mathbf{X}_1)} + e^{\frac{2}{d} h(\mathbf{X}_2)},
$$
(6.2)

see, e.g., [1, 7]. The EPI is widely used in electronics, i.e., consider a RV $\mathbf{Y}$ which
satisfies

$$
\mathbf{Y}_n = \sum_{i=0}^{\infty} a_i \mathbf{X}_{n-i}, n \in \mathbf{Z}^1, \quad \sum_{i=0}^{\infty} |a_i|^2 < \infty,
$$
(6.3)

where $a_i \in \mathbf{R}^1$, $\{\mathbf{X}_i\}$ are IID RVs. Then the EPI means

$$
h(\mathbf{Y}) \geq h(\mathbf{X}) + \frac{1}{2} \log\left(\sum_{i=0}^{\infty} |a_i|^2\right),
$$
(6.4)

with equality if and only if either **X** is Gaussian or if $\mathbf{Y}_n = \mathbf{X}_{n-k}$, for some k, that is, the filtering operation is a pure delay. Clearly, a possible extension of the EPI gives more flexibility in signal processing. We are interested in the weighted entropy power inequality (WEPI)

$$\kappa := \exp\left(\frac{2h_\varphi^{\mathrm{w}}(\mathbf{X}_1)}{d\mathbb{E}\varphi(\mathbf{X}_1)}\right) + \exp\left(\frac{2h_\varphi^{\mathrm{w}}(\mathbf{X}_2)}{d\mathbb{E}\varphi(\mathbf{X}_2)}\right) \leq \exp\left(\frac{2h_\varphi^{\mathrm{w}}(\mathbf{X})}{d\mathbb{E}\varphi(\mathbf{X})}\right). \tag{6.5}$$

Note that (6.5) coincides with (6.2) when $\varphi \equiv 1$. Let $d = 1$, we set

$$\alpha = \tan^{-1}\left[\exp\left(\frac{h_\varphi^{\mathrm{w}}(X_2)}{\mathbb{E}\varphi(X_2)} - \frac{h_\varphi^{\mathrm{w}}(X_1)}{\mathbb{E}\varphi(X_1)}\right)\right], \quad Y_1 = \frac{X_1}{\cos\alpha}, Y_2 = \frac{X_2}{\sin\alpha}. \tag{6.6}$$

Theorem 6.1. *Given independent RVs $X_1, X_2 \in \mathbb{R}^1$ with PDFs f_1, f_2, and the weight function φ, set $X = X_1 + X_2$. Assume the following conditions:*
(i)

$$\mathbb{E}\varphi(X_i) \geq \mathbb{E}\varphi(X) \quad \text{if } \kappa \geq 1, \ i = 1, 2,$$
$$\mathbb{E}\varphi(X_i) \leq \mathbb{E}\varphi(X) \quad \text{if } \kappa \leq 1, \ i = 1, 2. \tag{6.7}$$

(ii) With Y_1, Y_2 and α as defined in (6.6),

$$(\cos\alpha)^2 h_{\varphi_c}^{\mathrm{w}}(Y_1) + (\sin\alpha)^2 h_{\varphi_s}^{\mathrm{w}}(Y_2) \leq h_\varphi^{\mathrm{w}}(X), \tag{6.8}$$

where $\varphi_c(x) = \varphi(x\cos\alpha)$, $\varphi_s(x) = \varphi(x\sin\alpha)$ and

$$h_{\varphi_c}^{\mathrm{w}}(Y_1) = -\mathbb{E}\big[\varphi_c(Y_1)\log\big(f_{Y_1}(Y_1)\big)\big], \qquad h_{\varphi_s}^{\mathrm{w}}(Y_2) = -\mathbb{E}\big[\varphi_s(Y_2)\log\big(f_{Y_2}(Y_2)\big)\big]. \tag{6.9}$$

Then the WEPI holds.

Paying homage to [13] we call (6.8) weighted Lieb's splitting inequality (WLSI). In some cases the WLSI may be effectively checked.

Proof. Note that

$$h_\varphi^{\mathrm{w}}(X_1) = h_{\varphi_c}^{\mathrm{w}}(Y_1) + \mathbb{E}\varphi(X_1)\log\cos\alpha,$$
$$h_\varphi^{\mathrm{w}}(X_2) = h_{\varphi_s}^{\mathrm{w}}(Y_2) + \mathbb{E}\varphi(X_2)\log\sin\alpha. \tag{6.10}$$

Using (6.8), we have the following inequality

$$h_\varphi^{\mathrm{w}}(X) \geq (\cos\alpha)^2\big[h_\varphi^{\mathrm{w}}(X_1) - \mathbb{E}\varphi(X_1)\log\cos\alpha\big]$$
$$+ (\sin\alpha)^2\big[h_\varphi^{\mathrm{w}}(X_2) - \mathbb{E}\varphi(X_2)\log\sin\alpha\big]. \tag{6.11}$$

Furthermore, recalling the definition of κ in (6.5) we obtain

$$h_\varphi^{\mathrm{w}}(X) \geq \frac{1}{2\kappa}\big[\mathbb{E}\varphi(X_1)\log\kappa\big]\exp\left(\frac{2h_\varphi^{\mathrm{w}}(X_1)}{\mathbb{E}\varphi(X_1)}\right) + \frac{1}{2\kappa}\big[\mathbb{E}\varphi(X_1)\log\kappa\big]\exp\left(\frac{2h_\varphi^{\mathrm{w}}(X_2)}{\mathbb{E}\varphi(X_2)}\right). \tag{6.12}$$

By virtue of assumption (6.7), we derive

$$h_\varphi^{\mathrm{w}}(X) \geq \frac{1}{2}\mathbb{E}\varphi(X)\log\kappa. \tag{6.13}$$

The definition of κ in (6.5) leads directly to the result. $\square$

Example 6.2. Let $d = 1$ and $X_1 \sim N(0, \sigma_1^2)$, $X_2 \sim N(0, \sigma_2^2)$. Then the WLSI (6.8) takes the following form

$$\log[2\pi(\sigma_1^2 + \sigma_2^2)]\mathbb{E}\varphi(X) + \frac{\log e}{\sigma_1^2 + \sigma_2^2}\mathbb{E}[X^2\varphi(X)]$$

$$\geq (\cos\alpha)^2\left[\log\left(\frac{2\pi\sigma_1^2}{(\cos\alpha)^2}\right)\right]\mathbb{E}\varphi(X_1) + \frac{(\cos\alpha)^2\log e}{\sigma_1^2}\mathbb{E}[X_1^2\varphi(X_1)]$$

$$+ (\sin\alpha)^2\left[\log\left(\frac{2\pi\sigma_2^2}{(\sin\alpha)^2}\right)\right]\mathbb{E}\varphi(X_2) + \frac{(\sin\alpha)^2\log e}{\sigma_2^2}\mathbb{E}[X_2^2\varphi(X_2)]. \quad (6.14)$$

Example 6.3. Let $d = 1$, $X = X_1 + X_2$, $X_1 \sim U[a_1, b_1]$ and $X_2 \sim U[a_2, b_2]$ be independent. Denote by $\Phi(x) = \int_0^x \varphi(u)\mathrm{d}u$ and $L_i = b_i - a_i$, $i = 1, 2$. The WDE $h_\varphi^w(X_i) = \frac{\Phi(b_i) - \Phi(a_i)}{L_i}\log L_i$. Then the inequality $\kappa \geq (\leq)1$ takes the form $L_1^2 + L_2^2 \geq (\leq)1$. Suppose for definiteness that $L_2 \geq L_1$ or, equivalently, $C_1 := a_2 + b_1 \leq a_1 + b_2 =: C_2$. Inequalities (6.7) take the form

$$L_2[\Phi(b_1) - \Phi(a_1)], \qquad L_1[\Phi(b_2) - \Phi(a_2)] \geq (\leq)\mathbb{E}\varphi(X). \quad (6.15)$$

The WLSI takes the form

$$-\Lambda + \log(L_1 L_2)\mathbb{E}\varphi(X) \geq (\cos\alpha)^2\frac{\Phi(b_1) - \Phi(a_1)}{L_1}\log\left(\frac{L_1}{\cos\alpha}\right)$$

$$+ (\sin\alpha)^2\frac{\Phi(b_2) - \Phi(a_2)}{L_2}\log\left(\frac{L_2}{\sin\alpha}\right), \quad (6.16)$$

where

$$\Lambda = \frac{\log L_1}{L_2}[\Phi(C_1) - \Phi(C_2)] + \frac{1}{L_1 L_2}\left[\int_A^{C_1}\varphi(x)(x - A)\log(x - A)\mathrm{d}x\right.$$

$$\left. + \int_{C_2}^B\varphi(x)(B - x)\log(B - x)\mathrm{d}x\right],$$

$$A = a_1 + a_2, \quad B = b_1 + b_2. \quad (6.17)$$

Finally, define $\Phi^*(x) = \int_0^x u\varphi(u)\mathrm{d}u$ and note that

$$\mathbb{E}\varphi(X) = \frac{1}{L_1 L_2}\left[\Phi^*(C_1) - \Phi^*(A) - \Phi^*(B) + \Phi^*(C_2)\right]$$

$$- A[\Phi(C_1) - \Phi(A)] + L_1[\Phi(C_2) - \Phi(C_1)] + B[\Phi(B) - \Phi(C_2)]. \quad (6.18)$$

7 The WLSI for the WF close to a constant

Proposition 7.1. *Let $d = 1$, $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2$ be independent and $X = X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. Suppose that WF $x \to \varphi(x)$ is twice continuously differentiable and*

$$|\varphi''(x)| \leq \epsilon\varphi(x), \qquad |\varphi(x) - \bar{\varphi}| \leq \epsilon, \quad (7.1)$$

where $\epsilon > 0$ and $\bar{\varphi} > 0$ are constants. Then there exists $\epsilon_0 > 0$ such that for any WF φ satisfying (7.1) with $0 < \epsilon < \epsilon_0$ WLSI holds true. Hence, checking of the WEPI is reduced to condition (6.7).

For a RV $\mathbf{Z}$, $\gamma > 0$ and independent Gaussian RV $\mathbf{N} \sim \mathrm{N}(\mathbf{0}, \mathbf{I}_d)$ define

$$M(\mathbf{Z}; \gamma) = \mathbb{E}\big[\big\|\mathbf{Z} - \mathbb{E}[\mathbf{Z}|\mathbf{Z}\sqrt{\gamma} + \mathbf{N}]\big\|^2\big], \tag{7.2}$$

where $\|.\|$ stands for the Euclidean norm. According to [24, 8] the differential entropy

$$h(\mathbf{Z}) = h(\mathbf{N}) + \frac{1}{2} \int_0^\infty \big[M(\mathbf{Z}; \gamma) - \mathbf{1}_{\{\gamma < 1\}}\big]\mathrm{d}\gamma. \tag{7.3}$$

For $\mathbf{Z} = \mathbf{Y}_1, \mathbf{Y}_2, \mathbf{X}_1 + \mathbf{X}_2$ assume the following conditions

$$\mathbb{E}\big[\big|\log f_{\mathbf{Z}}(\mathbf{Z})\big|\big] < \infty, \mathbb{E}\big[\|\mathbf{Z}\|^2\big] < \infty \tag{7.4}$$

and the uniform integrability: for independent $\mathbf{N}, \mathbf{N}' \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$ and any $\gamma > 0$ there exist an integrable function $\xi(\mathbf{Z}, \mathbf{N})$ such that

$$\left|\log \mathbb{E}\left[f_{\mathbf{Z}}\left(\mathbf{Z} + \frac{\mathbf{N} - \mathbf{N}'}{\sqrt{\gamma}}|\mathbf{Z}, \mathbf{N}\right)\right]\right| \leq \xi(\mathbf{Z}, \mathbf{N}). \tag{7.5}$$

Theorem 7.2. *Let $d = 1$ and assume conditions (7.4), (7.5). Let γ_0 be a point of continuity of $M(Z; \gamma)$, $Z = Y_1, Y_2, X_1 + X_2$. Suppose that there exists $\delta > 0$ such that*

$$M(X_1 + X_2; \gamma_0) \geq M(Y_1, \gamma_0)(\cos\alpha)^2 + M(Y_2; \gamma_0)(\sin\alpha)^2 + \delta. \tag{7.6}$$

Suppose that for some $\bar{\varphi} > 0$ the WF satisfies

$$\big|\varphi(x) - \bar{\varphi}\big| < \epsilon. \tag{7.7}$$

Then there exists $\epsilon_0 = \epsilon_0(\gamma_0, \delta, f_1, f_2)$ such that for any WF satisfying (7.7) with $\epsilon < \epsilon_0$ the WLSI holds true.

Proof. For a constant WF $\bar{\varphi}$, the following inequality is valid (see [8], Lemma 4.2 or [24], Eqns (9) and (10))

$$(\cos\alpha)^2 h_{\bar{\varphi}}^{\mathrm{w}}(Y_1) + (\sin\alpha)^2 h_{\bar{\varphi}}^{\mathrm{w}}(Y_2) \leq h_{\bar{\varphi}}^{\mathrm{w}}(Y_1\cos\alpha + Y_2\sin\alpha). \tag{7.8}$$

However, in view of Theorem 4.1 from [8], the representation (7.3) and inequality (7.6) imply under conditions (7.4) and (7.5) a stronger inequality

$$(\cos\alpha)^2 h_{\bar{\varphi}}^{\mathrm{w}}(Y_1) + (\sin\alpha)^2 h_{\bar{\varphi}}^{\mathrm{w}}(Y_2) + c_0\delta \leq h_{\bar{\varphi}}^{\mathrm{w}}(Y_1\cos\alpha + Y_2\sin\alpha). \tag{7.9}$$

Here $c_0 > 0$ and the term of order δ appears from integration in (7.3) in a neighbourhood of the continuity point γ_0. Define $\varphi^*(x) = |\varphi(x) - \bar{\varphi}|$. It is easy to check that

$$h_{\varphi^*}^{\mathrm{w}}(Z) < c_1\epsilon, \quad Z = X_1, X_2, X_1 + X_2. \tag{7.10}$$

From (7.9) and (7.10) we obtain that for ϵ small enough

$$(\cos\alpha)^2 h_{\varphi}^{\mathrm{w}}(Y_1) + (\sin\alpha)^2 h_{\varphi}^{\mathrm{w}}(Y_2) \leq h_{\varphi}^{\mathrm{w}}(Y_1\cos\alpha + Y_2\sin\alpha), \tag{7.11}$$

i.e., the WLSI holds true. $\qquad\square$

As an example, consider the case where RVs X_1, X_2 are normal and WF $\varphi \in C^2$.

Proposition 7.3. *Let RVs $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2$ be independent, and $X = X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. Suppose that WF $x \in \mathbb{R} \to \varphi(x) \geq 0$ is twice contiuously differentiable and slowly varying in the sense that $\forall x$,*

$$\left|\varphi''(x)\right| \leq \epsilon \varphi(x), \qquad \left|\varphi(x) - \bar\varphi\right| < \epsilon, \tag{7.12}$$

where $\epsilon > 0$ and $\bar\varphi > 0$ are constants. Then there exists $\epsilon_0 = \epsilon_0(\mu_0, \mu_1, \sigma_0^2, \sigma_2^2) > 0$ such that for any $0 < \epsilon \leq \epsilon_0$, the WLSI (6.8) with the WF φ holds true.

Proof. Let α be as in (6.6); to check (6.8), we use Stein's formula: for $Z \sim N(0, \sigma^2)$

$$\mathbb{E}\left[Z^2 \varphi(Z)\right] = \sigma^2 \mathbb{E}\left[\varphi(Z)\right] + \sigma^4 \mathbb{E}\left[\varphi''(Z)\right]. \tag{7.13}$$

Owing the inequality $|\varphi(x) - \bar\varphi| < \epsilon$ we have

$$\begin{aligned} \alpha < \alpha_0 = \tan^{-1}\Big(&\exp\big((\bar\varphi + \epsilon)^2\big[h_+(X_2) - h_-(X_1)\big] \\ &- (\bar\varphi - \epsilon)^2\big[h_+(X_1) - h_-(X_2)\big]\big)\Big). \end{aligned} \tag{7.14}$$

Here

$$h_\pm(X_i) = -\mathbb{E}\big[\mathbf{1}\big(X_i \in A_\pm^i\big) \log f_{X_i}^{No}(X_i)\big], \quad i = 1, 2. \tag{7.15}$$

and

$$A_+^i = \big\{x \in \mathbf{R} : f_i^{No}(x) < 1\big\}, \qquad A_-^i = \big\{x \in \mathbf{R} : f_i^{No}(x) > 1\big\}, \quad i = 1, 2. \tag{7.16}$$

Evidently, under conditions $|\varphi'(x)|, |\varphi''(x)| < \epsilon \varphi(x)$ we have that $\alpha_0 < \frac{\pi}{2} - \epsilon$ and $0 < \epsilon < (\sin\alpha)^2$, $(\cos\alpha)^2 < 1 - \epsilon < 1$. We claim that inequality (6.14) is satisfied with φ replaced by $\bar\varphi$ and added $\delta > 0$:

$$\begin{aligned} \log\big[2\pi\,(\sigma_1^2 + \sigma_2^2)\big]\bar\varphi &+ \frac{\log e}{\sigma_1^2 + \sigma_2^2}\mathbb{E}\big[X^2\big]\bar\varphi \\ \geq (\cos\alpha)^2&\left[\log\left(\frac{2\pi\sigma_1^2}{(\cos\alpha)^2}\right)\right]\bar\varphi + \frac{(\cos\alpha)^2\log e}{\sigma_1^2}\mathbb{E}\big[X_1^2\big]\bar\varphi \\ + (\sin\alpha)^2&\left[\log\left(\frac{2\pi\sigma_2^2}{(\sin\alpha)^2}\right)\right]\bar\varphi + \frac{(\sin\alpha)^2\log e}{\sigma_2^2}\mathbb{E}\big[X_2^2\big]\bar\varphi + \delta. \end{aligned} \tag{7.17}$$

Here $\delta > 0$ is calculated through ϵ and increases to a limit $\delta_0 > 0$ as $\epsilon \downarrow 0$. Indeed, strict concavity of $\log y$ for $y \in [0, \frac{2\pi\sigma_1^2}{(\cos\alpha)^2} \vee \frac{2\pi\sigma_2^2}{(\sin\alpha)^2}]$ implies that

$$\begin{aligned} \log\big[2\pi\,(\sigma_1^2 + \sigma_2^2)\big] \geq (\cos\alpha)^2&\left[\log\left(\frac{2\pi\sigma_1^2}{(\cos\alpha)^2}\right)\right] \\ + (\sin\alpha)^2&\left[\log\left(\frac{2\pi\sigma_2^2}{(\sin\alpha)^2}\right)\right] + \delta. \end{aligned} \tag{7.18}$$

On the other hand,

$$\frac{1}{\sigma_1^2 + \sigma_2^2} \bar{\varphi} \mathbb{E}[X^2] = \frac{(\cos \alpha)^2}{\sigma_1^2} \bar{\varphi} \mathbb{E}[X_1^2] + \frac{(\sin \alpha)^2}{\sigma_2^2} \bar{\varphi} \mathbb{E}[X_2^2]. \qquad (7.19)$$

Combining (7.18) and (7.19) one gets (7.17). Now, to check (6.8) with WF φ, in view of (7.17) it suffices to verify

$$\log\left[2\pi\left(\sigma_1^2 + \sigma_2^2\right)\right]\left[\mathbb{E}\varphi(X) - \bar{\varphi}\right] + \frac{\log e}{\sigma_1^2 + \sigma_2^2}\mathbb{E}\left[X^2\left(\varphi(X) - \bar{\varphi}\right)\right]$$

$$- (\cos \alpha)^2\left[\log\left(\frac{2\pi\sigma_1^2}{(\cos \alpha)^2}\right)\right]\left[\mathbb{E}\varphi(X_1) - \bar{\varphi}\right] + \frac{(\cos \alpha)^2 \log e}{\sigma_1^2}\mathbb{E}\left[X_1^2\left(\varphi(X_1) - \bar{\varphi}\right)\right]$$

$$- (\sin \alpha)^2\left[\log\left(\frac{2\pi\sigma_2^2}{(\sin \alpha)^2}\right)\right]\left[\mathbb{E}\varphi(X_2) - \bar{\varphi}\right) + \frac{(\sin \alpha)^2 \log e}{\sigma_2^2}\mathbb{E}\left[X_2^2\left(\varphi(X_2) - \bar{\varphi}\right)\right] < \delta.$$

$$(7.20)$$

We check (7.20) by a brute force, claiming that each term in (7.20) has the absolute value $< \delta/6$ when ϵ is small enough. For the terms containing $\mathbb{E}[\varphi(Z) - \bar{\varphi}]$, $Z = X, X_1 X_2$, this follows since $|\varphi - \bar{\varphi}| < \epsilon$. For the terms containing factor $\mathbb{E}[Z^2(\varphi(Z) - \bar{\varphi})]$, we use Stein's formula (7.13) and the condition that $|\varphi''(x)| \leq \epsilon \varphi(x)$. $\qquad \square$

Similar assertions can be established for other examples of PDFs $f_1(x)$ and $f_2(x)$, i.e. uniform, exponential, Gamma, Cauchy, etc.

8 A weighted Fisher information inequality

Let $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$ be a pair of independent RVs $\mathbf{X}$ and $\mathbf{Y} \in \mathbb{R}^d$, with sample values $\mathbf{z} = (\mathbf{x}, \mathbf{y}) \in \mathbb{R}^d \times \mathbb{R}^d$ and marginal PDFs $f_1(\mathbf{x}, \underline{\theta})$, $f_2(\mathbf{y}, \underline{\theta})$, respectively. Let $f_{\mathbf{Z}|\mathbf{X}+\mathbf{Y}}(\mathbf{x}, \mathbf{y}|\mathbf{u})$ stand for the conditional PDF as

$$f_{\mathbf{Z}|\mathbf{X}+\mathbf{Y}}(\mathbf{x}, \mathbf{y}|\mathbf{u}) = \frac{f_1(\mathbf{x}) f_2(\mathbf{y}) \mathbf{1}(\mathbf{x} + \mathbf{y} = \mathbf{u})}{\int_{\mathbb{R}^n} f_1(\mathbf{v}) f_2(\mathbf{u} - \mathbf{v}) d\mathbf{v}}. \qquad (8.1)$$

Given a WF $\mathbf{z} = (\mathbf{x}, \mathbf{y}) \in \mathbb{R}^d \times \mathbb{R}^d \mapsto \varphi(\mathbf{z}) \geq 0$, we employ the following reduced WFs:

$$\varphi(\mathbf{u}) = \int \varphi(\mathbf{v}, \mathbf{u} - \mathbf{v}) f_{\mathbf{Z}|\mathbf{X}+\mathbf{Y}}(\mathbf{v}, \mathbf{u} - \mathbf{v}) d\mathbf{v},$$

$$\varphi_1(\mathbf{x}) = \int \varphi(\mathbf{x} + \mathbf{y}, \mathbf{y}) f_2(\mathbf{y}) d\mathbf{y}, \quad \varphi_2(\mathbf{y}) = \int \varphi(\mathbf{x}, \mathbf{x} + \mathbf{y}) f_1(\mathbf{x}) d\mathbf{x}. \qquad (8.2)$$

Next, let us introduce the matrices M_φ and G_φ:

$$M_\varphi = \int \varphi(\mathbf{x}, \mathbf{y}) f_1(\mathbf{x}) f_2(\mathbf{y}) \left(\frac{\partial \log f_1(\mathbf{x})}{\partial \underline{\theta}}\right)^{\mathrm{T}} \left(\frac{\partial \log f_2(\mathbf{x})}{\partial \underline{\theta}}\right) \mathbf{1}(f_1(\mathbf{x}) f_2(\mathbf{y}) > 0) d\mathbf{x} d\mathbf{y},$$

$$G_\varphi = \left(J_{\varphi_1}^{\mathrm{w}}(\mathbf{X})\right)^{-1} M_\varphi \left(J_{\varphi_2}^{\mathrm{w}}(\mathbf{Y})\right)^{-1}. \qquad (8.3)$$

Note that for $\varphi \equiv 1$ we have $M_\varphi = G_\varphi = 0$ and the classical Fisher information inequality emerges (cf. [27]). Finally, we define

$$\Xi := \Xi_{\varphi_1,\varphi_2}(\mathbf{X},\mathbf{Y}) = M_\varphi J_{\varphi_1}^{\mathrm{w}}(\mathbf{X}) G_\varphi (\mathbf{I} - M_\varphi G_\varphi)^{-1} M_\varphi \big[G_\varphi J_{\varphi_2}^{\mathrm{w}}(\mathbf{Y}) G_\varphi - J_{\varphi_1}^{\mathrm{w}}(\mathbf{X}) \big]$$
$$+ G_\varphi (\mathbf{I} - M_\varphi G_\varphi)^{-1} M_\varphi G_\varphi J_{\varphi_2}^{\mathrm{w}}(\mathbf{Y}) \big[M_\varphi^{-1} - G_\varphi \big] - G_\varphi J_{\varphi_2}^{\mathrm{w}}(\mathbf{Y}) G_\varphi - G_\varphi. \quad (8.4)$$

Theorem 8.1 (A weighted Fisher information inequality (WFII)). *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be independent RVs. Assume that* $f_{\mathbf{X}}^{(1)} = \frac{\partial}{\partial \underline{\theta}} f_1$ *is not a multiple of* $f_{\mathbf{Y}}^{(1)} = \frac{\partial}{\partial \underline{\theta}} f_2$. *Then*

$$J_\varphi^{\mathrm{w}}(\mathbf{X}+\mathbf{Y}) \le (\mathbf{I} - M_\varphi G_\varphi) \big[\big(J_{\varphi_1}^{\mathrm{w}}(\mathbf{X})\big)^{-1} + \big(J_{\varphi_2}^{\mathrm{w}}(\mathbf{Y})\big)^{-1} - \Xi_{\varphi_1,\varphi_2}(\mathbf{X},\mathbf{Y}) \big]^{-1}. \quad (8.5)$$

Proof. We use the same methodology as in Theorem 1 from [27]. Recalling Corollary 4.8, (iii) in [20] substitute $\mathrm{P} := [1,1]$. Therefore for $\mathbf{Z} = (\mathbf{X},\mathbf{Y})$, $J^{\mathrm{w}}(\mathbf{Z})$ is an $m \times m$ matrix

$$\big(J_\theta^{\mathrm{w}}(\mathbf{Z})\big)^{-1} = \begin{pmatrix} J_{\varphi_1}^{\mathrm{w}}(\mathbf{X}) & M_\varphi \\ M_\varphi & J_{\varphi_2}^{\mathrm{w}}(\mathbf{Y}) \end{pmatrix}^{-1}. \quad (8.6)$$

Next, we need the following well-known expression for the inverse of a block matrix

$$\begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{21} \\ \mathbf{C}_{12} & \mathbf{C}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{C}_{11}^{-1} + \mathbf{D}_{12}\mathbf{D}_{22}^{-1}\mathbf{D}_{21} & -\mathbf{D}_{12}\mathbf{D}_{22}^{-1} \\ -\mathbf{D}_{22}^{-1}\mathbf{D}_{21} & \mathbf{D}_{22}^{-1} \end{pmatrix}, \quad (8.7)$$

where

$$\mathbf{D}_{22} = \mathbf{C}_{22} - \mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{C}_{12}, \qquad \mathbf{D}_{12} = \mathbf{C}_{11}^{-1}\mathbf{C}_{12}, \qquad \mathbf{D}_{21} = \mathbf{C}_{21}\mathbf{C}_{11}^{-1},$$
$$\mathbf{C}_{11} = J_{\varphi_1}^{\mathrm{w}}(\mathbf{X}), \qquad \mathbf{C}_{22} = J_{\varphi_2}^{\mathrm{w}}(\mathbf{Y}), \quad \text{and} \quad \mathbf{C}_{12} = \mathbf{C}_{21} = M_\varphi.$$

By using the Schwarz inequality, we derive

$$M_\varphi^2 \le J_{\varphi_1}^{\mathrm{w}}(\mathbf{X})\, J_{\varphi_2}^{\mathrm{w}}(\mathbf{Y}), \quad \text{or } M_\varphi\, G_\varphi \le \mathbf{I}, \quad (8.8)$$

with equality iff $f_{\mathbf{X}}^{(1)}(x) = \frac{\partial \log f_1(x)}{\partial \underline{\theta}} \propto \frac{\partial \log f_2(y)}{\partial \underline{\theta}} = f_{\mathbf{Y}}^{(1)}(y)$.

Define

$$\delta := J_{\varphi_2}^{\mathrm{w}}(\mathbf{Y}) - \tilde{M}_\varphi \big(J_{\varphi_1}^{\mathrm{w}}(\mathbf{X})\big)^{-1} M_\varphi = (\mathbf{I} - M_\varphi\, G_\varphi) J_{\varphi_2}^{\mathrm{w}}(\mathbf{Y})$$
$$\Rightarrow \delta^{-1} = \big(J_{\theta_2}^{\mathrm{w}}(\mathbf{Y})\big)^{-1} (\mathbf{I} - M_\varphi\, G_\varphi)^{-1}. \quad (8.9)$$

Thus, owing to the (8.7), particularly for $\mathrm{P} = [1,1]$, we can write

$$\mathrm{P}\big(J_\varphi^{\mathrm{w}}(\mathbf{Z})\big)^{-1}\mathrm{P}^T = \big(J_{\varphi_1}^{\mathrm{w}}(\mathbf{X})\big)^{-1} + \big(J_{\varphi_1}^{\mathrm{w}}(\mathbf{X})\big)^{-1} M_\varphi\, \delta^{-1}\, M_\varphi \big(J_{\varphi_1}^{\mathrm{w}}(\mathbf{X})\big)^{-1}$$
$$- \big(J_{\varphi_1}^{\mathrm{w}}(\mathbf{X})\big)^{-1} M_\varphi\, \delta^{-1} - \delta^{-1}\, M_\varphi\, \big(J_{\varphi_1}^{\mathrm{w}}(\mathbf{X})\big)^{-1} + \delta^{-1}. \quad (8.10)$$

Substituting (8.9), in above expression, we have

$$\mathrm{P}\big(J_\varphi^{\mathrm{w}}(\mathbf{Z})\big)^{-1}\mathrm{P}^T$$
$$= \big(J_{\varphi_1}^{\mathrm{w}}(\mathbf{X})\big)^{-1} + G_\varphi (\mathbf{I} - M_\varphi\, G_\varphi)^{-1} M_\varphi \big(J_{\varphi_1}^{\mathrm{w}}(\mathbf{X})\big)^{-1}$$
$$- G_\varphi (\mathbf{I} - M_\varphi\, G_\varphi)^{-1} - \big(J_{\varphi_2}^{\mathrm{w}}(\mathbf{Y})\big)^{-1} (\mathbf{I} - M_\varphi\, G_\varphi)^{-1} M_\varphi \big(J_{\varphi_1}^{\mathrm{w}}(\mathbf{X})\big)^{-1}$$

$$+ \left(\mathsf{J}_{\varphi_2}^{\mathrm{w}}(\mathbf{Y})\right)^{-1}(\mathbf{I} - M_\varphi\ G_\varphi)^{-1}$$
$$= \Big\{\left(\mathsf{J}_{\varphi_1}^{\mathrm{w}}(\mathbf{X})\right)^{-1}(\mathbf{I} - M_\varphi\ G_\varphi)$$
$$\quad + G_\varphi(\mathbf{I} - M_\varphi\ G_\varphi)^{-1}M_\varphi\left(\mathsf{J}_{\varphi_1}^{\mathrm{w}}(\mathbf{X})\right)^{-1}(\mathbf{I} - M_\varphi\ G_\varphi) - G_\varphi$$
$$\quad - \left(\mathsf{J}_{\theta_2}^{\mathrm{w}}(\mathbf{Y})\right)^{-1}(\mathbf{I} - M_\varphi\ G_\varphi)^{-1}\tilde{M}_\varphi\left(\mathsf{J}_{\varphi_1}^{\mathrm{w}}(\mathbf{X})\right)^{-1}(\mathrm{I} - M_\varphi\ G_\varphi)$$
$$\quad + \left(\mathsf{J}_{\varphi_2}^{\mathrm{w}}(\mathbf{Y})\right)^{-1}\Big\}(\mathbf{I} - M_\varphi\ G_\varphi)^{-1}. \tag{8.11}$$

Consequently by simplifying (8.11), one yields

$$\mathsf{P}\left(\mathsf{J}_\varphi^{\mathrm{w}}(\mathbf{Z})\right)^{-1}\mathsf{P}^T$$
$$= \Big\{\left(\mathsf{J}_{\varphi_1}^{\mathrm{w}}(\mathbf{X})\right)^{-1} + \left(\mathsf{J}_{\varphi_2}^{\mathrm{w}}(\mathbf{Y})\right)^{-1} + \varXi_{\varphi_1,\varphi_2}(\mathbf{X},\mathbf{Y})\Big\}(\mathbf{I} - M_\varphi\ G_\varphi)^{-1}. \tag{8.12}$$

By using Corollary 3.4, (iii) from [20], we obtain the property claimed in (8.5):

$$\mathsf{J}_\varphi^{\mathrm{w}}(\mathbf{X} + \mathbf{Y}) \le \Big\{\big[\left(\mathsf{J}_{\varphi_1}^{\mathrm{w}}(\mathbf{X})\right)^{-1} + \left(\mathsf{J}_{\varphi_2}^{\mathrm{w}}(\mathbf{Y})\right)^{-1} + \varXi_{\varphi_1,\varphi_2}(\mathbf{X},\mathbf{Y})\big](\mathbf{I} - M_\varphi\ G_\varphi)^{-1}\Big\}^{-1}.$$

This concludes the proof. $\qquad\square$

Proposition 8.2. *Consider additive RV* $\mathbf{Z} = \mathbf{X} + \mathbf{N}_\Sigma$, *such that* $\mathbf{N}_\Sigma \sim \mathrm{N}(\mathbf{0}, \Sigma)$ *and* $\mathbf{N}_\Sigma$ *is independent of* $\mathbf{X}$. *Introduce matrices*

$$V_\varphi(\mathbf{X}|\mathbf{Z}) = \mathbb{E}\big[\varphi\ \big(\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Z}]\big)^{\mathsf{T}}\big(\mathbf{X} - \mathbb{E}(\mathbf{X}|\mathbf{Z})\big)\big],$$
$$E_\varphi = \mathbb{E}\big[\varphi\ \big(\mathbf{Z} - \mathbb{E}[\mathbf{X}|\mathbf{Z}]\big)^{\mathsf{T}}\big(\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Z}]\big)\big], \quad \overline{E}_\varphi = E_\varphi + E_\varphi^{\mathsf{T}}. \tag{8.13}$$

The WFIM of RV $\mathbf{Z}$ *can be written as*

$$\mathsf{J}_\varphi^{\mathrm{w}}(\mathbf{Z}) = \left(\Sigma^{-1}\right)^{\mathsf{T}}\big\{\mathbb{E}\big[\varphi\ \mathbf{N}_\Sigma^{\mathsf{T}}\mathbf{N}_\Sigma\big] + \overline{E}_\varphi - V_\varphi(\mathbf{X}|\mathbf{Z})\big\}\Sigma^{-1}. \tag{8.14}$$

9   The weighted entropy power is a concave function

Let $\mathbf{Z} = \mathbf{X} + \mathbf{Y}$ and $\mathbf{Y} \sim \mathrm{N}(\mathbf{0}, \sqrt{\gamma}\mathbf{I}_d)$. In the literature, several elegant proofs, employing the Fisher information inequality or basic properties of mutual information, have been proposed in order to prove that the entropy power (EP) is a concave function of γ [2, 25]. We are interested in the *weighted entropy power* (WEP) defined as follows:

$$\mathrm{N}_\varphi^{\mathrm{w}}(\mathbf{Z}) := \mathrm{N}_\varphi^{\mathrm{w}}(f_{\mathbf{Z}}) = \exp\left\{\frac{2\ h_\varphi^{\mathrm{w}}(\mathbf{Z})}{d\ \mathbb{E}[\varphi(\mathbf{Z})]}\right\}. \tag{9.1}$$

Compute the second derivative of the WEP

$$\frac{\mathrm{d}^2}{\mathrm{d}\gamma^2}\exp\left\{\frac{2}{d}\frac{h_\varphi^{\mathrm{w}}(\mathbf{Z})}{\mathbb{E}[\varphi(\mathbf{Z})]}\right\} = \exp\left\{\frac{2}{d}\frac{h_\varphi^{\mathrm{w}}(\mathbf{Z})}{\mathbb{E}[\varphi(\mathbf{Z})]}\right\}$$
$$\times \left[\left(\frac{2}{d}\frac{\mathrm{d}}{\mathrm{d}\gamma}\frac{h_\varphi^{\mathrm{w}}(\mathbf{Z})}{\mathbb{E}[\varphi(\mathbf{Z})]}\right)^2 + \left(\frac{2}{d}\frac{\mathrm{d}^2}{\mathrm{d}\gamma^2}\frac{h_\varphi^{\mathrm{w}}(\mathbf{Z})}{\mathbb{E}[\varphi(\mathbf{Z})]}\right)\right]$$
$$= \exp\left\{\frac{2}{d}\frac{h_\varphi^{\mathrm{w}}(\mathbf{Z})}{\mathbb{E}[\varphi(\mathbf{Z})]}\right\}\left[\left(\varLambda(\gamma)\right)^2 + \frac{\mathrm{d}}{\mathrm{d}\gamma}\varLambda(\gamma)\right], \tag{9.2}$$

where

$$\Lambda(\gamma) = \frac{2}{d} \frac{\mathrm{d}}{\mathrm{d}\gamma} \frac{h_\varphi^{\mathrm{w}}(\mathbf{Z})}{\mathbb{E}[\varphi(\mathbf{Z})]}. \tag{9.3}$$

In view of (9.2) the concavity of the WEP is equivalent to the inequality

$$\frac{\mathrm{d}}{\mathrm{d}\gamma}\big(\Lambda(\gamma)\big)^{-1} \geq 1. \tag{9.4}$$

In the spirit of the WEP, we shall present a new proof of concavity of EP. Regarding this, let us apply the WFII (8.5) to $\varphi \equiv 1$. Then a straightforward computation gives

$$\frac{\mathrm{d}}{\mathrm{d}\gamma} \frac{d}{\operatorname{tr} J(\mathbf{Z})} \geq 1. \tag{9.5}$$

Theorem 9.1 (A weighted De Bruijn's identity). *Let* $\mathbf{X} \sim f_\mathbf{X}$ *be a RV in* $\mathbb{R}^n$, *with a PDF* $f_\mathbf{X} \in C^2$. *For a standard Gaussian RV* $\mathbf{N} \sim \mathrm{N}(\mathbf{0}, \mathbf{I}_d)$ *independent of* $\mathbf{X}$, *and given* $\gamma > 0$, *define the RV* $\mathbf{Z} = \mathbf{X} + \sqrt{\gamma}\mathbf{N}$ *with PDF* $f_\mathbf{Z}$. *Let* $\mathbf{V}_r$ *be the d-sphere of radius r centered at the origin and having surface denoted by* $\mathbf{S}_r$. *Assume that for given WF* φ *and* $\forall \gamma \in (0, 1)$ *the relations*

$$\int f_\mathbf{Z}(\mathbf{x}) |\ln f_\mathbf{Z}(\mathbf{x})| \mathrm{d}\mathbf{x} < \infty, \qquad \int |\nabla f_\mathbf{Z}(\mathbf{y}) \ln f_\mathbf{Z}(\mathbf{y})| \mathrm{d}\mathbf{y} < \infty \tag{9.6}$$

and

$$\lim_{r\to\infty} \int_{\mathbb{S}_r} \varphi(\mathbf{y}) \log f_Z(\mathbf{y})\big(\nabla f_Z(\mathbf{y})\big)\mathrm{d}S_r = 0 \tag{9.7}$$

are fulfilled. Then

$$\frac{\mathrm{d}}{\mathrm{d}\gamma} h_\varphi^{\mathrm{w}}(\mathbf{Z}) = \frac{1}{2} \operatorname{tr} \mathsf{J}_\varphi^{\mathrm{w}}(\mathbf{Z}) - \frac{1}{2}\mathbb{E}\left[\varphi \frac{\Delta f_Z(\mathbf{Z})}{f_Z(\mathbf{Z})}\right] + \frac{\mathcal{R}(\gamma)}{2}. \tag{9.8}$$

Here

$$\mathcal{R}(\gamma) = \mathbb{E}\big[\nabla\varphi \, \log f_\mathbf{Z}(\mathbf{Z})\big(\nabla \log f_\mathbf{Z}(\mathbf{Z})\big)^{\mathrm{T}}\big]. \tag{9.9}$$

If we assume that $\varphi \equiv 1$, *then the equality (9.8) directly implies (9.4). Hence, the standard entropy power is a concave function of* γ.

Next, we establish the concavity of the WEP when the WF is close to a constant.

Theorem 9.2. *Assume conditions (9.6) and (9.7) and suppose that* $\forall \gamma \in (0, 1)$

$$\frac{\mathrm{d}}{\mathrm{d}\gamma} \frac{d}{\operatorname{tr} J(\mathbf{Z})} \geq 1 + \epsilon. \tag{9.10}$$

Then $\exists \delta = \delta(\epsilon)$ *such that any WF* φ *for which* $\exists \bar{\varphi} > 0$: $|\varphi - \bar{\varphi}| < \delta$, $|\nabla \varphi| < \delta$ *the WEP (9.1) is a concave function of* γ. *Under the milder assumption*

$$\frac{\mathrm{d}}{\mathrm{d}\gamma} \frac{d}{\operatorname{tr} J(\mathbf{Z})}\bigg|_{\gamma=0} \geq 1 + \epsilon, \tag{9.11}$$

the WEP is a concave function of γ *in a small neighbourhood of* $\gamma = 0$.

Proof. It is sufficient to check that

$$\frac{\mathrm{d}}{\mathrm{d}\gamma}\psi(\gamma) \geq 1 \quad \text{where } \psi(\gamma) = \left(\frac{2}{d}\frac{\mathrm{d}}{\mathrm{d}\gamma}\frac{h_\varphi^{\mathrm{w}}(\mathbf{Z})}{\mathbb{E}[\varphi(\mathbf{Z})]}]\right)^{-1} = \Lambda(\gamma)^{-1}. \tag{9.12}$$

By a straightforward calculation

$$\psi(\mathbf{Z}) = d\left(\mathbb{E}[\varphi(\mathbf{Z})]\right)^2\left[\frac{\mathrm{d}}{\mathrm{d}\gamma}h_\varphi^{\mathrm{w}}(\mathbf{Z})\mathbb{E}[\varphi(\mathbf{Z})] - h_\varphi^{\mathrm{w}}(\mathbf{Z})\frac{\mathrm{d}}{\mathrm{d}\gamma}\mathbb{E}\varphi(\mathbf{Z})\right]^{-1},$$

$$\frac{\mathrm{d}}{\mathrm{d}\gamma}h_\varphi^{\mathrm{w}}(\mathbf{Z}) = \frac{1}{2}\mathrm{tr}J_\varphi^w(\mathbf{Y}) - \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}\gamma}\mathbb{E}[\varphi(\mathbf{Y})] + \frac{1}{2}\mathcal{R}(\gamma). \tag{9.13}$$

These formulas imply

$$\psi(\gamma) = \frac{d}{\mathrm{tr}\,J_\varphi(\mathbf{Z})} + o(\delta).$$

$$\text{as } 1 - \delta < \mathbb{E}[\varphi(\mathbf{Z})] < 1 + \delta, \quad \left|\mathrm{tr}\,J_\varphi^{\mathrm{w}}(\mathbf{Z}) - \mathrm{tr}\,J(\mathbf{Z})\right| < \delta\,\mathrm{tr}\,J(\mathbf{Z}). \tag{9.14}$$

Next,

$$\frac{\mathrm{d}}{\mathrm{d}\gamma}\mathbb{E}[\varphi(\mathbf{Z})] = \frac{1}{2}\int\varphi(y)\Delta f_\mathbf{Z}(\mathbf{y})\mathrm{d}\mathbf{y} \tag{9.15}$$

and using the Stokes formula one can bound this term by δ. Finally, $|\mathcal{R}(\gamma)| \leq \delta$ in view of (9.7), which leads to the claimed result. $\qquad\square$

10   Rates of weighted entropy and information

This section follows [18]. The concept of a *rate* of the WE or WDE emerges when we work with outcomes in a context of a discrete-time random process (RP):

$$h_{\varphi_n}^{\mathrm{w}}(\mathbf{p}_n) = -\mathbb{E}\varphi_n(\mathbf{X}_0^{n-1})\log\mathbf{p}_n(\mathbf{X}_0^{n-1}) := \mathbb{E}I_{\varphi_n}^{\mathrm{w}}(\mathbf{X}_0^{n-1}). \tag{10.1}$$

Here the WF φ_n is made dependent on n: two immediate cases are where (a) $\varphi_n(\mathbf{x}_1^n) = \sum_{j=0}^n\psi(x_j)$ and (b) $\varphi_n(\mathbf{x}_1^n) = \prod_{j=0}^n\psi(x_j)$ (an additive and multiplicative WF, respectively). Next, $\mathbf{X}_0^{n-1} = (X_0, \ldots, X_{n-1})$ is a random *string* generated by an RP. For simplicity, let us focus on RPs taking values in a finite set $\mathcal{X}$. Symbol $\mathbb{P}$ stands for the probability measure of $\mathbf{X}$, and $\mathbb{E}$ denotes the expectation under $\mathbb{P}$. For an RP with IID values, the joint probability of a sample $\mathbf{x}_0^{n-1} = (x_0, \ldots, x_{n-1})$ is $\mathbf{p}_n(\mathbf{x}_0^{n-1}) = \prod_{j=0}^{n-1}p(x_j)$, $p(x) = \mathbb{P}(X_j = x)$ being the probability of an individual outcome $x \in \mathcal{X}$. In the case of a Markov chain, $\mathbf{p}_n(\mathbf{x}_0^{n-1}) = \lambda(x_0)\prod_{j=1}^n p(x_{j-1}, x_j)$. Here $\lambda(x)$ gives an initial distribution and $p(x, y)$ is the transition probability on $\mathcal{X}$; to reflect this fact, we will sometimes use the notation $h_{\varphi_n}^{\mathrm{w}}(\mathbf{p}_n, \lambda)$. The quantity

$$I_{\varphi_n}^{\mathrm{w}}(\mathbf{x}_0^{n-1}) := -\varphi_n(\mathbf{x}_0^{n-1})\log\mathbf{p}_n(\mathbf{x}_0^{n-1})$$

is interpreted as a *weighted information* (WI) contained in/conveyed by outcome $\mathbf{x}_0^{n-1}$.

In the IID case, the WI and WE admit the following representations. Define $S(p) = -\mathbb{E}[\log p(X)]$ and $H_\psi^{\mathrm{W}} = -\mathbb{E}[\psi(X)\log p(X)]$ to be the SE and the WE, of the one-digit distribution (the capital letter is used to make it distinct from $h_{\varphi_n}^{\mathrm{W}}$, the multi-time WE).

(A) For an additive WF:

$$I_{\varphi_n}^{\mathrm{W}}\left(\mathbf{x}_0^{n-1}\right) = -\sum_{j=0}^{n-1}\psi(x_j)\sum_{l=0}^{n-1}\log p(x_l) \tag{10.2}$$

and

$$h_{\varphi_n}^{\mathrm{W}}(\mathbf{p}_n) = n(n-1)S(p)\mathbb{E}\left[\psi(X)\right] + nH_\psi^{\mathrm{W}}(p) := n(n-1)\mathrm{A}_0 + n\mathrm{A}_1. \tag{10.3}$$

(B) For a multiplicative WF:

$$I_{\varphi_n}^{\mathrm{W}}\left(\mathbf{x}_0^{n-1}\right) = -\prod_{j=0}^{n-1}\psi(x_j)\sum_{l=0}^{n-1}\log p(x_l) \tag{10.4}$$

and

$$h_{\varphi_n}^{\mathrm{W}}(\mathbf{p}_n) = nH_\psi^{\mathrm{W}}(p)\left[\mathbb{E}\varphi(X)\right]^{n-1} := \mathrm{B}_0^{n-1} \times n\mathrm{B}_1. \tag{10.5}$$

The values A_0, B_0 and their analogs in a general situation are referred to as *primary* rates, and A_1, B_1 as *secondary* rates.

10.A   WI and WE rates for asymptotically additive WFs

Here we will deal with a stationary RP $\mathbf{X} = (X_j, j \in \mathbb{Z})$ and use the above notation $\mathbf{p}_n(\mathbf{x}_0^{n-1}) = \mathbb{P}(\mathbf{X}_0^{n-1} = \mathbf{x}_0^{n-1})$ for the joint probability. We will refer to the limit present in the Shannon–McMillan–Breiman (SMB) theorem (see, e.g., [1, 7]) taking place for an ergodic RP:

$$\lim_{n\to\infty}\left[-\frac{1}{n}\log \mathbf{p}_n\left(\mathbf{X}_0^{n-1}\right)\right] = -\mathbb{E}\log\mathbb{P}\left(X_0|\mathbf{X}_{-\infty}^{-1}\right) := S, \quad \mathbb{P}\text{-a.s.} \tag{10.6}$$

Here $\mathbb{P}(y|\mathbf{x}_{-\infty}^{-1})$ is the conditional PM/DF for $X_0 = y$ given $\mathbf{x}_{-\infty}^{-1}$, an infinite past realization of $\mathbf{X}$. An assumption upon WFs φ_n called *asymptotic additivity* (AA) is that

$$\lim_{n\to\infty}\frac{1}{n}\varphi_n\left(\mathbf{X}_0^{n-1}\right) = \alpha, \quad \mathbb{P}\text{-a.s. and/or in } L_2(\mathbb{P}). \tag{10.7}$$

Eqns (10.6), (10.7) lead to the identification of the primary rate: $\mathrm{A}_0 = \alpha S$.

Theorem 10.1. *Given an ergodic RP* $\mathbf{X}$*, consider the WI* $I_{\varphi_n}^{\mathrm{W}}(\mathbf{X}_0^{n-1})$ *and the WE* $H_{\varphi_n}^{\mathrm{W}}(\mathbf{p}_n)$ *as defined in* (10.2), (10.3). *Suppose that convergence in* (10.7) *holds* $\mathbb{P}$*-a.s. Then:*

(I) *We have that*

$$\lim_{n\to\infty}\frac{I_{\varphi_n}^{\mathrm{W}}(\mathbf{X}_0^{n-1})}{n^2} = \alpha S, \quad \mathbb{P}\text{-a.s.} \tag{10.8}$$

(II) *Furthermore,*

(a) *suppose that the WFs φ_n exhibit convergence* (10.7), *$\mathbb{P}$-a.s., with a finite α, and $|\varphi_n(\mathbf{X}_0^{n-1})/n| \leq c$ where c is a constant independent of n. Suppose also that convergence in Eqn* (10.6) *holds true. Then we have that*

$$\lim_{n\to\infty} \frac{h_{\varphi_n}^{\mathrm{w}}(\mathbf{p}_n)}{n^2} = \alpha S. \tag{10.9}$$

(b) *Likewise, convergence in Eqn* (10.9) *holds true whenever convergences* (10.7) *and* (10.6) *hold $\mathbb{P}$-a.s. and $|\log \mathbf{p}_n(\mathbf{X}_0^{n-1})/n| \leq c$ where c is a constant.*

(c) *Finally, suppose that convergence in* (10.6) *and* (10.7) *holds in $\mathrm{L}_2(\mathbb{P})$, with finite α and S. Then again, convergence in* (10.9) *holds true.*

Example 10.2. Clearly, the condition of stationarity cannot be dropped. Indeed, let $\varphi_n(x_0^{n-1}) = \alpha n$ be an additive WF and $\mathbf{X}$ be a (non-stationary) Gaussian process with covariances $C = \{C_{ij}, i, j \in \mathbb{Z}_+^1\}$. Let f_n be a n-dimensional PDF of the vector $(X_1, \ldots, X_n)$. Then

$$h_{\varphi_n}^{\mathrm{w}}(f_n) = \frac{\alpha n}{2} \left[n \log(2\pi e) + \log\big(\det(C_n)\big) \right]. \tag{10.10}$$

Suppose that the eigenvalues $\lambda_1 \leq \cdots \leq \lambda_j \leq \cdots \leq \lambda_n$ of C_n have the order $\lambda_j \approx cj$. Then by Stirling's formula the second term in (10.10) dominates and the scaling of $h_{\varphi_n}^{\mathrm{w}}(f_n)$ is $(n^2 \log n)^{-1}$ instead of n^{-2} as $n \to \infty$.

Theorem 10.1. can be considered as an analog of the SMB theorem for the primary WE rate in the case of an AA WF. A specification of the secondary rate A_1 is given in Theorem 10.3 for an additive WF. The WE rates for multiplicative WFs are studied in Theorem 10.4 for the case where $\mathbf{X}$ is a stationary ergodic Markov chain on $\mathcal{X}$.

Theorem 10.3. *Suppose that $\varphi_n(\mathbf{x}_0^{n-1}) = \sum_{j=0}^{n-1} \psi(x_j)$. Let $\mathbf{X}$ be a stationary RP with the property that $\forall i \in \mathbb{Z}$ there exists the limit*

$$\lim_{n\to\infty} \sum_{j\in\mathbb{Z}: |j+i|\leq n} \mathbb{E}\big[\psi(X_0) \log p^{(n+i+j)}\big(X_j | \mathbf{X}_{-n-i}^{j-1}\big)\big]$$

$$= \sum_{j\in\mathbb{Z}} \mathbb{E}\big[\psi(X_0) \log p\big(X_j | \mathbf{X}_{-\infty}^{j-1}\big)\big] := -A_1 \tag{10.11}$$

and the last series converges absolutely. Then $\lim_{n\to\infty} \frac{1}{n} H_{\varphi_n}^{\mathrm{w}}(\mathbf{p}_n) = A_1$.

10.B WI and WE rates for asymptotically multiplicative WFs

The WI rate is given in Theorem 10.3. Here we use the condition of asymptotic multiplicativity:

$$\lim_{n\to\infty} \big[\varphi_n(\mathbf{X}_0^{n-1})\big]^{1/n} = \beta, \quad \mathbb{P}\text{-a.s.} \tag{10.12}$$

Theorem 10.4. *Given an ergodic RP* $\mathbf{X}$ *with a probability distribution* $\mathbb{P}$, *consider the WI* $I_{\varphi_n}^{\mathrm{w}}(\mathbf{x}_0^{n-1}) = -\varphi_n(\mathbf{x}_0^{n-1}) \log \mathbf{p}_n(\mathbf{x}_0^{n-1})$. *Suppose that convergence in* (10.12) *holds* $\mathbb{P}$-*a.s. Then the following limit holds true:*

$$\lim_{n \to \infty} \frac{1}{n} \log I_{\varphi_n}^{\mathrm{w}}(\mathbf{X}_0^{n-1}) = \beta, \quad \mathbb{P}\text{-}a.s.$$

Theorem 10.5. *Assume that* $\varphi(\mathbf{x}_0^{n-1}) = \prod_{j=0}^{n-1} \psi(x_j)$, *with* $\psi(x) > 0$, $x \in \mathcal{X}$. *Let* $\mathbf{X}$ *be a stationary Markov chain with transition probabilities* $p(x, y) > 0$, $x, y \in \mathcal{X}$. *Then, for all initial distribution* λ,

$$\lim_{n \to \infty} \frac{1}{n} \log h_{\varphi_n}^{\mathrm{w}}(\mathbf{p}_n, \lambda) = \mathrm{B}_0. \tag{10.13}$$

Here

$$\mathrm{B}_0 = \log \mu \tag{10.14}$$

and $\mu > 0$ *is the Perron–Frobenius eigenvalue of the matrix* $\mathrm{M} = (\psi(x)p(x, y))$ *coinciding with the norm of* M.

The secondary rate B_1 in this case is identified through the invariant probabilities $\pi(x)$ of the Markov chain and the Perron–Frobenius eigenvectors of matrices M and M^{T}.

Example 10.6. Consider a stationary sequence $X_{n+1} = \alpha X_n + Z_{n+1}, n \geq 0$, where $Z_{n+1} \sim \mathrm{N}(0, \sigma^2)$ are independent, and $X_0 \sim \mathrm{N}(0, c)$, $c = \frac{1}{1-\alpha^2}$. Then

$$\begin{aligned}
h_{\varphi_n}^{w}(f_n) = \frac{1}{2}\mathbb{E}\bigg[\prod_{j=0}^{n-1} \psi(X_j) \bigg(& X_0^2 - 2\alpha X_0 X_1 \\
& + (1+\alpha^2)X_1^2 - 2\alpha X_1 X_2 + (1+\alpha^2)X_2^2 - \cdots \\
& + (1+\alpha^2)X_{n-2}^2 - 2\alpha X_{n-2}X_{n-1} + X_{n-1}^2 - 2\log\Big(\frac{\sqrt{1-\alpha^2}}{(2\pi)^{n/2}}\Big)\bigg)\bigg].
\end{aligned}$$

$$(10.15)$$

Conditions of Theorem 10.5 may be checked under some restrictions on the WF ψ, see [18].

Acknowledgments

References

[1] Cover, T., Thomas, J.: Elements of Information Theory. John Wiley, New York (2006). MR2239987

[2] Dembo, A.: Simple proof of the concavity of the entropy power with respect to added Gaussian noise. IEEE Trans. Inf. Theory **35**(4), 887–888 (1989). MR1013698. doi:10.1109/18.32166

[3] Dembo, A., Cover, T., Thomas, J.: Information theoretic inequalities. IEEE Trans. Inf. Theory **37**, 1501–1518 (1991). MR1134291. doi:10.1109/18.104312

[4] Frizelle, G., Suhov, Y.: An entropic measurement of queueing behaviour in a class of manufacturing operations. Proc. R. Soc. Lond., Ser A **457**, 1579–1601 (2001). doi:10.1098/rspa.2000.0731

[5] Frizelle, G., Suhov, Y.: The measurement of complexity in production and other commercial systems. Proc. R. Soc. Lond., Ser A **464**, 2649–2668 (2008). doi:10.1098/rspa.2007.0275

[6] Guiasu, S.: Weighted entropy. Rep. Math. Phys. **2**, 165–179 (1971). MR0289206. doi:10.1016/0034-4877(71)90002-4

[7] Kelbert, M., Suhov, Y.: Information Theory and Coding by Example. Cambridge University Press, Cambridge (2013). MR3137525. doi:10.1017/CBO9781139028448

[8] Kelbert, M., Suhov, Y.: Continuity of mutual entropy in the limiting signal-to-noise ratio regimes. In: Stochastic Analysis, pp. 281–299. Springer, Berlin (2010). MR2789089. doi:10.1007/978-3-642-15358-7_14

[9] Kelbert, M., Stuhl, I., Suhov, Y.: Weighted entropy and optimal portfolios for risk-averse Kelly investments. Aequationes Mathematicae **91** (2017). in press

[10] Kelbert, M., Stuhl, I., Suhov, Y.: Weighted entropy and its use in Computer Science and beyond. Lecture Notes in Computer Science. (in press)

[11] Khan, J.F., Bhuiyan, S.M.: Weighted entropy for segmentation evaluation. Opt. Laser Technol. **57**, 236–242 (2014). doi:10.1016/j.optlastec.2013.07.012

[12] Lai, W.K., Khan, I.M., Poh, G.S.: Weighted entropy-based measure for image segmentation. Proc. Eng. **41**, 1261–1267 (2012). doi:10.1016/j.proeng.2012.07.309

[13] Lieb, E.: Proof of entropy conjecture of Wehrl. Commun. Math. Phys. **62**, 35–41 (1978). MR0506364. doi:10.1007/BF01940328

[14] Nawrockia, D.N., Harding, W.H.: State-value weighted entropy as a measure of investment risk. Appl. Econ. **18**, 411–419 (1986). doi:10.1080/00036848600000038

[15] Paksakis, C., Mermigas, S., Pirourias, S., Chondrokoukis, G.: The role of weighted entropy in security quantification. Int. Journ. Inf. Electron. Eng. **3**(2), 156–159 (2013)

[16] Rioul, O.: Information theoretic proofs of entropy power inequality. IEEE Trans. Inf. Theory **57**(1), 33–55 (2011). MR2810269. doi:10.1109/TIT.2010.2090193

[17] Shockley, K.R.: Using weighted entropy to rank chemicals in quantitative high throughput screening experiments. J. Biomol. Screen. **19**, 344–353 (2014). doi:10.1177/1087057113505325

[18] Suhov, Y.: Stuhl I. Weighted information and entropy rates (2016). arXiv:1612.09169v1

[19] Suhov, Y., Sekeh, S.: An extension of the Ky-Fan inequality. arXiv:1504.01166

[20] Suhov, Y., Stuhl, I., Sekeh, S., Kelbert, M.: Basic inequalities for weighted entropy. Aequ. Math. **90**(4), 817–848 (2016). MR3523101. doi:10.1007/s00010-015-0396-5

[21] Suhov, Y., Sekeh, S., Kelbert, M.: Entropy-power inequality for weighted entropy. arXiv:1502.02188

[22] Suhov, Y., Yasaei Sekeh, S.: Stuhl I. Weighted Gaussian entropy and determinant inequalities. arXiv:1505.01753v1

[23] Tsui, P.-H.: Ultrasound detection of scatterer concentration by weighted entropy. Entropy **17**, 6598–6616 (2015). doi:10.3390/e17106598

[24] Verdú, S., Guo, D.: A simple proof of the entropy-power inequality. IEEE Trans. Inf. Theory **52**(5), 2165–2166 (2006). MR2234471. doi:10.1109/TIT.2006.872978

[25] Villani, C.: A short proof of the "concavity of entropy power". IEEE Trans. Inf. Theory **46**, 1695–1696 (2000). MR1768665. doi:10.1109/18.850718

[26] Yang, L., Yang, J., Peng, N., Ling, J.: Weighted information entropy: A method for estimating the complex degree of infrared images' backgrounds. In: Kamel, M., Campilho, A. (eds.) Image Analysis and Recognition, vol. 3656, Springer, Berlin/Heidelberg, pp. 215–22 (2005). MR3157460. doi:10.1007/978-3-642-39094-4

[27] Zamir, R.: A proof of the Fisher information inequality via a data processing argument. IEEE Trans. Inf. Theory **44**(3), 1246–1250 (1998). MR1616672. doi:10.1109/18.669301