Zero-Entropy Minimization for Blind Extraction of Bounded Sources (BEBS)

Frédéric Vrins[1], Deniz Erdogmus[2], Christian Jutten[3], and Michel Verleysen[1,⋆]

[1] Machine Learning Group, Université catholique de Louvain, Louvain-la-Neuve, Belgium
[2] Dep. of CSEE, OGI Oregon, Health and Science University, Portland, Oregon, USA
[3] Laboratoire des Images et des Signaux, Institut National Polytechnique de Grenoble (INPG), France
{vrins, verleysen}@dice.ucl.ac.be, derdogmus@ieee.org, christian.jutten@inpg.fr

Abstract. Renyi's entropy can be used as a cost function for blind source separation (BSS). Previous works have emphasized the advantage of setting Renyi's exponent to a value different from one in the context of BSS. In this paper, we focus on zero-order Renyi's entropy minimization for the blind extraction of bounded sources (BEBS). We point out the advantage of choosing the *extended* zero-order Renyi's entropy as a cost function in the context of BEBS, when the sources have non-convex supports.

1 Introduction

Shannon's entropy is a powerful quantity in information theory and signal processing; it can be used e.g. in blind source separation (BSS) applications. Shannon's entropy can be seen as a particular case of *Renyi's entropy*, defined as [1]:

$$h_r[f_X] = \begin{cases} \frac{1}{1-r}\log\left\{\int f_X^r(x)dx\right\} & \text{for } r \in \{[0,1) \cup (1,\infty)\} \\ -\mathrm{E}\left\{\log f_X(x)\right\} & \text{for } r = 1 \end{cases} . \tag{1}$$

The above integrals are evaluated on the support $\Omega(X)$ of the probability distribution (pdf) f_X. The first-order Renyi's entropy $h_1[f_X]$ corresponds to Shannon's entropy; function $h_r[f_X]$ is continuous in r.

Previous works have already emphasized that advantages can be taken by considering the general form of Renyi's entropy rather than Shannon's in the BSS context [2]. For instance, it is interesting to set $r = 2$ in specific cases. Using kernel density estimates leads to a simple estimator for $h_2[.]$.

This paper points out that in particular situations, e.g. when dealing with the blind extraction of bounded sources (BEBS) application, zero-Renyi's entropy (Renyi's entropy with $r = 0$) should be preferred to other Renyi's entropies.

Renyi's entropy with $r = 0$ is a very specific case; it simply reduces to the logarithm of the support volume of $\Omega(X)$: $h_0[f_X] = \log \mathrm{Vol}[\Omega(X)]$ [3]. In the BEBS

⋆ Michel Verleysen is Research Director of the Belgian F.N.R.S.

J. Rosca et al. (Eds.): ICA 2006, LNCS 3889, pp. 747–754, 2006.

context, it can be shown that the global minimum of the output zero-Renyi's entropy is reached when the output is proportional to the source with the lowest support measure [4], under the whiteness constraint. The other sources can then be iteratively extracted, minimizing the output zero-Renyi's entropy in directions orthogonal to the previously extracted signals. A similar conclusion has been independently drawn in [5], where it is also shown that the output support convex hull volume has a local minimum when the output is proportional to one of the sources. The main advantage in considering zero-Renyi's entropy is that, under mild conditions, this cost function is free of local minima. Hence gradient-based methods yield the optimal solution of the BEBS problem. When the sources have strongly multimodal pdfs, this property is not shared by the most popular information-theoretic cost functions, like e.g. mutual information, maximum-likelihood and Shannon's marginal entropy (see [6,7,8,9]).

This contribution aims at analyzing the condition for which the "spurious minima-free" property of zero-Renyi's entropy $h_0[f_X]$ holds in the context of BEBS. First, it is shown that the output zero-Renyi's entropy has no spurious minimum in the BEBS application when the volume of the non-convex part of the sources support is zero. Second, it is shown that the support $\Omega[.]$ should be replaced by its convex hull $\overline{\Omega}[.]$ in Renyi's entropy definition (1), in order to avoid spurious minima when the source supports have non-convex parts having a strictly positive volume measure. These two last claims are based on the Brunn-Minkowski inequality.

The following of the paper is organized as follows. The impact of choosing the support pdf or its convex hull when computing Renyi's entropy is first analyzed in Section 2. Section 3 recalls the Brunn-Minkowski inequality. The latter is used to discuss the existence of spurious zero-Renyi's entropy minima depending of the convexity of the source supports in Section 4. The theoretical results are illustrated on a simple example in Section 5.

2 Support, Convex Hull and Renyi's Entropy

The density f_X of a one-dimensional bounded random variable (r.v.) X satisfies $f_X(x) = 0$ for all $x > \sup(X)$ and $x < \inf(X)$. The support of the density is defined as the set where the r.v. *lives* [10]: $\Omega(X) \triangleq \{x : f_X(x) > 0\}$. Another viewpoint is e.g. to consider that the r.v. lives for x such that $0 < F_X(x) < 1$, where F_X is the cumulative distribution of X. Therefore, an extended definition of the support could be : $\overline{\Omega}(X) \triangleq \{x \in [\inf\{x : f_X(x) > 0\}, \sup\{x : f_X(x) > 0\}]\}$. Then, $\overline{\Omega}(X)$ can be seen as the *closed bounded convex hull* of $\Omega(X)$, and obviously: $\Omega(X) \subseteq \overline{\Omega}(X)$.

Let us abuse notation by writing $h_{r,\Omega(X)}[f_X]$ for $h_r[f_X]$. Consider the slightly modified Renyi's entropy (called in the following *extended Renyi's entropy*), defined as $h_{r,\overline{\Omega}(X)}[f_X]$: f_X^r is now integrated on the set $\overline{\Omega}(X)$ rather than on the support $\Omega(X)$ in eq. (1). For $r \neq 0$, one gets $h_{r,\Omega(X)}[f_X] = h_{r,\overline{\Omega}(X)}[f_X]$, because $0^r = 0$ for $r \neq 0$ and $0 \log 0 = 0$ by convention [10]. Conversely, $h_{0,\overline{\Omega}(X)}[f_X] > h_{0,\Omega(X)}[f_X]$ if $\mathrm{Vol}[\overline{\Omega}(X) \setminus \Omega(X)] > 0$ (the support contains 'holes' with non-zero volume measure). Indeed, consider the Lebesgue measure $\mu[.]$, which is the standard way of assigning a volume to subsets of the Euclidean space. Let us assume that $\Omega(X)$ can be written as the union of I disjoint intervals $\Omega_i(X)$ of strictly positive volume. Using the properties of Lebesgue measure, zero-Renyi's entropy becomes : $h_{0,\Omega(X)}[f_X] = \log \sum_{i=1}^{I} \mu[\Omega_i(X)]$. This quantity is strictly lower than $h_{0,\overline{\Omega}(X)}[f_X] = \log \mu[\overline{\Omega}(X)]$ if $\mu[\overline{\Omega}(X) \setminus \Omega(X)] > 0$. In summary, we have:

$$\begin{cases} h_{r,\Omega(X)}[f_X] = h_{r,\overline{\Omega}(X)}[f_X] \text{ for } r \neq 0. \\ \lim_{r\to 0} h_{r,\Omega(X)}[f_X] = h_{0,\Omega(X)}[f_X] \leq h_{0,\overline{\Omega}(X)}[f_X] \end{cases} . \tag{2}$$

The $r = 0$ case is thus very specific when considering Renyi's entropies; for other values of r, $h_{r,\Omega(X)}[f_X] = h_{r,\overline{\Omega}(X)}[f_X]$. The $r = 0$ value is also the only one for which $h_{r,\overline{\Omega}(X)}[f_X]$ can be not continuous in r. The impact of choosing $h_{0,\overline{\Omega}(X)}[f_X]$ rather than $h_{0,\Omega(X)}[f_X]$ as BEBS cost function is analyzed in Section 4. The study is based on Brunn-Minkowski's inequality [11], which is introduced below.

3 Brunn-Minkowski Revisited

The following theorem presents the original Brunn-Minkowski inequality [11].

Theorem 1 (Brunn-Minkowski Inequality). *If $\mathcal{X}$ and $\mathcal{Y}$ are two compact convex sets with nonempty interiors (i.e. mesurable) in $\mathbb{R}^n$, then for any $s, t > 0$:*

$$\mathrm{Vol}^{1/n}[s\mathcal{X} + t\mathcal{Y}] \geq s\mathrm{Vol}^{1/n}[\mathcal{X}] + t\mathrm{Vol}^{1/n}[\mathcal{Y}] \ . \tag{3}$$

The operator Vol[.] *stands for* volume. *The operator "+" means that $\mathcal{X} + \mathcal{Y} = \{x + y : x \in \mathcal{X}, y \in \mathcal{Y}\}$. The equality holds when $\mathcal{X}$ and $\mathcal{Y}$ are equal up to translation and dilatation (i.e. when they are* homothetic*).*

As explained in the previous section, we use the Lebesgue measure $\mu[.]$ as the volume Vol[.] operator. Obviously, one has $\mu[\overline{\Omega}(X)] \geq \mu[\Omega(X)] \geq 0$.

Inequality (3) has been extended in [10,12] to non-convex bodies; in this case however, to the authors knowledge, the *strict equality* and *strict inequality* cases were not discussed in the literature. Therefore, the following lemma, which is an extension of the Brunn-Minkowski theorem in the $n = 1$ case, states sufficient conditions so that the strict equality holds (the proof is relegated to the appendix).

Lemma 1. *Suppose that $\Omega(X) = \cup_{i=1}^{I} \Omega_i(X)$ with $\mu[\Omega_i(X)] > 0$ and $\Omega(Y) = \cup_{j=1}^{J} \Omega_j(Y)$ with $\mu[\Omega_j(Y)] > 0$, with $\Omega(X) \subset \mathbb{R}$, $\Omega(Y) \subset \mathbb{R}$. Then:*

$$\mu[\Omega(X + Y)] \geq \mu[\Omega(X)] + \mu[\Omega(Y)] \ ,$$

with equality if and only if $\mu[\overline{\Omega}(X) \setminus \Omega(X)] = \mu[\overline{\Omega}(Y) \setminus \Omega(Y)] = 0$.

4 Zero-Renyi's *vs* Extended Zero-Renyi's Entropy for BEBS

Consider the linear instantaneous BEBS application, and let $S_1, S_2, \cdots, S_K$ be the independent source signals. If we focus on the extraction of a single output Z, we can write $Z = \sum_{i=1}^{K} \mathbf{c}(i) S_i$, where $\mathbf{c}$ is the vector of the transfer weights between Z and the S_i. The vector $\mathbf{c}$ is the row of the transfer matrix $\mathbf{C}$ associated to the output Z. The latter matrix is obtained by left-multiplying the unknown mixing matrix by the unmixing matrix that has to be estimated. The unmixing matrix row associated to $\mathbf{c}$ can be blindly found by minimizing $h_{0,\Omega(Z)}[f_Z]$, under a fixed-norm constraint to avoid that $\mathrm{var}(Z)$ diverges (see [4,5]).

The following subsections discuss the impact of minimizing zero-Renyi's entropy $h_{0,\Omega(Z)}[f_Z]$ or its extended definition $h_{0,\overline{\Omega}(Z)}[f_Z]$ for the BEBS application.

4.1 Convex Supports

If the sources have convex supports $\Omega(S_i)$ (Theorem 1), or if $\mu[\overline{\Omega}(S_i) \setminus \Omega(S_i)] = 0$ (Lemma 1) for all $1 \leq i \leq K$, then both approaches are identical: $\mu[\Omega(Z)] = \mu[\overline{\Omega}(Z)]$. Brunn-Minkowski equality holds, and the following relation comes: $\mu[\Omega(Z)] = \sum_{i=1}^{K} |\mathbf{c}(i)|.\mu[\Omega(S_i)]$. It is known that in the $K = 2$ case, we can freely parametrize $\mathbf{c}$ by a single angle: $\mathbf{c}$ can be written as $[\sin\theta, \cos\theta]$, where θ is the transfer angle. This parametrization of $\mathbf{c}$ forces the vector to have a unit Euclidean norm. In this case $\mu[\Omega(Z)] = \mu[\overline{\Omega}(Z)]$ is concave w.r.t. θ in each quadrant [5]. Since $\log f$ is concave if f is concave, $\log\mu[\Omega(Z)] = \log\mu[\overline{\Omega}(Z)]$ is also concave. In other words, the minima of $\mu[\Omega(Z)]$ w.r.t. θ can only occur at $\theta \in \{k\pi/2 | k \in \mathbb{Z}\}$: all the minima of $h_{0,\Omega(Z)}[f_Z]$ are non-mixing (corresponding to non-spurious solutions of the BEBS problem). This last result holds for higher dimensions, i.e. for $K \geq 2$ (see [5] for more details).

4.2 Non-convex Supports

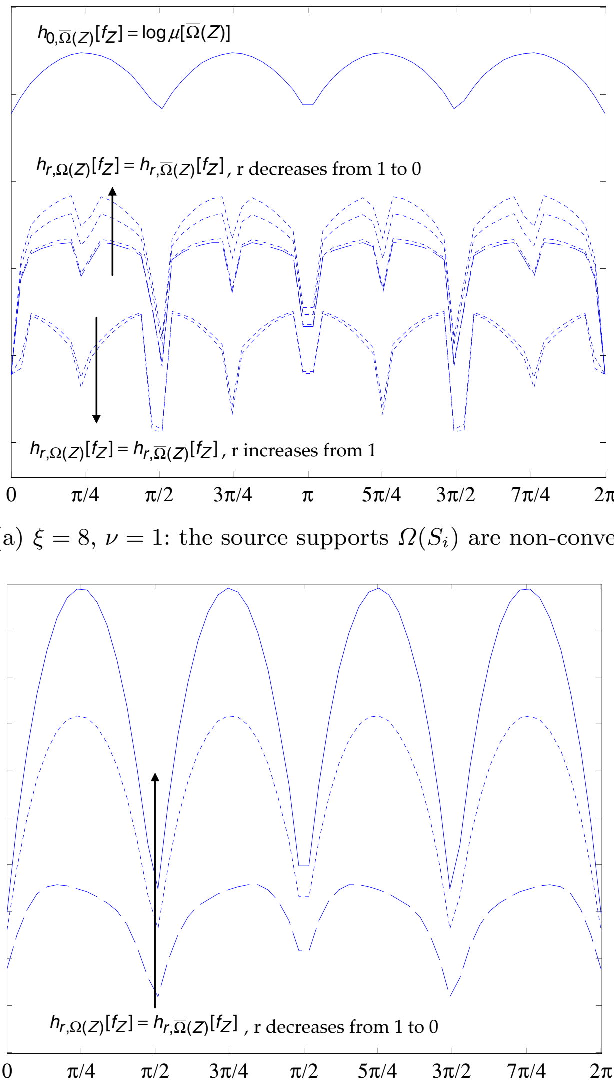
In the non-convex situation, Brunn-Minkowski equality holds for the set $\overline{\Omega}(.)$ (by Theorem 1):

$$\mu[\overline{\Omega}(Z)] = \sum_{i=1}^{K} |\mathbf{c}(i)|.\mu[\overline{\Omega}(S_i)] \ . \tag{4}$$

It can be shown that all the minima of the above quantity w.r.t. vector $\mathbf{c}$ are relevant; as in the convex-support case, they all correspond to non-spurious solutions of the BEBS problem [5]. By contrast, the strict Brunn-Minkowski inequality holds when a source has a support $\Omega(S_i)$ such that $\mu[\overline{\Omega}(S_i) \setminus \Omega(S_i)] > 0$. Lemma 1 gives $\mu[\Omega(Z)] > \sum_{i=1}^{K} |\mathbf{c}(i)|.\mu[\Omega(S_i)]$. In this case, there is no more guarantee that $\mu[\Omega(Z)]$ does not have mixing minima when a source has a non-convex support. The next section will presents simulation results showing on a simple example that spurious minima of $\mu[\Omega(Z)]$ may exist.

As a conclusion, the best integration domain for evaluating Renyi's entropy for the blind separation of bounded sources seems to be $\overline{\Omega}(Z)$, the convex hull of the output support $\Omega(Z)$. Remark that contrarily to $h_{r,\Omega(Z)}[f_Z]$, $h_{r,\overline{\Omega}(Z)}[f_Z]$ is not rigourously speaking a Renyi's entropy. Nevertheless, while $h_{0,\Omega(Z)}[f_Z]$ is the log of the volume of $\Omega(Z)$, *extended* zero-Renyi's entropy $h_{0,\overline{\Omega}(Z)}[f_Z]$ is the log of the volume of $\Omega(Z)$'s *convex hull*.

In the BEBS application, the output volume must be estimated directly from Z, since neither $\mathbf{c}$, nor the $\mu[\Omega(S_i)]$ are known. Therefore evaluating zero-Renyi's entropy requires the estimation of $\mu[\Omega(Z)]$ and computing extended zero-Renyi's



(a) $\xi = 8$, $\nu = 1$: the source supports $\Omega(S_i)$ are non-convex

(b) $\xi = 0.5$, $\nu = 1$: the source supports $\Omega(S_i)$ are convex

Fig. 1. Extended zero-Renyi (–), Shannon (- -), and r-Renyi entropies with $r \neq \{0, 1\}$ (..)

entropy requires the estimation of $\mu[\overline{\Omega}(Z)]$. In [5], the support of $\overline{\Omega}(Z)$ is approximated by $\max(\hat{Z}) - \min(\hat{Z})$ ($\hat{Z}$ is the set of observations of Z), which is also a good approximation of $\mu[\Omega(Z)]$ (i.e. of $\exp\{h_{0,\Omega(Z)}[f_Z]\}$) when the source supports are convex.

5 Example

Let p_{S_1} and p_{S_2} be two densities of independent random variables $S_i = U_i + D_i$ where U_1 and U_2 are independent uniform variables taking non-zero values in $[-\nu, \nu]$ ($\nu > 0$) and D_1, D_2 are independent discrete random variables taking values $[\alpha, 1-\alpha]$ at $\{-\xi, \xi\}$ ($\xi > 0$). Suppose further that $\xi > \nu$. Then, both sources S_i have the same density p_S:

$$p_S(s) = \begin{cases} \frac{\alpha}{2\nu} & \text{for } x \in [-\xi-\nu, -\xi+\nu] \\ \frac{1-\alpha}{2\nu} & \text{for } x \in [\xi-\nu, \xi+\nu] \\ 0 & \text{elsewhere.} \end{cases} \tag{5}$$

It results that $\Omega(S_i) = \{x \in [-\xi-\nu, -\xi+\nu] \cup [\xi-\nu, \xi+\nu]\}$ and $\overline{\Omega}(S_i) = \{x \in [-\xi-\nu, \xi+\nu]\}$, which implies $\mu[\Omega(S_i)] = 4\nu$ and $\mu[\overline{\Omega}(S_i)] = 2\xi + 2\nu$. By Lemma 1, we have $\mu[\overline{\Omega}(S_1 + S_2)] = \mu[\overline{\Omega}(S_1)] + \mu[\overline{\Omega}(S_2)]$ and $\mu[\Omega(S_1 + S_2)] > \mu[\Omega(S_1)] + \mu[\Omega(S_2)]$.

Let us note $Z = \cos\theta S_1 + \sin\theta S_2$. Equation (4) guarantees that $\mu[\overline{\Omega}(Z)]$ is concave with respect to θ. By contrast, according to Section 4.2, there is no guarantee that $\mu[\Omega(Z)]$ has no minima for $\theta \notin \{k\pi/2 | k \in \mathbb{Z}\}$.

Figure 1 illustrates the effect of the source support convexity on $h_r[f_Z]$ w.r.t. θ for various values of r in the above example. Note that the omitted scales of vertical axes are common to all curves. We can observe that $h_r[f_Z]$ has spurious minima regardless of r; there exist local minima of the zero-entropy criterion for which Z is not proportional to one of the sources. By contrast, when considering the extended zero-Renyi's entropy $h_{r,\overline{\Omega}(Z)}[f_Z]$, no spurious minimum exists: all the $h_{r,\overline{\Omega}(Z)}[f_Z]$ local minima correspond to $Z = \pm S_i$. Note that in Figure 1 (a), $h_{r_1}[f_Z] < h_{r_2}[f_Z]$ if $r_1 > r_2$. This result can be theoretically proven by Hölder's inequality: Renyi's entropy $h_r[f]$ is decreasing in r, and strictly decreasing unless f is a uniform density [13].

6 Conclusion and Perspectives

This paper focusses on zero-Renyi's entropy for blind extraction of bounded sources. Theoretical results show that if the sources have convex supports, both zero-Renyi and extended zero-Renyi's entropies are free of spurious minima.

However, this is no more true when the source support contains "holes" of positive volume. In this case, simulation results seem to indicate that the order of Renyi's entropy (i.e. parameter r) has no influence on the existence of local spurious minima, see Figure 1 (a). Nevertheless, when considering the extended zero-Renyi's entropy, Brunn-Minkowski inequality shows that this cost function is free of spurious minima, when the support is correctly estimated.

Finally, new perspectives for Renyi entropy-based BSS/BEBS algorithms arise from the results presented in this paper. Despite the "no spurious minima" property of the extended zero-Renyi's entropy which is not shared by Shannon's one, both the output support volume and Shannon's entropy BEBS can be used in *deflation* algorithms for source separation. Indeed, it is known that the *Entropy Power inequality* shows that Shannon's entropy can be used in deflation procedures for BSS. On the other hand, this paper shows that *Brunn-Minkowski inequality* justifies the use of zero-Renyi's entropy for the sequential extraction of bounded sources. Conversely, to the authors knowledge, there is no proof to date justifying the use of Renyi's entropy for $r \neq 0$ and $r \neq 1$ in *deflation* BSS/BEBS algorithms. It is thus intriguing to remark that the two aforementioned information-theoretic inequalities are closely related [12]. By contrast, the sum of output Renyi's entropies can be seen as a cost function for *symmetric* BSS (all sources are extracted simultaneously), as explained in [2]. As it is known that values of r different from 0 and 1 are also interesting in specific BSS applications, future work should then study deflation methods based on general Renyi's entropy definition (of order $r \in \mathbb{R}^+$).

Acknowledgment

The authors would like to thank Prof. Erwin Lutwak from Polytechnic University (Brooklyn, NY, USA) for discussion on Hölder's inequality.

References

1. Principe, J.C Xu, D. & Fisher III, J.W. (2000) "Information-Theoretic Learning." In *Unsup. Adapt. Filtering* **I**: 265–319, Haykin, S. (edt), Wiley, New York.
2. Erdogmus, D., Hild, K.E. & Principe, J.C. (2002) Blind Source Separation Using Renyi's α-marginal Entropies. *Neurocomputing*, **49**: 25–38.
3. Guleryuz, O.G., Lutwak, E., Yang, D., & Zhang, G. (2000) Information-Theoretic Inequalities for Contoured Probability Distributions. *IEEE Trans. Info. Theory*, **48**(8): 2377–2383.
4. Cruces, S. & Duran, I. (2004) The Minimum Support Criterion for Blind Source Extraction: a Limiting Case of the Strengthened Young's Inequality. In proc. ICA, *Int. Conf. on Ind. Comp. Anal. and Blind Sig. Sep.*: 57–64, Granada (Spain).
5. Vrins, F., Jutten, C. & Verleysen, M. (2005) SWM : a Class of Convex Contrasts for Source Separation. In proc. ICASSP, *IEEE Int. Conf. on Acoustics, Speech and Sig. Process.*: V.161-V.164, Philadelphia (USA).
6. Vrins, F. & Verleysen, M . (2005) On the Entropy Minimization of a Linear Mixture of Variables for Source Separation. *Sig. Process.*, **85**(5): 1029–1044, Elsevier.
7. Learned-Miller, E.G. & Fisher III, J.W. (2003) ICA Using Spacings Estimates of Entropy. *Journal of Machine Learning Research*, **4**: 1271–1295, MIT press.
8. Pham, D.T. & Vrins, F. (2005) Local Minima of Information-Theoretic Criteria in Blind Source Separation. *IEEE Sig. Process. Lett.*, **12**(11): 788-791.

9. Cardoso, J.-C. (2000) “Entropic Contrast for Source Separation: Geometry & Stability.” In *Unsup. Adapt. Filtering* **I**: 265–319, Haykin, S. (edt), Wiley, New York.
10. Cover, Th. & Thomas, J. A. (1991) “Elements of Information Theory.” Wiley, New-York.
11. Gardner, R.J. (2002) The Brunn-Minkowski Inequality. *Bull. of Am. Math. Soc.*, **39**(3): 355–405.
12. Costa, M. & Cover Th. (1984) On the Similarity of the Entropy Power Inequality and the Brunn-Minkowski Inequality. *IEEE Trans. Info. Theory*, **30**(6): 837–839.
13. Lutwak, E., Yang, D., & Zhang, G. (2005) Cramér-Rao and Moment-Entropy Inequalities for Renyi Entropy and Generalized Fisher Information. *IEEE Trans. Info. Theory*, **51**(2): 473–478.

Appendix. Proof of Lemma 1

Suppose that $\mu[\Omega(X)] = \mu[\overline{\Omega}(X)] > 0$ and $\mu[\Omega(Y)] = \mu[\overline{\Omega}(Y)] > 0$. This means that $\mu[\overline{\Omega}(X) \setminus \Omega(X)] = \mu[\overline{\Omega}(Y) \setminus \Omega(Y)] = 0$. Therefore, the sets $\Omega(X)$ and $\Omega(Y)$ can be expressed as

$$\begin{cases} \Omega(X) = [\inf X, \sup X] \setminus \cup_{i=1}^{I'} \{x_i\} \\ \Omega(Y) \ = [\inf Y, \sup Y] \setminus \cup_{j=1}^{J'} \{y_j\} \end{cases} \tag{6}$$

where x_i, y_i are isolated points. Then,

$$\begin{aligned} \mu[\Omega(X+Y)] &= \mu[\overline{\Omega}(X+Y)] \\ &= (\sup X + \sup Y) - (\inf X + \inf Y) \\ &= \mu[\Omega(X)] + \mu[\Omega(Y)] \ , \end{aligned}$$

which yields the first result of the Lemma.

To prove the second claim, suppose that $X^\star = \cup_{i=1}^{I-1} [X_i^m, X_i^M] \setminus \cup_{i'=1}^{I'} \{x_{i'}\}$, $Y^\star = \cup_{j=1}^{J-1} [Y_j^m, Y_j^M] \setminus \cup_{j'=1}^{J'} \{y_{j'}\}$ and $X = X^\star \cup [X_I^m, X_I^M] \setminus \cup_{i^\star=1}^{I^\star} \{x_{i^\star}\}$, $Y = Y^\star \cup [Y_J^m, Y_J^M] \setminus \cup_{j^\star=1}^{J^\star} \{y_{j^\star}\}$ where $X_i^m < X_i^M < X_{i+1}^m$, $Y_i^m < Y_i^M < Y_{i+1}^m$ and $X_I^m = X_{I-1}^M + \epsilon$, $\epsilon > 0$. Then, if we note $\Delta_X \triangleq X_I^M - X_I^m$ and $\Delta_Y \triangleq Y_J^M - Y_J^m$, we have:

$$\mu[\Omega(X+Y)] \geq \mu[\Omega(X^\star+Y)] + \left\{(Y_J^M + X_I^M) - \max(X_{I-1}^M + Y_J^M, Y_J^m + X_I^m)\right\} \ ,$$

where the term into brackets is a lower bound of the sub-volume of $\Omega(X+Y)$ due to the interval $[X_I^m, X_I^M]$; it can be rewritten as $\min\{\Delta_X + \epsilon, \Delta_X + \Delta_Y\}$. Finally, having the Brunn-Minkowski inequality in mind, one gets:

$$\begin{aligned} \mu[\Omega(X+Y)] &\geq \mu[\Omega(X^\star+Y)] + \min\{\Delta_X + \epsilon, \Delta_X + \Delta_Y\} \\ &\geq \mu[\Omega(X)] - \Delta_X + \mu[\Omega(Y)] + \min\{\Delta_X + \epsilon, \Delta_X + \Delta_Y\} \\ &> \mu[\Omega(X)] + \mu[\Omega(Y)] \ . \end{aligned}$$