UDC 519.722

Weighted entropy: basic facts and properties

M. Kelbert*, **Y. Suhov**†‡, **I. Stuhl**§

* *Moscow Higher School of Economics, RF*

† *Math Dept, Penn State University, PA, USA*

‡ *DPMMS, University of Cambridge, UK*

§ *Math Dept, University of Denver, USA*

Abstract. The concept of weighted entropy takes into account values of different outcomes, i.e., makes entropy context-dependent, through the weight function. We analyse an analog of the entropy-power inequality for the weighted entropy and discuss connections with weighted Lieb's splitting inequality.

Keywords: weighted entropy, Gibbs inequality, Ky-Fan inequality, entropy power inequality, Lieb splitting inequality.

1. Introduction

We all know that the Shannon entropy of a probability distribution $\mathbf{p}$

$$h(\mathbf{p}) = -\sum p(x_i)\log p(x_i)$$

is context-free, i.e., it does not depend on the nature of outcomes x_i, only upon probabilities $p(x_i)$. However, imagine two equally rare medical conditions, occurring with probability $p << 1$, one of which carries a major health risk while the other is just a peculiarity. Formally, they provide the same amount of information $-\log p$ but the value of this information can be very different. So, we may modify the definition to make it context dependent. The weighted entropy is defined as

$$h^w_\phi(\mathbf{p}) = -\sum \phi(x_i)p(x_i)\log p(x_i).$$

A positive function $x_i \to \phi(x_i) \in \mathbf{R}_+$ represents weights of outcomes x_i. A popular example is $\phi(x) = \mathbf{1}(x \in A)$ where A is a particular subcollection of outcomes. A similar approach can be proposed for differential entropy of the probability density function (PDF) f_Z of random variable (RV) Z. Define the weighted differential entropy (WDE) as

$$h^w_\phi(f_Z) = h^w_\phi(Z) = -\mathbf{E}\big[\phi(Z)\log f_Z(Z)\big] = -\int \phi(x)f_Z(x)\log f_Z(x)\mathrm{d}x. \tag{1}$$

We say that WDE is finite if RV Z has a density and the integral in (1) is absolutely convergent. Basic properties of WDE are studied in [4]. As an example, take $f(x)=f^{No}_C(x)$ where $f^{No}_C(x)$ stands for $d-$dimensional Gaussian distribution with mean 0 and covariance matrix C

$$h^w_\phi(f^{No}_C)=\frac{\alpha_\phi(C)}{2}\log\big[(2\pi)^d\det(C)\big]+\frac{\log e}{2}\mathrm{tr}\big[C^{-1}\Phi_{C,\phi}\big]\quad \textit{where}$$
$$\alpha_\phi(C)=\int_{\mathbf{R}^d}\phi(x)f^{No}_C(x)\mathrm{d}x,\ \Phi_{C,\phi}=\int_{\mathbf{R}^d}xx^{\mathrm{T}}\phi(x)f^{No}_C(x)\mathrm{d}x.$$

2. The weighted Gibbs inequality

Given two non-negative functions f,g define the weighted Kullback-Leibler divergence (or relative WDE) as

$$D^w_\phi(f||g)=\int\phi(x)f(x)\log\frac{f(x)}{g(x)}\mathrm{d}x.$$

Proposition 1 Suppose that

$$\int\phi(x)[f(x)-g(x)]\mathrm{d}x\geq 0.$$

Then $D^w_\phi(f||g)\geq 0$.

3. Concavity and convexity of weighted entropy

Theorem 2 (a) The function $f\to h^w_\phi(f)$ is concave in argument f. Namely, for any PDFs $f_1(x),f_2(x)$, a non-negative function $x\to\phi(x)$, and $\lambda_1,\lambda_2\in[0,1]$ such that $\lambda_1+\lambda_2=1$

$$h^w_\phi(\lambda_1 f_1+\lambda_2 f_2)\geq\lambda_1 h^w_\phi(f_1)+\lambda_2 h^w_\phi(f_2).$$

This inequality is strict unless $\phi(x)[f_1(x)-f_2(x)]=0$ for $(\lambda_1 f_1+\lambda_2 f_2)-$almost all x. (b) However, the relative WDE is convex: given two pairs of PDFs (f_1,f_2) and (g_1,g_2)

$$\lambda_1 D^w_\phi(f_1||g_1)+\lambda_2 D^w_\phi(f_2||g_2)\geq D^w_\phi(\lambda_1 f_1+\lambda_2 f_2||\lambda_1 g_1+\lambda_2 g_2).$$

4. Ky-Fan type inequalities

It is well-known that $C\to\delta(C)=\log\det(C)$ is a concave function of a (strictly) positive-definite matrix C

$$\delta(\lambda_1 C_1+\lambda_2 C_2)\geq\lambda_1\delta(C_1)+\lambda_2\delta(C_2)$$

where $\lambda_1+\lambda_2=1, \lambda_i \geq 0$. This is the well-known Ky-Fan inequality. It terms of differential entropies it is equivalent to the inequality

$$h(f_C^{No}) \geq \lambda_1 h(f_{C_1}^{No}) + \lambda_2 h(f_{C_2}^{No}).$$

where $C=\lambda_1 C_1+\lambda_2 C_2$. Theorem 3 below presents a previously unknown series of bounds of Ky-Fan type. The most explicit results are available for $\phi(\mathbf{x})=\exp\big(\mathbf{x}^T\mathbf{t}\big), \mathbf{t}\in\mathbf{R}^d$ in view of identity $h_\phi^w(f^{No})=\exp\big(\frac{1}{2}\mathbf{t}^TC\mathbf{t}\big)h(f^{No})$. Introduce a set

$$\mathcal{S}=\{\mathbf{t}\in\mathbf{R}^d: F^{(1)}(\mathbf{t})\geq 0, F^{(2)}(\mathbf{t})\leq 0\} \quad where$$

$$\begin{aligned}
&F^{(1)}(\mathbf{t})=\sum_{i=1}^{2}\lambda_i\exp\big(\tfrac{1}{2}\mathbf{t}^TC_i\mathbf{t}\big)-\exp\big(\tfrac{1}{2}\mathbf{t}^TC\mathbf{t}\big),\\
&F^{(2)}(\mathbf{t})=\Big[\sum_{i=1}^{2}\lambda_i\exp\big(\tfrac{1}{2}\mathbf{t}^TC_i\mathbf{t}\big)-\exp\big(\tfrac{1}{2}\mathbf{t}^TC\mathbf{t}\big)\Big]\log\big[(2\pi)^d\mathrm{det}(C)\big]\\
&+\sum_{i=1}^{2}\lambda_i\exp\big(\tfrac{1}{2}\mathbf{t}^TC_i\mathbf{t}\big)\mathrm{tr}\big[C^{-1}C_i\big]-d\exp\big(\tfrac{1}{2}\mathbf{t}^TC\mathbf{t}\big).
\end{aligned}$$

Theorem 3 Given positive-definite matrices C_1, C_2 and $\lambda_1,\lambda_2\in[0,1]$ with $\lambda_1+\lambda_2=1$, set $C=\lambda_1C_1+\lambda_2C_2$. Assume $\mathbf{t}\in\mathcal{S}$. Then

$$h(f_C^{No})\exp\big(\frac{1}{2}\mathbf{t}^TC\mathbf{t}\big)-h(f_{C_1}^{No})\exp\big(\frac{1}{2}\mathbf{t}^TC_1\mathbf{t}\big)-h(f_{C_2}^{No})\exp\big(\frac{1}{2}\mathbf{t}^TC_2\mathbf{t}\big)\geq 0,$$

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

5. Weighted entropy-power inequality (WEPI)

Let X_1, X_2 be independent RVs with PDF f_1, f_2 and $X=X_1+X_2$. The famous Shannon entropy power inequality states that

$$h(X_1+X_2)\geq h(N_1+N_2)$$

where N_1, N_2 are Gaussian RVs such that $h(X_i)=h(N_i), i=1,2$. Equivalently,

$$e^{2h(X_1+X_2)}\geq e^{2h(X_1)}+e^{2h(X_2)} \tag{2}$$

see, e.g., [1]. We are interested in the Weighted Entropy Power inequality (WEPI)

$$\kappa:=\exp\big(\frac{2h_\phi^w(X_1)}{\mathbf{E}\phi(X_1)}\big)+\exp\big(\frac{2h_\phi^w(X_2)}{\mathbf{E}\phi(X_2)}\big)\leq\exp\big(\frac{2h_\phi^w(X)}{\mathbf{E}\phi(X)}\big). \tag{3}$$

Note that (5) coinsides with (5) when $\phi \equiv 1$. We set

$$\alpha = \tan^{-1}\Big[\exp\Big(\frac{h^w_\phi(X_2)}{\mathbf{E}\phi(X_2)} - \frac{h^w_\phi(X_1)}{\mathbf{E}\phi(X_1)}\Big)\Big], Y_1 = \frac{X_1}{\cos\alpha}, Y_2 = \frac{X_2}{\sin\alpha}. \qquad (4)$$

Theorem 4 Given independent RVs X_1, X_2 with PDFs f_1, f_2, and the weight function ϕ, set $X = X_1 + X_2$. Assume the following conditions: (i)

$$\begin{aligned} &\mathbf{E}\phi(X_i) \geq \mathbf{E}\phi(X)\ if\ \kappa \geq 1,\ i = 1,2, \\ &\mathbf{E}\phi(X_i) \leq \mathbf{E}\phi(X)\ if\ \kappa \leq 1,\ i = 1,2. \end{aligned} \qquad (5)$$

(ii) With Y_1, Y_2 and α as defined in (5)

$$(\cos\alpha)^2 h^w_{\phi_c}(Y_1) + (\sin\alpha)^2 h^w_{\phi_s}(Y_2) \leq h^w_\phi(X) \qquad (6)$$

where $\phi_c(x) = \phi(x\cos\alpha), \phi_s(x) = \phi(x\sin\alpha)$ and

$$h^w_{\phi_c}(Y_1) = -\mathbf{E}\big[\phi_c(Y_1)\log(f_{Y_1}(Y_1))\big], h^w_{\phi_s}(Y_2) = -\mathbf{E}\big[\phi_s(Y_2)\log(f_{Y_2}(Y_2))\big].$$

Then WEPI holds.

Paying homage to [3] we call (5) weighted Lieb's splitting inequality (WLSI). In some cases WLSI may be effectively checked.

Example 5 Let $d = 1$ and $X_1 \sim$N$(0, \sigma_1^2)$, $X_2 \sim$N$(0, \sigma_2^2)$. Then the WLSI (5) takes the following form

$$\begin{aligned} &\log\big[2\pi(\sigma_1^2 + \sigma_2^2)\big]\mathbf{E}\phi(X) + \tfrac{\log e}{\sigma_1^2+\sigma_2^2}\mathbf{E}[X^2\phi(X)] \geq \\ &(\cos\alpha)^2\Big[\log\big(\tfrac{2\pi\sigma_1^2}{(\cos\alpha)^2}\big)\Big]\mathbf{E}\phi(X_1) + \tfrac{(\cos\alpha)^2\log e}{\sigma_1^2}\mathbf{E}[X_1^2\phi(X_1)] \\ &+(\sin\alpha)^2\Big[\log\big(\tfrac{2\pi\sigma_2^2}{(\sin\alpha)^2}\big)\Big]\mathbf{E}\phi(X_2) + \tfrac{(\sin\alpha)^2\log e}{\sigma_2^2}\mathbf{E}[X_2^2\phi(X_2)]. \end{aligned}$$

5.1. WLSI for the weight function close to a constant

Proposition 6 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 $x \to \phi(x)$ is twice continuously differentiable and

$$|\phi''(x)| \leq \epsilon\phi(x), |\phi(x) - \bar{\phi}| \leq \epsilon \qquad (7)$$

where $\epsilon > 0$ and $\bar{\phi} > 0$ are constants. Then there exists $\epsilon_0 > 0$ such that for all WF ϕ satisfying (5.1) with $0 < \epsilon < \epsilon_0$ WLSI holds true. Hence, the checking of WEPI is reduced to condition (5).

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

$$M(Z;\gamma_0) = \mathbf{E}\Big[||Z - \mathbf{E}[Z|Z\sqrt{\gamma} + N]||^2\Big]$$

where $||.||$ stands for Euclidean norm. According to [2, 5] the differential entropy

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

For $Z = Y_1, Y_2, X_1 + X_2$ assume the following conditions

$$\mathbf{E}[|\log f_Z(Z)|] < \infty, \mathbf{E}[||Z||^2] < \infty. \qquad (8)$$

Theorem 7 Assume conditions (5.1). 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

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

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

$$|\phi(x) - \bar{\phi}| < \epsilon. \qquad (9)$$

Then there exists $\epsilon_0 = \epsilon_0(\gamma_0, \delta, f_1, f_2)$ such that for all WFs satisfying (5.1) with $\epsilon < \epsilon_0$ WLSI holds true.

References

1. *Kelbert M., Suhov Y.* Information Theory and Coding by Example. — Cambridge: Cambridge University Press, 2013.
2. *Kelbert M., Suhov Y.* Continuity of mutual entropy in the limiting signal-to-noise ratio regimes / Stochastic Analysis. — Springer-Verlag: Berlin, 2010. — P. 281-299.
3. *Lieb E.* Proof of entropy conjecture of Wehrl // Commun. Math. Phys. — 1978. — Vol. 62. — P. 35–41.
4. *Suhov Y., Stuhl I, Sekeh S., Kelbert M.* Basic inequalities for weighted entropy // Aequationes Mathematicae. — 2016. — Vol. 90, no. 4. — P. 817–848.
5. *Verdú S., Guo D.* A simple proof of the entropy-power inequality // IEEE Transaction on Information Theory. — 2006. — Vol. 52, no. 5. — P. 2165–2166.