Probability in the Engineering and Informational Sciences, **33**, 2019, 186–204.
doi:10.1017/S0269964818000098

BOUNDS ON EXTROPY WITH VARIATIONAL DISTANCE CONSTRAINT

JIANPING YANG
Department of Mathematical Sciences, School of Science, Zhejiang Sci-Tech University, Hangzhou, Zhejiang 310018, People's Republic of China,
E-mail: yangjp@zstu.edu.cn

WANWAN XIA AND TAIZHONG HU
Department of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei Anhui 230026, People's Republic of China,
E-mails: xiaww@mail.ustc.edu.cn; thu@ustc.edu.cn

The relation between extropy and variational distance is studied in this paper. We determine the distribution which attains the minimum or maximum extropy among these distributions within a given variation distance from any given probability distribution, obtain the tightest upper bound on the difference of extropies of any two probability distributions subject to the variational distance constraint, and establish an analytic formula for the confidence interval of an extropy. Such a study parallels to that of Ho and Yeung [3] concerning entropy. However, the proofs of the main results in this paper are different from those in Ho and Yeung [3]. In fact, our arguments can simplify several proofs in Ho and Yeung [3].

Keywords: confidence interval, entropy, extropy, extropy estimation, majorization, variational distance

1. INTRODUCTION

For any positive integer n or infinite n, define

$$\Gamma_n = \left\{ \boldsymbol{p} = (p_1, p_2, \ldots, p_n) : \sum_{i=1}^{n} p_i = 1, p_j \geq 0, 1 \leq j \leq n \right\}.$$

For any $\boldsymbol{p} \in \Gamma_n$, $\boldsymbol{p}$ can be regarded as a probability mass function (pmf) of some discrete random variable X taking values in the set $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$, that is, $p_i = \mathbb{P}(X = x_i)$ for each i. $\mathcal{X}$ is termed as the alphabet of X. The support of $\boldsymbol{p}$ is defined as

$$\mathrm{supp}(\boldsymbol{p}) = \{x_i : p_i > 0, 1 \leq i \leq n\}.$$

© Cambridge University Press 2018 0269-9648/18 $25.00

https://doi.org/10.1017/S0269964818000098 Published online by Cambridge University Press

For any positive integers $\ell < m$ and any $\boldsymbol{p} \in \Gamma_\ell$, $\boldsymbol{p}$ can be regarded as

$$(\boldsymbol{p}, \underbrace{0, \ldots, 0}_{m-\ell}) \in \Gamma_m,$$

equivalently. So we have $\Gamma_\ell \subset \Gamma_m$ for all $1 \leq \ell < m$. For $\boldsymbol{p}, \boldsymbol{q} \in \Gamma_n$ with possibly different supports, the variational distance between $\boldsymbol{p}$ and $\boldsymbol{q}$ is defined as

$$V(\boldsymbol{p}, \boldsymbol{q}) = \sum_{i=1}^{n} |p_i - q_i|.$$

For $\boldsymbol{p} \in \Gamma_\ell$ and $\boldsymbol{q} \in \Gamma_m$ with $m > \ell$, $V(\boldsymbol{p}, \boldsymbol{q})$ is interpreted as

$$V(\boldsymbol{p}, \boldsymbol{q}) = \sum_{i=1}^{\ell} |p_i - q_i| + \sum_{j=\ell+1}^{m} q_j.$$

We first recall the definitions of two information measures, the Shannon entropy (Shannon [8], Cover and Thomas [2]) and the extropy (Lad et al. [4]), for discrete distribution functions. Throughout this paper, the term 'log' will denote the natural logarithm. By convention, $x \log x = 0$ when $x = 0$.

DEFINITION 1.1: *Let X be a discrete random variable, taking values in the set $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$, with pmf $\boldsymbol{p} = (p_1, p_2, \ldots, p_n) \in \Gamma_n$.*

(i) *The Shannon entropy of X is defined as*

$$H(\boldsymbol{p}) = -\sum_{i=1}^{n} p_i \log p_i;$$

(ii) *The extropy of X is defined as*

$$J(\boldsymbol{p}) = -\sum_{i=1}^{n} (1 - p_i) \log(1 - p_i).$$

In Definition 1.1, the Shannon entropy $H(\boldsymbol{p})$ describes the uncertainty of the random variable X and measures how much information is required, on average, to identify a random variable X with pmf $\boldsymbol{p}$. $H(\boldsymbol{p})$ is maximized by the discrete uniform distribution on the set $\mathcal{X}$. For two pmfs $\boldsymbol{p}$ and $\boldsymbol{q}$ of two random variables X and Y taking values in $\mathcal{X}$, $H(\boldsymbol{p}) \leq H(\boldsymbol{q})$ means that the distribution of $\boldsymbol{q}$ is closer to the discrete uniform distribution on $\mathcal{X}$ than the distribution of $\boldsymbol{p}$ in some sense, and that the random behavior of Y is more difficult to predict than that of X. The extropy $J(\boldsymbol{p})$ is a complementary dual of the Shannon entropy $H(\boldsymbol{p})$. An axiomatic characterization and several intriguing properties of the extropy were displayed in Lad et al. [4], in which the motivations of introducing the extropy were also given. The extropy has several potential applications, for example, in the theory of proper scoring rules for alternative forecast distributions.

Two information measures, entropy and extropy, are intertwined with each other. More precisely, for finite $n > 1$, $J(\boldsymbol{p})$ equals a location and scale transform of the entropy of

https://doi.org/10.1017/S0269964818000098 Published online by Cambridge University Press

another pmf

$$\boldsymbol{p}^* = \left(\frac{1-p_1}{n-1}, \frac{1-p_2}{n-1}, \ldots, \frac{1-p_n}{n-1}\right), \tag{1.1}$$

that is,

$$J(\boldsymbol{p}) = (n-1)[H(\boldsymbol{p}^*) - \log(n-1)]. \tag{1.2}$$

For more details of extropy, see Lad et al. [4].

The relation between the Shannon entropy and variational distance was studied extensively in the literature of information; see Ho and Yeung [3], Barbour et al. [1], Ley and Swan [5], Sason and Verdú [7], Verdú [10] and references therein. The main important and interesting results in Ho and Yeung [3] can be summarized as follows:

R1. they determine the distribution which attains the minimum or maximum entropy among these distributions within a given variation distance from any given probability distribution;

R2. obtaine the tightest upper bound on the difference of entropies of two probability distributions in terms of their alphabet sizes and variational distance; and

R3. establish an analytic formula for the confidence interval of an entropy.

It should be pointed out that the proofs concerning Results R1 and R2 are lengthy, complicated and not straightforward.

The purposes of this paper are two-fold. First, parallel to the discussions in Ho and Yeung [3], we will apply majorization theory to establish the above results R1 to R3 for the extropy. Second, the tool (majorization theory) used in this paper can be used to simplify the proofs of Results R1 and R2 concerning entropy in Ho and Yeung [3]. In fact, the arguments of this paper also apply to the following measure

$$K_\psi(\boldsymbol{p}) = -\sum_{i=1}^{n} \psi(p_i), \quad \boldsymbol{p} \in \Gamma_n,$$

where $\psi : [0,1] \to \mathbb{R}$ is any convex function. $K_\psi(\boldsymbol{p})$ reduces to $H(\boldsymbol{p})$ and $J(\boldsymbol{p})$, respectively, when $\psi(x) = x \log x$ and $\psi(x) = -(1-x)\log(1-x)$.

The rest of this paper is organized as follows. The corresponding results R1, R2 and R3 for extropy are presented in Sections 2–4, respectively. In Section 2, we also discuss the discontinuity of extropy on Γ_∞ with respect to the variational distance and obtain bounds on the derivative of extropy. The proofs of the main results in Section 3 are put in the Appendix, as well as the definition of majorization order.

2. BOUNDS ON EXTROPY DIFFERENCE WITH ONE DISTRIBUTION FIXED

2.1. Upper Bound

Let $0 < \epsilon < 2$ and $\boldsymbol{p} = (p_1, p_2, \ldots, p_n) \in \Gamma_n$ with finite $n > 1$. Assume that the p_i are arranged in descending order, that is, $p_1 \geq p_2 \geq \cdots \geq p_n$. Let μ and ν be real numbers

https://doi.org/10.1017/S0269964818000098 Published online by Cambridge University Press

such that

$$\sum_{i=1}^{n}(p_i-\mu)_+=\frac{\epsilon}{2} \tag{2.1}$$

and

$$\sum_{i=1}^{n}(\nu-p_i)_+=\frac{\epsilon}{2}, \tag{2.2}$$

where $x_+=\max\{x,0\}$ for any $x\in\mathbb{R}$. If $\nu\geq\mu$, define $r_i=1/n$ for each i; and if $\nu<\mu$, define

$$r_i=\begin{cases}\mu, & \text{if } p_i>\mu,\\ p_i, & \text{if } \nu\leq p_i\leq\mu,\\ \nu, & \text{if } p_i<\nu,\end{cases} \tag{2.3}$$

for each i. Denote $\boldsymbol{r}(\boldsymbol{p},n,\epsilon)=(r_1,r_2,\ldots,r_n)$. Clearly, it can be checked that $\boldsymbol{r}(\boldsymbol{p},n,\epsilon)\in\Gamma_n$. Ho and Yeung [3] proved that

- $V(\boldsymbol{p},\boldsymbol{r}(\boldsymbol{p},n,\epsilon))\leq\epsilon$. Furthermore, if $0<\epsilon\leq\sum_{i=1}^{n}|p_i-1/n|$, then $V(\boldsymbol{p},\boldsymbol{r}(\boldsymbol{p},n,\epsilon))=\epsilon$;
- If $V(\boldsymbol{p},\boldsymbol{q})\leq\epsilon$, then $\boldsymbol{r}(\boldsymbol{p},n,\epsilon)\preceq_{\mathrm{m}}\boldsymbol{q}$;
- For $0<\epsilon_1<\epsilon_2$, $\boldsymbol{r}(\boldsymbol{p},n,\epsilon_2)\preceq_{\mathrm{m}}\boldsymbol{r}(\boldsymbol{p},n,\epsilon_1)$.

Here $\preceq_{\mathrm{m}}$ denotes the majorization order. The formal definitions of the majorization order and Schur-convex [Schur-concave] functions are given in Appendix A.1.

In view of these observations and the fact that the extropy $J(\boldsymbol{p})$ is Schur-concave in $\boldsymbol{p}$, we conclude the following result.

THEOREM 2.1: *Let $\boldsymbol{p}\in\Gamma_n$ with finite $n>1$. Then, for $\epsilon\in(0,2)$,*

$$\max_{\boldsymbol{q}\in\Gamma_n:V(\boldsymbol{p},\boldsymbol{q})\leq\epsilon}J(\boldsymbol{q})=J(\boldsymbol{r}(\boldsymbol{p},n,\epsilon)).$$

Remark 2.2: Let $\boldsymbol{p},\boldsymbol{q}\in\Gamma_n$ with finite $n>1$. Define

$$\boldsymbol{p}^*=\left(\frac{1-p_1}{n-1},\frac{1-p_2}{n-1},\ldots,\frac{1-p_n}{n-1}\right)$$

and

$$\boldsymbol{q}^*=\left(\frac{1-q_1}{n-1},\frac{1-q_2}{n-1},\ldots,\frac{1-q_n}{n-1}\right).$$

Then $\boldsymbol{p}^*,\boldsymbol{q}^*\in\Gamma_n$ and

$$V(\boldsymbol{p},\boldsymbol{q})=(n-1)V(\boldsymbol{p}^*,\boldsymbol{q}^*).$$

From (1.2), it follows that

$$\max_{\boldsymbol{q}:V(\boldsymbol{p},\boldsymbol{q})\leq\epsilon}J(\boldsymbol{q})=(n-1)\left[\max_{\boldsymbol{q}^*:V(\boldsymbol{p}^*,\boldsymbol{q}^*)\leq\epsilon/(n-1)}H(\boldsymbol{q}^*)-\log(n-1)\right], \tag{2.4}$$

where the probability masses in $\boldsymbol{p}^*$ and $\boldsymbol{q}^*$ are less than or equal to $1/(n-1)$. Therefore, Theorem 2.1 can not deduced from Theorem 2 in Ho and Yeung [3] immediately.

https://doi.org/10.1017/S0269964818000098 Published online by Cambridge University Press

Define

$$\Psi_{\boldsymbol{p},n}(\epsilon) = J(\boldsymbol{r}(\boldsymbol{p}, n, \epsilon)) \tag{2.5}$$

From Lemma 1 in Ho and Yeung [3], it follows that $\Psi_{\boldsymbol{p},n}(\epsilon)$ is concave, continuous and strictly increasing in ϵ, where $0 < \epsilon \le \sum_{i=1}^{n} |p_i - 1/n|$.

2.2. Lower Bound

Suppose $\epsilon \in (0, 2)$ and $\boldsymbol{p} \in \Gamma_n$, where n can be infinity. Assume that the p_i are in descending order. If $1 - p_1 \le \epsilon/2$, let $\boldsymbol{s}(\boldsymbol{p}, n, \epsilon) = (1, 0, \ldots, 0) \in \Gamma_n$; if $1 - p_1 > \epsilon/2$, let k be the largest integer such that

$$\sum_{i=k}^{n} p_i \ge \frac{\epsilon}{2},$$

and denote $\boldsymbol{s}(\boldsymbol{p}, n, \epsilon) = (s_1, s_2, \ldots, s_n)$, where

$$s_i = \begin{cases} p_1 + \epsilon/2, & \text{if } \ i = 1, \\ p_i, & \text{if } \ 1 < i < k, \\ \sum_{j=k}^{n} p_j - \epsilon/2, & \text{if } \ i = k, \\ 0, & \text{if } \ i > k. \end{cases} \tag{2.6}$$

Ho and Yeung [3] proved that

- $V(\boldsymbol{p}, \boldsymbol{s}(\boldsymbol{p}, n, \epsilon)) \le \epsilon$. Furthermore, for $0 < \epsilon \le 2(1 - p_1)$, we have $V(\boldsymbol{p}, \boldsymbol{s}(\boldsymbol{p}, n, \epsilon)) = \epsilon$.
- If $V(\boldsymbol{p}, \boldsymbol{q}) \le \epsilon$, then $\boldsymbol{q} \preceq_{\mathrm{m}} \boldsymbol{s}(\boldsymbol{p}, n, \epsilon)$.
- For $\epsilon_1 < \epsilon_2$, $\boldsymbol{s}(\boldsymbol{p}, n, \epsilon_1) \preceq_{\mathrm{m}} \boldsymbol{s}(\boldsymbol{p}, n, \epsilon_2)$.

In view of these observations and the fact that the extropy $J(\boldsymbol{p})$ is Schur-concave in $\boldsymbol{p}$, we conclude the following result.

THEOREM 2.3: *Let $\boldsymbol{p} \in \Gamma_n$ with $1 < n \le +\infty$. Then, for $\epsilon \in (0, 2)$,*

$$\min_{\boldsymbol{q} \in \Gamma_n : V(\boldsymbol{p}, \boldsymbol{q}) \le \epsilon} J(\boldsymbol{q}) = J(\boldsymbol{s}(\boldsymbol{p}, n, \epsilon)).$$

Note that $J(\boldsymbol{s}(\boldsymbol{p}, n, \epsilon))$ depends only on $\boldsymbol{p}$ and ϵ, which does not depend on n. So define

$$\Phi_{\boldsymbol{p}}(\epsilon) = J(\boldsymbol{s}(\boldsymbol{p}, n, \epsilon)).$$

Then $\Phi_{\boldsymbol{p}}(\epsilon)$ is continuous and strictly decreasing in ϵ, where $0 < \epsilon \le 2(1 - p_1)$.

Remark 2.4: By Theorems 2.1 and 2.3, we can derive the upper bound of the following quantities on the extropy difference when one distribution is fixed. For any given $\boldsymbol{p} \in \Gamma_n$ and $\epsilon \in (0, 2)$,

$$\begin{aligned} &\sup_{\boldsymbol{q} \in \Gamma_n : V(\boldsymbol{p}, \boldsymbol{q}) \le \epsilon} |J(\boldsymbol{q}) - J(\boldsymbol{p})| \\ &\quad = \max \left\{ \sup_{\boldsymbol{q} \in \Gamma_n : V(\boldsymbol{p}, \boldsymbol{q}) \le \epsilon} [J(\boldsymbol{q}) - J(\boldsymbol{p})], \sup_{\boldsymbol{q} \in \Gamma_n : V(\boldsymbol{p}, \boldsymbol{q}) \le \epsilon} [J(\boldsymbol{p}) - J(\boldsymbol{q})] \right\} \\ &\quad = \max\{\Psi_{\boldsymbol{p},n}(\epsilon) - J(p),\ J(p) - \Phi_{\boldsymbol{p}}(\epsilon)\}. \end{aligned}$$

https://doi.org/10.1017/S0269964818000098 Published online by Cambridge University Press

2.3. Discontinuity

Obviously, for any finite n, $J(\cdot)$ is continuous on Γ_n with respect to the variational distance. However, the extropy $J(\cdot)$ is not continuous on Γ_∞ with respect to the variation distance, as shown by the following theorem.

THEOREM 2.5: *Suppose that $\delta > 0$ and $\epsilon > 0$ are given arbitrarily. For any pmf $\boldsymbol{p} \in \Gamma_n$, there exists a sufficiently large integer $m > n$ and a pmf $\boldsymbol{q} \in \Gamma_m$ such that the variational distance $V(\boldsymbol{p}, \boldsymbol{q}) < \epsilon$, but $J(\boldsymbol{q}) - J(\boldsymbol{p}) > \delta$.*

PROOF: We use an idea from the proof of Theorem 1 in Ho and Yeung [3]. Choose an integer $m > n$, and define a pmf

$$\boldsymbol{q} = \left(p_1 - \frac{p_1}{\sqrt{\log m}},\ p_2 + \frac{p_1}{m\sqrt{\log m}},\ \ldots,\ p_n + \frac{p_1}{m\sqrt{\log m}},\ \frac{p_1}{m\sqrt{\log m}},\ \ldots,\ \frac{p_1}{m\sqrt{\log m}}\right)$$

with $m + 1$ probability masses. It is obvious that

$$V(\boldsymbol{p}, \boldsymbol{q}) = 2\frac{p_1}{\sqrt{\log m}}$$

and

$$J(\boldsymbol{q}) - J(\boldsymbol{p}) = \zeta_1(\boldsymbol{p}, m) + \zeta_2(\boldsymbol{p}, m) + \zeta_3(\boldsymbol{p}, m),$$

where

$$\zeta_1(\boldsymbol{p}, m) = -\left(1 - p_1 + \frac{p_1}{\sqrt{\log m}}\right)\log\left(1 - p_1 + \frac{p_1}{\sqrt{\log m}}\right) + (1 - p_1)\log(1 - p_1),$$

$$\zeta_2(\boldsymbol{p}, m) = -\sum_{i=2}^{n}\left[\left(1 - p_i - \frac{p_1}{\sqrt{n\log n}}\right)\log\left(1 - p_i - \frac{p_1}{\sqrt{m\log m}}\right) - (1 - p_i)\log(1 - p_i)\right],$$

and

$$\zeta_3(\boldsymbol{p}, m) = (m + 1 - n)(1 - \frac{p_1}{\sqrt{m\log m}})\log(1 - \frac{p_1}{\sqrt{m\log m}}).$$

It is easy to verify that

$$\lim_{m\to\infty} V(\boldsymbol{p}, \boldsymbol{q}) = 0, \qquad \lim_{m\to\infty} \zeta_1(\boldsymbol{p}, m) = 0, \qquad \lim_{m\to\infty} \zeta_2(\boldsymbol{p}, m) = 0,$$

and

$$\lim_{m\to\infty} \zeta_3(\boldsymbol{p}, m) = +\infty.$$

Hence, for any $\epsilon > 0$ and $\delta > 0$, we have $V(\boldsymbol{p}, \boldsymbol{q}) < \epsilon$ but $J(\boldsymbol{q}) - J(\boldsymbol{p}) > \delta$ when m is large enough. ■

It has been established in Topsøe [9] that entropy is lower-semi continuous with respect to pointwise convergence of pmfs. A simple proof of this lower-semi continuity of entropy was presented in Ho and Yeung [3]. A careful check finds that the argument in the proof of Corollary 1 in Ho and Yeung [3] can also apply to the extropy by exploiting Theorem 2.3. We state the corresponding result as follows.

PROPOSITION 2.6: *For any $\boldsymbol{p} \in \Gamma_\infty$ with $J(\boldsymbol{p}) < \infty$, $J(\cdot)$ is lower-semi continuous at $\boldsymbol{p}$ with respect to pointwise convergence, that is, if a sequence of pmfs $\boldsymbol{p}^{(k)}$ which converges to $\boldsymbol{p}$ pointwisely, then*

$$\liminf_{k\to\infty} J\left(\boldsymbol{p}^{(k)}\right) \geq J(\boldsymbol{p}).$$

https://doi.org/10.1017/S0269964818000098 Published online by Cambridge University Press

2.4. Bounds on the Derivative of Extropy

Next, we will consider the derivative of $J(\boldsymbol{p})$, $\boldsymbol{p} \in \Gamma_n$. Let $\boldsymbol{t} = (t_1.t_2, \ldots, t_n) \in \mathbb{R}^n$ such that $\sum_{i=1}^n t_i = 0$ and $\sum_{i=1}^n |t_i| = 1$. For $\boldsymbol{t}$ such that $(\boldsymbol{p} + \epsilon \cdot \boldsymbol{t}) \in \Gamma_n$ when ϵ is small enough, define

$$D_{\boldsymbol{t}} J(\boldsymbol{p}) = \lim_{\epsilon \to 0} \frac{J(\boldsymbol{p} + \epsilon \cdot \boldsymbol{t}) - J(\boldsymbol{p})}{\epsilon}.$$

THEOREM 2.7: *For any pmf* $\boldsymbol{p} = (p_1, p_2, \ldots, p_n) \in \Gamma_n$ *where* $p_1 \geq p_2 \geq \cdots \geq p_n > 0$ *and* n *is finite, if* $D_{\boldsymbol{t}} J(\boldsymbol{p})$ *is defined, then*

$$\frac{1}{2} \log \frac{1 - p_1}{1 - p_n} \leq D_{\boldsymbol{t}} J(\boldsymbol{p}) \leq \frac{1}{2} \log \frac{1 - p_n}{1 - p_1}. \tag{2.7}$$

PROOF: The proof parallels to that of Theorem 5 in Ho and Yeung [3] concerning the bounds on derivative of entropy. We first prove the upper bound. If $\boldsymbol{p} = (1/n, 1/n, \ldots, 1/n) \in \Gamma_n$, then $J(\cdot)$ attains its maximum at $\boldsymbol{p}$ and, hence, $D_{\boldsymbol{t}} J(\boldsymbol{p}) \leq 0$ for all $\boldsymbol{t}$. In this case, the upper bound is verified. In the following, we assume $\boldsymbol{p} \neq (1/n, 1/n, \ldots, 1/n)$ and prove the upper bound. Consider only $\boldsymbol{t}$ such that $\boldsymbol{p} + \epsilon \cdot \boldsymbol{t} \in \Gamma_n$ when ϵ is small enough. Note that

$$V(\boldsymbol{p}, \boldsymbol{p} + \epsilon \cdot \boldsymbol{t}) = \sum_{i=1}^n |p_i - p_i - \epsilon t_i| = \epsilon \sum_{i=1}^n |r_i| = \epsilon.$$

Then

$$D_{\boldsymbol{t}} J(\boldsymbol{p}) = \lim_{\epsilon \to 0} \frac{J(\boldsymbol{p} + \epsilon \cdot \boldsymbol{t}) - J(\boldsymbol{p})}{\epsilon} \leq \lim_{\epsilon \to 0} \frac{J(\boldsymbol{r}(\boldsymbol{p}, n, \epsilon)) - J(\boldsymbol{p})}{\epsilon},$$

where $\boldsymbol{r}(\boldsymbol{p}, n, \epsilon)$ is defined by (2.3). Denote

$$v = \#\{i : p_i = p_1 . 1 \leq i \leq n\} \quad \text{and} \quad w = \#\{i : p_i = p_n, 1 \leq i \leq n\}.$$

Choose ϵ such that $0 < \epsilon \leq \sum_{i=1}^n |p_i - 1/n|$ and $\epsilon \leq \min\{p_i - p_j : p_i - p_j > 0, 1 \leq i < j \leq n\}$. Then $\boldsymbol{r}(\boldsymbol{p}, n, \epsilon)$ can be taken

$$r_i = \begin{cases} p_1 - \epsilon/(2v) & \text{if } \ i \leq v, \\ p_i, & \text{if } \ v < i \leq n - w, \\ p_n + \epsilon/(2w), & \text{if } \ n - w < i \leq n. \end{cases}$$

By L'Hôpital's rule, we have

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[-v \left(1 - p_1 + \frac{\epsilon}{2v}\right) \log \left(1 - p_1 + \frac{\epsilon}{2v}\right) + v(1 - p_1) \log(1 - p_1) \right] = -\frac{1}{2} \log(1 - p_1) - \frac{1}{2}$$

and

$$\begin{aligned} &\lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[-w \left(1 - p_n - \frac{\epsilon}{2w}\right) \log \left(1 - p_n - \frac{\epsilon}{2w}\right) + w(1 - p_n) \log(1 - p_n) \right] \\ &\quad = \frac{1}{2} \log(1 - p_n) + \frac{1}{2}. \end{aligned}$$

Hence,

$$D_{\boldsymbol{t}} J(\boldsymbol{p}) \leq \frac{1}{2} \log \frac{1 - p_n}{1 - p_1},$$

which proves the upper bound in (2.7).

https://doi.org/10.1017/S0269964818000098 Published online by Cambridge University Press

Now, we prove the lower bound of (2.7). If $\boldsymbol{p} = (1, 0, \ldots, 0) \in \Gamma_n$, then $J(\cdot)$ attains its minimum at $\boldsymbol{p}$ and $D_{\boldsymbol{t}} J(\boldsymbol{p}) \geq 0$ for all $\boldsymbol{t}$. In this case, the lower bound (2.7) is verified. In the following, we assume $\boldsymbol{p} \neq (1, 0, \ldots, 0)$. Then

$$D_{\boldsymbol{t}} J(\boldsymbol{p}) = \lim_{\epsilon \to 0} \frac{J(\boldsymbol{p} + \epsilon \cdot \boldsymbol{t}) - J(\boldsymbol{p})}{\epsilon} \geq \lim_{\epsilon \to 0} \frac{J(\boldsymbol{s}(\boldsymbol{p}, n, \epsilon)) - J(\boldsymbol{p})}{\epsilon},$$

where $\boldsymbol{s}(\boldsymbol{p}, n, \epsilon)$ is defined by (2.6). Choose $\epsilon \leq 2(1 - p_1)$ such that $p_n > \epsilon/2$. Therefore, $\boldsymbol{s}(\boldsymbol{p}, n, \epsilon)$ is

$$s_i = \begin{cases} p_1 + \epsilon/2, & \text{if } \ i = 1, \\ p_i, & \text{if } \ 1 < i \leq n-1, \\ p_n - \epsilon/2, & \text{if } \ i = n. \end{cases}$$

By L'Hôpital's rule, we have

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[-\left(1 - p_1 - \frac{\epsilon}{2}\right) \log \left(1 - p_1 - \frac{\epsilon}{2}\right) + (1 - p_1) \log(1 - p_1) \right] = \frac{1}{2} \log(1 - p_1),$$

and

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[-\left(1 - p_n + \frac{\epsilon}{2}\right) \log \left(1 - p_n + \frac{\epsilon}{2}\right) + (1 - p_n) \log(1 - p_n) \right] = -\frac{1}{2} \log(1 - p_n).$$

Hence,

$$D_{\boldsymbol{t}} J(\boldsymbol{p}) \geq \frac{1}{2} \log \frac{1 - p_1}{1 - p_n},$$

which proves the lower bound in (2.7). ■

3. BOUNDS ON EXTROPY DIFFERENCE WITH NO DISTRIBUTION FIXED

THEOREM 3.1: *If $\boldsymbol{p} \in \Gamma_\ell$ and $\boldsymbol{q} \in \Gamma_m$ with $m \geq \ell$ are two pmfs such that $V(\boldsymbol{p}, \boldsymbol{q}) \leq \epsilon < 2$, and the probability masses in $\boldsymbol{p}$ and $\boldsymbol{q}$ are less than or equal to $1/n$ for a positive integer $n \leq \ell$, then*

$$\max_{\boldsymbol{p} \in \Gamma_\ell, \boldsymbol{q} \in \Gamma_m : V(\boldsymbol{p}, \boldsymbol{q}) \leq \epsilon} |J(\boldsymbol{q}) - J(\boldsymbol{p})| = \begin{cases} g_1(m, n, \epsilon), & \text{if } \ \epsilon < \dfrac{2(m-n)}{m}, \\ g_2(m, n, \epsilon), & \text{if } \ \epsilon \geq \dfrac{2(m-n)}{m}, \end{cases} \tag{3.1}$$

where

$$g_1(m, n, \epsilon) = n f\left(\frac{n-1}{n}, \frac{\epsilon}{2n}\right) + (m - n) f\left(1, -\frac{\epsilon}{2(m-n)}\right),$$

$$g_2(m, n, \epsilon) = (1 - n) \log\left(\frac{n-1}{n}\right) - (m - 1) \log\left(\frac{m-1}{m}\right),$$

and $f(p, d) = p \log p - (p + d) \log(p + d)$. For a fixed ϵ, the above maximum is attained by

$$\boldsymbol{p} = \left(\underbrace{\frac{1}{n}, \ \ldots, \ \frac{1}{n}}_{n}, \ 0, \ \ldots, \ 0 \right) \in \Gamma_\ell \tag{3.2}$$

https://doi.org/10.1017/S0269964818000098 Published online by Cambridge University Press

and $\boldsymbol{q} = (q_1, q_2, \ldots, q_m)$ *such that when* $\epsilon < 2(m-n)/m$,

$$q_i = \begin{cases} \dfrac{2-\epsilon}{2n}, & \text{for } 1 \le i \le n, \\ \dfrac{\epsilon}{2(m-n)}, & \text{for } n+1 \le i \le m. \end{cases} \tag{3.3}$$

Otherwise, $\boldsymbol{q} = (1/m, 1/m, \ldots, 1/m) \in \Gamma_m$.

The proof of Theorem 3.1 is postponed in Appendix A.2. The argument in the proof of Theorem 3.1 can be applied to simplify the proof of Theorem 7 in Ho and Yeung [3]. An immediate consequence of Theorem 3.1 with $n = 1$ is as follows.

COROLLARY 3.2: *Let* $m \ge \ell \ge 1$ *be integers and* $0 < \epsilon < 2$. *Then*

$$\max_{\boldsymbol{p}\in\Gamma_\ell, \boldsymbol{q}\in\Gamma_m : V(\boldsymbol{p},\boldsymbol{q}) \le \epsilon} |J(\boldsymbol{q}) - J(\boldsymbol{p})|$$
$$= \begin{cases} -\dfrac{\epsilon}{2}\log\dfrac{\epsilon}{2} - \left(m - 1 - \dfrac{\epsilon}{2}\right)\log\left(1 - \dfrac{\epsilon}{2(m-1)}\right), & \text{for } 0 \le \epsilon < \dfrac{2(m-1)}{m}, \\ (m-1)\log\dfrac{m}{m-1}, & \text{for } \epsilon \ge \dfrac{2(m-1)}{m}. \end{cases}$$

For a fixed ϵ, *the above maximum is attained by* $\boldsymbol{p} = (1, 0, \ldots, 0) \in \Gamma_\ell$ *and*

$$\boldsymbol{q} = \begin{cases} \left(1 - \dfrac{\epsilon}{2}, \dfrac{\epsilon}{2(m-1)}, \ldots, \dfrac{\epsilon}{2(m-1)}\right), & \text{for } 0 < \epsilon < \dfrac{2(m-1)}{m}, \\ \left(\dfrac{1}{m}, \dfrac{1}{m}, \ldots, \dfrac{1}{m}\right), & \text{for } \epsilon \ge \dfrac{2(m-1)}{m}. \end{cases}$$

Remark 3.3: For any fixed integers $1 \le \ell \le m$, by Corollary 3.2, we have

$$\lim_{\epsilon\to 0} \max_{\boldsymbol{p}\in\Gamma_\ell, \boldsymbol{q}\in\Gamma_m : V(\boldsymbol{p},\boldsymbol{q}) \le \epsilon} |J(\boldsymbol{q}) - J(\boldsymbol{p})|$$
$$= \lim_{\epsilon\to 0}\left[-\frac{\epsilon}{2}\log\frac{\epsilon}{2} - \left(m - 1 - \frac{\epsilon}{2}\right)\log\left(1 - \frac{\epsilon}{2(m-1)}\right)\right] = 0.$$

This means that for any finite n, $J(\cdot)$ is continuous on Γ_n with respect to the variational distance.

Corollary 3.2 gives the upper bound of

$$\max_{\boldsymbol{q}\in\Gamma_m : V(\boldsymbol{p},\boldsymbol{q}) \le \epsilon} |J(\boldsymbol{q}) - J(\boldsymbol{p})| \tag{3.4}$$

for all $\boldsymbol{p} \in \Gamma_\ell$. The next theorem gives the lower bound of (3.4) for all $\boldsymbol{p} \in \Gamma_\ell$. The proof of Theorem 3.4 is postponed in Appendix A.3.

https://doi.org/10.1017/S0269964818000098 Published online by Cambridge University Press

THEOREM 3.4: *Let $m \geq \ell \geq 1$ be two integers and $\epsilon \in (0, 2)$. Then*

$$\min_{\boldsymbol{p}\in\Gamma_\ell} \max_{\boldsymbol{q}\in\Gamma_m : V(\boldsymbol{p},\boldsymbol{q})\leq\epsilon} \big[J(\boldsymbol{q}) - J(\boldsymbol{p})\big] = \begin{cases} h_1(\ell, m, \epsilon), & 0 < \epsilon \leq \dfrac{2(m-\ell)}{m}, \\ h_2(\ell, m, \epsilon), & \epsilon > \dfrac{2(m-\ell)}{m}, \end{cases} \tag{3.5}$$

where

$$h_1(\ell, m, \epsilon) = \ell f\left(1 - \frac{1}{\ell}, \frac{\epsilon}{2\ell}\right) + (m-\ell) f\left(1, -\frac{\epsilon}{2(m-\ell)}\right), \tag{3.6}$$

$$h_2(\ell, m, \epsilon) = (1-m)\log\left(1 - \frac{1}{m}\right) - (1-\ell)\log\left(1 - \frac{1}{\ell}\right), \tag{3.7}$$

and $f(p, d) = p\log p - (p+d)\log(p+d)$.

In the end of this section, we extend Corollary 3.2 to generalized pmfs. $\tilde{\boldsymbol{p}} = (\tilde{p}_1, \tilde{p}_2, \ldots, \tilde{p}_n)$ is said to be a generalized pmf if $\tilde{p}_i \geq 0$ for each i and $\sum_{i=1}^n \tilde{p}_i \leq 1$. Define

$$\tilde{\Gamma}_n = \{\tilde{\boldsymbol{p}} = (\tilde{p}_1, \tilde{p}_2, \ldots, \tilde{p}_n) : \ \tilde{\boldsymbol{p}} \text{ is a generalized pmf}\}.$$

The extropy of $\tilde{\boldsymbol{p}}$ is defined by

$$J(\tilde{\boldsymbol{p}}) = -\sum_{i=1}^n (1-p_i)\log(1-p_i).$$

Denote $\alpha = \sum_{i=1}^n p_i$. Then $(\alpha/n, \alpha/n, \ldots, \alpha/n) \preceq_{\mathrm{m}} \tilde{\boldsymbol{p}}$. Thus,

$$J(\tilde{\boldsymbol{p}}) \leq J\left(\frac{\alpha}{n}, \frac{\alpha}{n}, \ldots, \frac{\alpha}{n}\right) = -(n-\alpha)\log\frac{n-\alpha}{n} \leq -(n-1)\log\frac{n-1}{n}. \tag{3.8}$$

For generalized pmfs, we have an analogous result to Corollary 3.2. The proof of Theorem 3.5 is postponed in Appendix A.4.

THEOREM 3.5: *Let $m \geq \ell \geq 1$ be two integers and $0 < \epsilon < \min\{m/e, 1\}$. Then*

$$\sup_{\tilde{\boldsymbol{p}}\in\tilde{\Gamma}_\ell, \tilde{\boldsymbol{q}}\in\tilde{\Gamma}_m : V(\tilde{\boldsymbol{p}},\tilde{\boldsymbol{q}})\leq\epsilon} \big|J(\tilde{\boldsymbol{p}}) - J(\tilde{\boldsymbol{q}})\big| \leq -\epsilon\log\epsilon + \epsilon\log m. \tag{3.9}$$

4. APPLICATIONS

In this section, we give one application of the main results about the bounds on extropy obtained in Sections 2 and 3, that is, determining the confidence interval for $J(\boldsymbol{p})$ when $\boldsymbol{p} \in \Gamma_m$ is unknown.

Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with generic pmf $\boldsymbol{p} \in \Gamma_m$. Denote by $\hat{\boldsymbol{p}}(\boldsymbol{X}) = (\hat{p}_1(\boldsymbol{X}), \hat{p}_2(\boldsymbol{X}), \ldots, \hat{p}_m(\boldsymbol{X}))$ the empirical pmf in terms of the random sample $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$, where

$$\hat{p}_i(\boldsymbol{X}) = N_i/n \quad \text{and} \quad N_i = \#\{k : X_k = i, 1 \leq k \leq n\}$$

for $i = 1, \ldots, m$. Here, without loss of generality, we assume that the alphabet $\mathcal{X} = \{1, 2, \ldots, m\}$. Ho and Yeung [3] derived a formula for the confidence interval (CI) of entropy

https://doi.org/10.1017/S0269964818000098 Published online by Cambridge University Press

estimation, that is, obtaining an analytic bound on

$$\mathbb{P}\big(|H(\hat{\boldsymbol{p}}(\boldsymbol{X}) - H(\boldsymbol{p})| > \gamma\big),$$

where $\gamma > 0$ is fixed. All arguments in Section IV of Ho and Yeung [3] are also valid for extropy. We state all the corresponding results for extropy without proofs. The following lemma plays an important role in deriving the CI for $H(\boldsymbol{p})$ and $J(\boldsymbol{p})$ in terms of $\hat{\boldsymbol{p}}(\boldsymbol{X})$.

LEMMA 4.1 (Weissman et al. [11]): *For $\boldsymbol{p} \in \Gamma_m$ and $\epsilon > 0$,*

$$\mathbb{P}(V(\boldsymbol{p}, \hat{\boldsymbol{p}}(\boldsymbol{X})) \le \epsilon) \ge 1 - (2^m - 2)\exp\left\{-\frac{n}{2}\epsilon^2\right\}.$$

THEOREM 4.2: *Let $\boldsymbol{p} \in \Gamma_m$ be unknown with m finite.*

(i) *If $V(\hat{\boldsymbol{p}}(\boldsymbol{X}), \boldsymbol{p}) \le \epsilon$, then*

$$J\big(\boldsymbol{s}(\hat{\boldsymbol{p}}(\boldsymbol{X}), m, \epsilon)\big) \le J(\boldsymbol{p}) \le J\big(\boldsymbol{r}(\hat{\boldsymbol{p}}(\boldsymbol{X}), m, \epsilon)\big).$$

(ii) *For any $\alpha \in (0, 1)$,*

$$\mathbb{P}(J(\boldsymbol{s}(\hat{\boldsymbol{p}}(\boldsymbol{X}), m, \epsilon)) \le J(\boldsymbol{p}) \le J(\boldsymbol{r}(\hat{\boldsymbol{p}}(\boldsymbol{X}), m, \epsilon))) \ge 1 - \alpha,$$

where

$$\epsilon = \sqrt{\frac{2}{n}\log\frac{2^m - 2}{\alpha}}.$$

(iii) *For any $0 < \gamma < (1 - m)\log(1 - 1/m)$, let $\epsilon > 0$ be the smallest root of*

$$\gamma = -\frac{\epsilon}{2}\log\frac{\epsilon}{2} - \left(m - 1 - \frac{\epsilon}{2}\right)\log\left(1 - \frac{\epsilon}{2(m-1)}\right). \tag{4.1}$$

Then

$$\mathbb{P}(|J(\hat{\boldsymbol{p}}(\boldsymbol{X})) - J(\boldsymbol{p})| > \gamma) \le (2^m - 2)\exp\left\{-\frac{n}{2}\epsilon^2\right\}.$$

(iv) *For any $0 < \gamma < (1 - m)\log(1 - 1/m)$, let $\epsilon > 0$ be the smallest root of (4.1). For $0 < \alpha < 1$, choose*

$$n \ge \frac{2}{\epsilon^2}\log\frac{2^m - 2}{\alpha}.$$

Then

$$\mathbb{P}(|J(\hat{\boldsymbol{p}}(\boldsymbol{X})) - J(\boldsymbol{p})| \le \gamma) \ge 1 - \alpha.$$

Part (iv) of Theorem 4.2 can be used to determine the sample size n such that the CI for $J(\boldsymbol{p})$ is small enough. It is worth pointing out that it is an interesting problem to compare CIs determined by Theorem 4.2, the Markov Chain Monte Carlo technique and other methods. We will present such a comparison study in the near future.

Acknowledgments

J. Yang was supported by the NNSF of China (No. 11701518) and Zhejiang Provincial Natural Science Foundation (No. LQ17A010011). T. Hu was supported by the NNSF of China (Nos. 71671176, 11371340, 71631006).

https://doi.org/10.1017/S0269964818000098 Published online by Cambridge University Press

References

1. Barbour, A.D., Johnson, O., Kontoyiannis, I., & Madiman, M. (2010). Compound Poisson approximation via information functionals. *Electronic Journal of Probability*, 15: paper no 42, 1344–1368.
2. Cover, T.M. & Thomas, J.A. (2006). *Elements of Information Theory*, 2nd ed. New York: Wiley-Interscience.
3. Ho, S.-W. & Yeung, R.W. (2010). The interplay between entropy and variational distance. *IEEE Transactions on Information Theory* 56(12): 5906–5929.
4. Lad, F., Sanfilippo, G., & Agrò, G. (2015). Extropy: Complementary dual of entropy. *Statistical Science* 30(1): 40–58.
5. Ley, C. & Swan, Y. (2013). Local Pinsker inequalities via Stein's discrete density approach. *IEEE Transactions on Information Theory* 59(9): 5584–5591.
6. Marshall, A.W., Olkin, I., & Arnold, B.C. (2011). *Inequalities: Theory of Majorization and Its Applications*, 2nd ed. New York: Springer.
7. Sason, I. & Verdu, S. (2015) Upper bounds on the relative entropy and Rényi divergence as a function of total variation distance for finite alphabets. Proceedings of the 2015 IEEE Information Theory Workshop, Jeju, Korea, October 2015, pp. 214–218.
8. Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal* 27: 379–423 and 623–656.
9. Topsøe, F. (2001). Basic concepts, identities and inequalities—The toolkit of information theory. *Entropy* 3: 162–190.
10. Verdú, S. (2014) Total variation distance and the distribution of relative information. *Information Theory and Applications Workshop (ITA)*, San Diego, CA, USA, February 2014, pp. 1–3.
11. Weissman, T., Ordentlich, E., Seroussi, G., Verdú, S., & Weinberger, M. (2005). Universal discrete denoising: Known chanell. *IEEE Transactions on Information Theory* 51(1): 5–28.

APPENDIX A

A.1. Majorization

The majorization order enables us to compare the diversity of two real vectors (see Chapter 1 in Marshall et al. [6]). For fixed $n \geq 2$, let $\boldsymbol{a} = (a_1, a_2, \ldots, a_n)$ and $\boldsymbol{b} = (b_1, b_2, \ldots, b_n)$ be two real vectors, and denote by $a_{[1]} \geq a_{[2]} \geq \cdots \geq a_{[n]}$ and $b_{[1]} \geq b_{[2]} \geq \cdots \geq b_{[n]}$ the ordered components of $\boldsymbol{a}$ and $\boldsymbol{b}$, respectively. For $\boldsymbol{a}$ and $\boldsymbol{b} \in \mathbb{R}^n$, $\boldsymbol{a}$ is said to be majorized by $\boldsymbol{b}$, denoted by $\boldsymbol{a} \preceq_{\mathrm{m}} \boldsymbol{b}$, if

$$\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} b_i \quad \text{and} \quad \sum_{i=1}^{k} a_{[i]} \leq \sum_{i=1}^{k} b_{[i]} \quad \text{for } k = 1, \ldots, n-1.$$

A function $\phi : \mathbb{R}^n \to \mathbb{R}$ is said to be Schur-convex [Schur-concave] if $\boldsymbol{x} \preceq_{\mathrm{m}} \boldsymbol{y}$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ imply that $\phi(\boldsymbol{x}) \leq [\geq] \phi(\boldsymbol{y})$.

Note that a Schur function is necessarily permutation-invariant. A necessary and sufficient condition for a permutation symmetric differentiable function to be Schur-convex is that

$$\left(\frac{\partial \phi(\boldsymbol{x})}{\partial x_i} - \frac{\partial \phi(\boldsymbol{x})}{\partial x_j} \right) (x_i - x_j) \geq 0 \quad \text{for all } \boldsymbol{x} \text{ and } i \neq j.$$

The inequality is reversed for Schur-concave functions.

The following characterization of the majorization order is well-known; see Propositions 3.C.1 and 4.B.1 in Marshall et al. [6]. Let $\mathbb{I}$ be an interval of $\mathbb{R}$. Then the inequality $\sum_{i=1}^{n} \psi(x_i) \leq \sum_{i=1}^{n} \psi(y_i)$ holds for all continuous convex functions $\psi : \mathbb{I} \to \mathbb{R}$ if and only if $\boldsymbol{x} \preceq_{\mathrm{m}} \boldsymbol{y}$ when $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{I}^n$.

A.2. Proof of Theorem 3.1

Proof: We use an idea from the proof of Theorem 7 in Ho and Yeung [3]. Consider two pmfs $\boldsymbol{p} = (p_1, p_2, \ldots, p_\ell)$ and $\boldsymbol{q} = (q_1, q_2, \ldots, q_m)$. Assume that the p_i are in descending order. Define

https://doi.org/10.1017/S0269964818000098 Published online by Cambridge University Press

two pmfs

$$\boldsymbol{\gamma} = \Big(\underbrace{\frac{1}{n}, \ldots, \frac{1}{n}}_{n}, \ 0, \ldots, 0 \Big) \in \Gamma_\ell, \qquad \boldsymbol{\eta} = \Big(\underbrace{\frac{1}{m}, \frac{1}{m}, \ldots, \frac{1}{m}}_{m} \Big) \in \Gamma_m.$$

Obviously,

$$V(\boldsymbol{\gamma}, \boldsymbol{\eta}) = \frac{2(m-n)}{m}.$$

Since $p_i \le 1/n$ for each i, it follows that $\boldsymbol{p} \preceq_{\mathrm{m}} \boldsymbol{\gamma}$. Observing that the extropy function $J(\cdot)$ is Schur-concave, we have

$$J(\boldsymbol{p}) \ge J(\boldsymbol{\gamma}) = -(n-1)\log\left(1 - \frac{1}{n}\right).$$

Also, $\boldsymbol{\eta} \preceq_{\mathrm{m}} \boldsymbol{q}$ implies that

$$J(\boldsymbol{q}) \le J(\boldsymbol{\eta}) = -(m-1)\log(1 - \frac{1}{m}).$$

Thus, when $\epsilon \ge 2(m-n)/m$,

$$|J(\boldsymbol{q}) - J(\boldsymbol{p})| \le J(\boldsymbol{\eta}) - J(\boldsymbol{\gamma}) = g_2(m, n, \epsilon).$$

In the following, we will consider the case that $0 < \epsilon < 2(m-n)/m$. Note that $\boldsymbol{\gamma}$ can be viewed as $\boldsymbol{\gamma} \in \Gamma_m$. From the theorem 2.1, we have

$$\begin{aligned} \max_{\boldsymbol{q} \in \Gamma_m, V(\boldsymbol{\gamma}, \boldsymbol{q}) \le \epsilon} |J(\boldsymbol{q}) - J(\boldsymbol{\gamma})| &= J(\boldsymbol{r}(\boldsymbol{\gamma}, m, \epsilon)) - J(\boldsymbol{\gamma}) \\ &= nf\left(1 - \frac{1}{n}, \frac{\epsilon}{2n}\right) + (m-n)f\left(1, -\frac{\epsilon}{2(m-n)}\right), \end{aligned} \tag{A.1}$$

where

$$\boldsymbol{r}(\boldsymbol{\gamma}, m, \epsilon) = \Big(\underbrace{\frac{2-\varepsilon}{2n}, \ \ldots, \ \frac{2-\varepsilon}{2n}}_{n}, \ \underbrace{\frac{\epsilon}{2(m-n)}, \ \ldots, \ \frac{\epsilon}{2(m-n)}}_{m-n} \Big), \tag{A.2}$$

and $V(\boldsymbol{r}(\boldsymbol{\gamma}, m, \epsilon), \boldsymbol{\gamma}) = \epsilon$. Similarly, for any $\boldsymbol{p} \in \Gamma_\ell$, we have

$$\max_{\boldsymbol{q} \in \Gamma_m : V(\boldsymbol{q}, \boldsymbol{p}) \le \epsilon} |J(\boldsymbol{q}) - J(\boldsymbol{p})| \le J(\boldsymbol{r}(\boldsymbol{p}, m, \epsilon)) - J(\boldsymbol{p}),$$

where $\boldsymbol{r}(\boldsymbol{p}, m, \epsilon)$ is specified by (2.3). In view of (A.1), it suffices to prove that

$$J(\boldsymbol{r}(\boldsymbol{p}, m, \epsilon)) - J(\boldsymbol{p}) \le J(\boldsymbol{r}(\gamma, m, \epsilon)) - J(\boldsymbol{\gamma}). \tag{A.3}$$

Let $\mu_{\boldsymbol{\gamma}}$ and $\mu_{\boldsymbol{p}}$ be the μ's when $\boldsymbol{r}(\boldsymbol{\gamma}, m, \epsilon)$ and $\boldsymbol{r}(\boldsymbol{p}, m, \epsilon)$ are evaluated, respectively. Then,

$$\mu_{\boldsymbol{\gamma}} = \frac{2-\epsilon}{2n}.$$

Under the assumption that $p_i \le 1/n$ for all i, we have $\mu_{\boldsymbol{p}} \le \mu_{\boldsymbol{\gamma}}$ (see Eq. (115) in Ho and Yeung [3]). Denote

$$\boldsymbol{r}(\boldsymbol{p}, m, \epsilon) = (r_1, r_2, \ldots, r_m) = (p_1 + d_1, \ldots, p_m + d_m).$$

Then

$$r_i \le \mu_{\boldsymbol{p}} \le \mu_{\boldsymbol{\gamma}} = \frac{2-\epsilon}{2n}$$

https://doi.org/10.1017/S0269964818000098 Published online by Cambridge University Press

for all i. Similarly, let $\nu_{\boldsymbol{\gamma}}$ and $\nu_{\boldsymbol{p}}$ be the ν's when $\boldsymbol{r}(\boldsymbol{\gamma}, m, \epsilon)$ and $\boldsymbol{r}(\boldsymbol{p}, m, \epsilon)$ are evaluated, respectively. Then, we have for all i,

$$r_i \geq \nu_{\boldsymbol{p}} \geq \frac{\epsilon}{2(m-n)}.$$

Note that

$$\begin{aligned} J(\boldsymbol{r}(\boldsymbol{p}, m, \epsilon)) - J(\boldsymbol{p}) &= \sum_{i:p_i > \mu_{\boldsymbol{p}}} f(1-p_i, -d_i) + \sum_{i:p_i < \nu_{\boldsymbol{p}}} f(1-p_i, -d_i) \\ &= \sum_{i=1}^{\tau} f(1-p_i, -d_i) + \sum_{i=k}^{m} f(1-p_i, -d_i), \end{aligned} \tag{A.4}$$

where $\tau = \max\{i : d_i < 0, 1 \leq i \leq m\}$ and $k = \min\{i : d_i > 0, 1 \leq i \leq n\}$, that is, $d_\tau < 0 \leq d_{\tau+1}$ and $d_{k-1} \leq 0 < d_k$. Observe that

- for all $1 \leq i \leq \tau$, we have

$$p_i + d_i < p_i \leq \frac{1}{n} \quad \text{and} \quad p_i + d_i < \frac{2-\epsilon}{2n};$$

- for all $k \leq i \leq m$, we have

$$p_i + d_i > p_i > 0 \quad \text{and} \quad p_i + d_i \geq \frac{\epsilon}{2(m-n)};$$

- and $\sum_{i=1}^{\tau} d_i = -\epsilon/2$ and $\sum_{i=k}^{m} d_i = \epsilon/2$.

In view of these observations, it can be checked that

$$\begin{aligned} &(1-p_1, \ldots, 1-p_\tau, \underbrace{1 - \frac{2-\epsilon}{2n}, \ldots, 1 - \frac{2-\epsilon}{2n}}_{n}) \\ &\quad \preceq_{\mathrm{m}} (1-p_1-d_1, \ldots, 1-p_\tau-d_\tau, \underbrace{1 - \frac{1}{n}, \ldots, 1 - \frac{1}{n}}_{n}) \end{aligned}$$

and

$$\begin{aligned} &(1-p_k, \ldots, 1-p_m, \underbrace{1 - \frac{\epsilon}{2(m-n)}, \ldots, 1 - \frac{\epsilon}{2(m-n)}}_{m-n}) \\ &\quad \preceq_{\mathrm{m}} (1-p_k-d_k, \ldots, 1-p_m-d_m, \underbrace{1, \ldots, 1}_{m-n}). \end{aligned}$$

Note that $x \log x$ is convex on $\mathbb{R}_+$. By Proposition 3.C.1 in Marshall et al. [6], it follows that

$$\begin{aligned} &\sum_{i=1}^{\tau} (1-p_i) \log(1-p_i) + n \left(1 - \frac{2-\epsilon}{2n}\right) \log \left(1 - \frac{2-\epsilon}{2n}\right) \\ &\quad \leq \sum_{i=1}^{\tau} (1-p_i-d_i) \log(1-p_i-d_i) + (n-1) \log \left(1 - \frac{1}{n}\right), \end{aligned} \tag{A.5}$$

https://doi.org/10.1017/S0269964818000098 Published online by Cambridge University Press

and

$$\sum_{i=k}^{m}(1-p_i)\log(1-p_i)+(m-n)\left(1-\frac{\epsilon}{2(m-n)}\right)\log\left(1-\frac{\epsilon}{2(m-n)}\right)$$
$$\leq \sum_{i=k}^{m}(1-p_i-d_i)\log(1-p_i-d_i). \tag{A.6}$$

Obviously, (A.5) and (A.6) can be written, respectively, as

$$\sum_{i=1}^{\tau} f(1-p_i,-d_i) \leq nf\left(1-\frac{1}{n},\frac{\epsilon}{2n}\right)$$

and

$$\sum_{i=k}^{m} f\,(1-p_i,-d_i) \leq (m-n)f(1,-\frac{\epsilon}{2(m-n)}).$$

From (A.1) and (A.4), we get (A.3). Also from (A.1), it follows that the maximum in the left-hand side of (3.1) is attained by $\boldsymbol{p}$ and $\boldsymbol{q}$ given by (3.2) and (3.3), respectively. This completes the proof of Theorem 3.1. ■

A.3. Proof of Theorem 3.4

Proof: The proof follows the line of the argument in the proof of Theorem 9 in Ho and Yeung [3]. First, we will show that the lower bound on

$$\sup_{\boldsymbol{q}\in\Gamma_m:V(\boldsymbol{p},\boldsymbol{q})\leq\epsilon}\big[J(\boldsymbol{q})-J(\boldsymbol{p})\big]$$

is given by taking $\boldsymbol{p}\in\Gamma_\ell$ to be the uniform distribution $\boldsymbol{p}'=(1/\ell,1/\ell,\ldots,1/\ell)$.

For any $\epsilon\in(0,2)$ and $\boldsymbol{p}=(p_1,p_2,\ldots,p_\ell)\in\Gamma_\ell$, assume that the p_i are sorted in descending order, and let

$$\boldsymbol{q}=\boldsymbol{r}(\boldsymbol{p},m,\epsilon)=(p_1+d_1,p_2+d_2,\ldots,p_\ell+d_\ell,d_{\ell+1},\ldots,d_m)\in\Gamma_m$$

be specified as in Theorem 2.1. For the above $\boldsymbol{p}'$, denote its corresponding $\boldsymbol{r}(\boldsymbol{p}',m,\epsilon)$ by

$$\boldsymbol{q}'=\boldsymbol{r}(\boldsymbol{p}',m,\epsilon)=\left(\frac{1}{\ell}+d_1',\frac{1}{\ell}+d_2',\ldots,\frac{1}{\ell}+d_\ell',d_{\ell+1}',\ldots,d_m'\right).$$

We now show that, for any $\boldsymbol{p}\in\Gamma_\ell$,

$$J(\boldsymbol{q})-J(\boldsymbol{p})\geq J(\boldsymbol{q}')-J(\boldsymbol{p}'). \tag{A.7}$$

Three cases arise.

Case 1. Suppose that $V(\boldsymbol{p},\boldsymbol{q})<\epsilon$. It was shown by Ho and Yeung [3] that

$$\boldsymbol{q}=\left(\frac{1}{m},\frac{1}{m},\ldots,\frac{1}{m}\right)\in\Gamma_m \quad\text{and}\quad \epsilon>2\left(1-\frac{\ell}{m}\right).$$

https://doi.org/10.1017/S0269964818000098 Published online by Cambridge University Press

Clearly,

$$V(\boldsymbol{p}', \boldsymbol{q}) = 2\left(1 - \frac{\ell}{m}\right) < \epsilon.$$

Thus, by Theorem 2.1, we have

$$J(\boldsymbol{q}') \geq J(\boldsymbol{q}) = (1-m)\log\left(1 - \frac{1}{m}\right).$$

On the other hand, $\boldsymbol{q} \preceq_{\mathrm{m}} \boldsymbol{q}'$ implies that $J(\boldsymbol{q}') \leq J(\boldsymbol{q})$. Then

$$J(\boldsymbol{q}') = J(\boldsymbol{q}) = (1-m)\log\left(1 - \frac{1}{m}\right).$$

Also, $\boldsymbol{p}' \preceq_{\mathrm{m}} \boldsymbol{p}$ implies that $J(\boldsymbol{p}) \leq J(\boldsymbol{p}') = (1-\ell)\log(1-1/\ell)$. Therefore,

$$J(\boldsymbol{q}) - J(\boldsymbol{p}) \geq J(\boldsymbol{q}') - J(\boldsymbol{p}') = h_2(\ell, m, \epsilon),$$

implying (A.7).

Case 2. Suppose $V(\boldsymbol{p}, \boldsymbol{q}) = V(\boldsymbol{p}', \boldsymbol{q}') = \epsilon$. It is trivial for $m = \ell$. So assume that $m > \ell$. By the construction of $\boldsymbol{q}'$ and $\boldsymbol{q}$, we have

$$d_1' = d_2' = \cdots = d_\ell' = -\frac{\epsilon}{2\ell}, \qquad d_{\ell+1}' = d_{\ell+2}' = \cdots = d_m' = \frac{\epsilon}{2(m-\ell)},$$

and

$$\sum_{i:d_i<0} d_i = -\frac{\epsilon}{2}, \quad \sum_{i:d_i>0} d_i = \frac{\epsilon}{2}.$$

Furthermore, $d_j > 0$ for $j = \ell+1, \ldots, m$, and $\tau \leq \ell$, where $\tau = \max\{i : d_i < 0, 1 \leq i \leq m\}$. Let $\mu_{\boldsymbol{p}}$ and $\mu_{\boldsymbol{p}'}$ be the μ's when $\boldsymbol{r}(\boldsymbol{p}, m, \epsilon)$ and $\boldsymbol{r}(\boldsymbol{p}', m, \epsilon)$ are evaluated as in Theorem 2.1, respectively. Then $\mu_{\boldsymbol{p}} \geq \mu_{\boldsymbol{p}'}$. From the construction as in (2.3), we know that

$$p_1 + d_1 = \cdots = p_\tau + d_\tau = \mu_{\boldsymbol{p}}, \quad \frac{1}{\ell} + d_1' = \cdots = \frac{1}{\ell} + d_\ell' = \mu_{\boldsymbol{p}'},$$

and $p_1 \geq \cdots \geq p_\tau \geq 1/\ell$. Thus,

$$\Big(p_1 + d_1, \ldots, p_\tau + d_\tau, \underbrace{\frac{1}{\ell}, \ldots, \frac{1}{\ell}}_{\ell}\Big) \preceq_{\mathrm{m}} \Big(p_1, \ldots, p_\tau, \ \frac{1}{\ell} + d_1', \ldots, \frac{1}{\ell} + d_\ell'\Big),$$

which implies that

$$\sum_{i=1}^{\tau} \zeta(p_i + d_i) + \sum_{i=1}^{\ell} \zeta\left(\frac{1}{\ell}\right) \geq \sum_{i=1}^{\tau} \zeta(p_i) + \sum_{i=1}^{\ell} \zeta\left(\frac{1}{\ell} + d_i'\right). \tag{A.8}$$

Here, $\zeta(x) = -(1-x)\log(1-x)$ is concave in $x \in [0,1]$.

Define

$$\tilde{\boldsymbol{q}} = (p_1 + d_1, p_2 + d_2, \ldots, p_\tau + d_\tau, \ p_{\tau+1}, \ldots, p_\ell, \ d_{\ell+1}', \ldots, d_m').$$

It is easy to see that $\tilde{\boldsymbol{q}} \in \Gamma_m$ and $V(\tilde{\boldsymbol{q}}, \boldsymbol{p}) = \epsilon$. From Theorem 2.1, it follows that $J(\boldsymbol{q}) \geq J(\tilde{\boldsymbol{q}})$. Thus, to prove (A.7), it suffices to prove that

$$J(\tilde{\boldsymbol{q}}) - J(\boldsymbol{p}) \geq J(\boldsymbol{q}') - J(\boldsymbol{p}') = h_1(\ell, m, \epsilon) \tag{A.9}$$

or, equivalently,

$$\sum_{i=1}^{\tau} \zeta(p_i + d_i) - \sum_{i=1}^{\tau} \zeta(p_i) \geq \sum_{i=1}^{\ell} \zeta\left(\frac{1}{\ell} + d_i'\right) - \sum_{i=1}^{\ell} \zeta\left(\frac{1}{\ell}\right).$$

This is (A.8). Thus, (A.9) and, hence, (A.7) holds.

https://doi.org/10.1017/S0269964818000098 Published online by Cambridge University Press

Case 3. Suppose that $V(\boldsymbol{p},\boldsymbol{q}) = \epsilon > \delta = V(\boldsymbol{p}',\boldsymbol{q}')$. By theorem 2.1, there exists $\boldsymbol{q}^* = \boldsymbol{r}(\boldsymbol{p},m,\delta)$ such that

$$\begin{aligned} J(\boldsymbol{q}) - J(\boldsymbol{p}) &= \max_{\boldsymbol{q}''\in\Gamma_m : V(\boldsymbol{p},\boldsymbol{q}'')\le\epsilon} \big[J(\boldsymbol{q}'') - J(\boldsymbol{p})\big] \\ &\ge \max_{\boldsymbol{q}''\in\Gamma_m : V(\boldsymbol{p},\boldsymbol{q}'')\le\delta} \big[J(\boldsymbol{q}'') - J(\boldsymbol{p})\big] \\ &= J(\boldsymbol{q}^*) - J(\boldsymbol{p}) \\ &\ge J(\boldsymbol{q}') - J(\boldsymbol{p}'), \end{aligned}$$

where the last inequality follows from Case 2. This proves (A.7) in Case 3.

Combining the above three cases, we have proved that for any $\boldsymbol{p}\in\Gamma_\ell$,

$$\sup_{\boldsymbol{q}\in\Gamma_m : V(\boldsymbol{p},\boldsymbol{q})\le\epsilon} (J(\boldsymbol{q}) - J(\boldsymbol{p})) \ge J(\boldsymbol{q}') - J(\boldsymbol{p}').$$

Thus,

$$\inf_{\boldsymbol{p}\in\Gamma_\ell} \sup_{\boldsymbol{q}\in\Gamma_m : V(\boldsymbol{p},\boldsymbol{q})\le\epsilon} (J(\boldsymbol{q}) - J(\boldsymbol{p})) \ge J(\boldsymbol{q}') - J(\boldsymbol{p}').$$

Since $\boldsymbol{p}'\in\Gamma_\ell$ and $\boldsymbol{q}'\in\Gamma_m$, we have

$$\begin{aligned} &\min_{\boldsymbol{p}\in\Gamma_\ell} \max_{\boldsymbol{q}\in\Gamma_m : V(\boldsymbol{p},\boldsymbol{q})\le\epsilon} \big[J(\boldsymbol{q}) - J(\boldsymbol{p})\big] \\ &\quad = J(\boldsymbol{q}') - J(\boldsymbol{p}') = \begin{cases} h_1(\ell,m,\epsilon), & 0 < \epsilon \le \dfrac{2(m-\ell)}{m}, \\ h_2(\ell,m,\epsilon), & \epsilon > \dfrac{2(m-\ell)}{m}, \end{cases} \end{aligned}$$

where $h_1(\ell,m,\epsilon)$ and $h_2(\ell,m,\epsilon)$ are defined by (3.6) and (3.7), respectively. This completes the proof of Theorem 3.4. ■

A.4. Proof of Theorem 3.5

Proof: For any $\tilde{\boldsymbol{p}} = (\tilde{p}_1,\tilde{p}_2,\ldots,\tilde{p}_\ell)$ and $\tilde{\boldsymbol{q}} = (\tilde{q}_1,\tilde{q}_2,\ldots,\tilde{q}_m)$ such that $V(\tilde{\boldsymbol{p}},\tilde{\boldsymbol{q}}) \le \epsilon$, let $\tilde{p}_i = 0$ for $\ell < i \le m$ and denote $d_i^* = \tilde{q}_i - \tilde{p}_i$. Obviously, $-1 \le d_i^* \le 1$ for each i. Without any loss generality, assume that the d_i^* are in ascending order, and denote

$$\tau = \max\{i : d_i^* < 0, 1 \le i \le m\} \quad \text{and} \quad k = \min\{i : d_i^* > 0, 1 \le i \le m\}.$$

Then, for $1 \le i \le \tau$,

$$(1-\tilde{p}_i)\log(1-\tilde{p}_i) - (1-\tilde{p}_i-d_i^*)\log(1-\tilde{p}_i-d_i^*) \le d_i^*\log(-d_i^*), \tag{A.10}$$

and for $k \le i \le m$,

$$(1-\tilde{p}_i)\log(1-\tilde{p}_i) - (1-\tilde{p}_i-d_i^*)\log(1-\tilde{p}_i-d_i^*) \le -(1-d_i^*)\log(1-d_i^*). \tag{A.11}$$

Note that $(1-x)\log(1-x) \ge x\log x$ for $0 \le x \le 1/2$. From (A.11), it follows that, for $0 \le d_i^* \le 1/2$,

$$(1-\tilde{p}_i)\log(1-\tilde{p}_i) - (1-\tilde{p}_i-d_i^*)\log(1-\tilde{p}_i-d_i^*) \le -d_i^*\log d_i^*. \tag{A.12}$$

Define a vector $\tilde{\boldsymbol{d}} = (\tilde{d}_1,\tilde{d}_2,\ldots,\tilde{d}_m)$ with

$$\tilde{d}_i = \begin{cases} -d_i^*, & -1 \le d_i^* < 0, \\ d_i^*, & 0 \le d_i^* < 1/2 \\ 1-d_i^*, & 1/2 \le d_i^* \le 1. \end{cases}$$

https://doi.org/10.1017/S0269964818000098 Published online by Cambridge University Press

Clearly, $\tilde{\boldsymbol{d}} \in \tilde{\Gamma}_m$ with $\gamma := \sum_{i=1}^m \tilde{d}_i \le \sum_{i=1}^m |d_i^*| \le \min\{1, \epsilon\}$. From (A.10), (A.11) and (A.12), we conclude that for all $1 \le i \le m$,

$$(1-\tilde{p}_i)\log(1-\tilde{p}_i) - (1-\tilde{p}_i-d_i^*)\log(1-\tilde{p}_i-d_i^*) \le -\tilde{d}_i \log \tilde{d}_i.$$

Thus,

$$\begin{aligned}
J(\tilde{\boldsymbol{q}}) - J(\tilde{\boldsymbol{p}}) &= -\sum_{i=1}^m (1-\tilde{p}_i-d_i^*)\log(1-\tilde{p}_i-d_i^*) + \sum_{i=1}^{\ell}(1-\tilde{p}_i)\log(1-\tilde{p}_i) \\
&= \sum_{i=1}^{\ell}\Big[-(1-\tilde{p}_i-d_i^*)\log(1-\tilde{p}_i-d_i^*) + (1-\tilde{p}_i)\log(1-\tilde{p}_i)\Big] \\
&\quad + \sum_{i=\ell+1}^{m}(1-d_i^*)\log(1-d_i^*) \\
&\le -\sum_{i=1}^m \tilde{d}_i \log \tilde{d}_i \\
&= H(\tilde{\boldsymbol{d}}) \\
&\le -\gamma\log\gamma + \gamma \log m \stackrel{\text{def}}{=} g(\gamma).
\end{aligned}$$

where the last inequality follows from Lemma 9 in Ho and Yeung [3]. Since $g(\gamma)$ is increasing when $\gamma \le \min\{m/e, 1\}$, we have

$$J(\tilde{\boldsymbol{q}}) - J(\tilde{\boldsymbol{p}}) \le -\epsilon\log\epsilon + \epsilon\log m.$$

Similarly,

$$J(\tilde{\boldsymbol{p}}) - J(\tilde{\boldsymbol{q}}) \le -\epsilon\log\epsilon + \epsilon\log m.$$

Thus, the desired result follows. ■

A.5. Simple Proofs of Two Lemmas

In this subsection of the appendix, we give simple proofs of Lemmas 7 and 8 in Ho and Yeung [3]. Lemma A.1 plays an important role in the proof of Theorem 9 in Ho and Yeung [3].

LEMMA A.1 (Lemma 7 in Ho and Yeung [3]): *Let* $\{(p_i, d_i) : i = 1, \ldots, n\}$ *and* $\{(p_j^*, d_j^*) : j = 1, \ldots, n\}$ *be two sets of real vectors such that for all* i *and* j,

$$0 \le p_i + d_i \le p_i \le 1 \quad \text{and} \quad 0 \le p_j^* + d_j^* \le p_j^* \le 1,$$

where $d_i, d_j^* \le 0$ *for all* i *and* j. *If* $\sum_{i=1}^n d_i = \sum_{i=1}^n d_i^*$ *and*

$$\min_{1\le i\le n}(p_i + d_i) \ge \max_{1\le j\le n} p_j^*,$$

then

$$\sum_{i=1}^n f(p_i, d_i) \ge \sum_{j=1}^n f(p_j^*, d_j^*), \tag{A.13}$$

where $f(p, d) = p\log p - (p+d)\log(p+d)$.

https://doi.org/10.1017/S0269964818000098 Published online by Cambridge University Press

PROOF: The proof given in Ho and Yeung [3] is not straightforward, and their construction method is not clear. Here we give a simple proof. It is trivial for $n = 1$. First consider $n = 2, 3$. It is easy to check that

$$(p_1^*, p_2^*, p_1 + d_1, p_2 + d_2) \preceq_{\mathrm{m}} (p_1, p_2, p_1^* + d_1^*, p_2^* + d_2^*) \tag{A.14}$$

and

$$\big(p_1^*, p_2^*, p_3^*, p_1 + d_1, p_2 + d_2, p_3 + d_3\big) \preceq_{\mathrm{m}} \big(p_1, p_2, p_3, p_1^* + d_1^*, p_2^* + d_2^*, p_3^* + d_3^*\big). \tag{A.15}$$

Note that $x \log x$ is convex on $\mathbb{R}_+$. By Proposition 3.C.1 in Marshall et al. [6], it follows that

$$\sum_{i=1}^{k} p_i^* \log p_i^* + \sum_{i=1}^{k} (p_i + d_i) \log(p_i + d_i) \le \sum_{i=1}^{k} p_i \log p_i + \sum_{i=1}^{k} (p_i^* + d_i^*) \log(p_i^* + d_i^*), \quad k = 2, 3,$$

which implies that (A.13) holds for $n = 2, 3$.

It is well-known that if $\boldsymbol{a}_1 \preceq_{\mathrm{m}} \boldsymbol{b}_1$ and $\boldsymbol{a}_2 \preceq_{\mathrm{m}} \boldsymbol{b}_2$, then $\boldsymbol{a}_1 + \boldsymbol{a}_2 \preceq_{\mathrm{m}} \boldsymbol{b}_1 + \boldsymbol{b}_2$, where $\boldsymbol{a}_i \in \mathbb{R}^r$ and $\boldsymbol{b}_i \in \mathbb{R}^s$ for $i = 1, 2$ and all positive integers r and s. From (A.14) and (A.15), it follows that, for $n \ge 2$,

$$(p_1^*, \ldots, p_n^*, p_1 + d_1, \ldots, p_n + d_n) \preceq_{\mathrm{m}} (p_1, \ldots, p_3, p_1^* + d_1^*, \ldots, p_n^* + d_n^*). \tag{A.16}$$

Therefore,

$$\sum_{i=1}^{n} p_i^* \log p_i^* + \sum_{i=1}^{n} (p_i + d_i) \log(p_i + d_i) \le \sum_{i=1}^{n} p_i \log p_i + \sum_{i=1}^{n} (p_i^* + d_i^*) \log(p_i^* + d_i^*),$$

implying (A.13). This completes the proof of the lemma. ■

LEMMA A.2 (Lemma 8 in Ho and Yeung [3]): *Let* $\{(p_i, d_i) : i = 1, \ldots, n\}$ *and* $\{(p_j^*, d_j^*) : j = 1, \ldots, n\}$ *be two sets of real vectors such that for all* i *and* j,

$$0 \le p_i \le p_i + d_i \le 1 \quad \textit{and} \quad 0 \le p_j^* \le p_j^* + d_j^* \le 1,$$

where $d_i, d_j^* \ge 0$ *for all* i *and* j. *If* $\sum_{i=1}^n d_i = \sum_{i=1}^n d_i^*$ *and*

$$\min_{1 \le i \le n} p_i^* \ge \max_{1 \le j \le n} (p_j + d_j),$$

then (A.13) *holds, where* $f(p, d) = p \log p - (p + d) \log(p + d)$.

PROOF: The proof is similar to that of Lemma A.1. Under the above conditions, it can be checked that (A.14), (A.15) and hence (A.16) hold. Therefore, (A.13) follows. ■

https://doi.org/10.1017/S0269964818000098 Published online by Cambridge University Press