

Correspondence

Convergence of Differential Entropies

Mahesh Godavarti, *Member, IEEE*, and Alfred Hero, *Fellow, IEEE*

Abstract—Calculation of the differential entropy of the limiting density of a sequence of probability density functions (pdfs) is an interesting mathematical problem and is important in asymptotic analysis of communication systems. In such cases, it would be of interest to know if the limit of the differential entropies \mathcal{H}_n , corresponding to the sequence of pdfs f_n , is equal to the differential entropy \mathcal{H} , of the limiting pdf f . In this correspondence, we establish sufficient conditions under which $\mathcal{H}_n \rightarrow \mathcal{H}$.

Index Terms—Convergence, differential entropy, probability density function (pdf).

I. INTRODUCTION

The concept of convergence of differential entropy can be traced to the problem of asymptotic analysis of communication systems where either the maximum asymptotic rate of communication [4], [9], [15], [16], [18], [19], [20], the asymptotic storage capacity [14], or the asymptotically optimal source compression [7], [8] is considered. The asymptotic analysis of communication systems is valid if the entropy of the limiting density function f is equal to the limit of the entropies of the converging density functions f_n . In this correspondence, we establish sufficient conditions for which $\mathcal{H}(f_n) \rightarrow \mathcal{H}(f)$.

A related area where the convergence of differential entropies is studied is that of entropy estimation [1]–[3], [10]–[13], [17]. In [2], [12], the authors use either the Gram–Charlier expansion or expansions based on moments to approximate the density and hence the entropy. In [1], [10], [13], [17] the density is estimated from a finite number of realizations X_1, \dots, X_n of the source X and the estimate is refined as $n \rightarrow \infty$. An estimate of the entropy \mathcal{H}_n is obtained from this estimate of the density f_n and is required to converge to $\mathcal{H}(f)$ as $f_n \rightarrow f$. The problem of entropy estimation finds application in independent component analysis and projection pursuit. For more applications, an overview of existing entropy estimators and the conditions on the final density function f required for convergence see [3].

However, the problem we are looking at in this correspondence is not of entropy estimation but that of entropy convergence where a sequence of random variables are converging to a final random variable and we are interested in the convergence of the corresponding differential entropies. Unlike in the problem described in the previous paragraph, the entropy \mathcal{H}_n is not a random variable but a deterministic quantity. Hence, the convergence we are interested in is deterministic rather than stochastic.

Manuscript received November 29, 2001; revised July 20, 2003. This work was supported by the IEEE. This work was completed while Mahesh Godavarti was a Ph.D. candidate at the University of Michigan under the supervision of Prof. Alfred Hero. The material in this correspondence was presented in part at the IEEE International Symposium on Information theory, Laussane, Switzerland, June/July 2002.

M. Godavarti is with Ditech Communications, Inc. Mountain View, CA USA (e-mail: mgodavar@hotmail.com) (**Au: Please give zip code for Mountain View, CA. Thank you.**)

A. Hero is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: hero@eecs.umich.edu).

Communicated by P. Narayan, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2003.821979

Some conditions known for the convergence of differential entropies are given below [5].

Proposition 1: Let $f_n(x) \rightarrow f(x)$ pointwise and let $f_n(x)$ be bounded from above and below for all n over the support of $f_n(x)$ then $\mathcal{H}(f_n) \rightarrow \mathcal{H}(f)$.

Proof: We have $|\log f_n(x)| \leq A$ for x in the support of $f_n(x)Sf_n$, for all n . Let Sf denote the support of $f(x)$ and let $Sf \setminus Sf_n$ denote the set of all x such that $x \in Sf$ and $x \notin Sf_n$. Then

$$\begin{aligned} & \left| \lim_{n \rightarrow \infty} \int_{Sf_n} f_n(x) \log f_n(x) dx - \int_{Sf} f(x) \log f(x) dx \right| \\ & \leq \left| \lim_{n \rightarrow \infty} \int_{Sf_n} f(x) \log \frac{f(x)}{f_n(x)} dx \right| \\ & \quad + \left| \lim_{n \rightarrow \infty} \int_{Sf \setminus Sf_n} f(x) \log f(x) dx \right| \\ & \quad + \left| \lim_{n \rightarrow \infty} \int_{Sf_n} (f(x) - f_n(x)) \cdot \log f_n(x) dx \right| \\ & \leq \left| \lim_{n \rightarrow \infty} \int_{Sf_n} f(x) \log \frac{f(x)}{f_n(x)} dx \right| \\ & \quad + \left| \lim_{n \rightarrow \infty} \int_{Sf \setminus Sf_n} f(x) \log f(x) dx \right| \\ & \quad + \lim_{n \rightarrow \infty} \int_{Sf_n} |f(x) - f_n(x)| A dx \\ & \rightarrow 0. \quad \square \end{aligned}$$

The sufficient conditions given in Proposition 1 exclude the case of converging Gaussian densities given below.

Example 1: Let

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

and

$$f_n(x) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{x^2}{2\sigma_n^2}\right).$$

If $\sigma_n \rightarrow \sigma$ then we can easily see that $\mathcal{H}(f_n) \rightarrow \mathcal{H}(f)$

Gaussian densities occur in asymptotic analysis of Rayleigh/Rician fading channels and hence it is useful to have conditions that facilitate analysis of such channels. In that regard, we show convergence under weaker conditions on the converging and final density functions. We start with the following examples.

Example 2: Consider the sequence of probability density functions (pdfs) $f_n(x)$ defined over the real line as follows:

$$f_n(x) = \begin{cases} 1 - \frac{1}{n}, & \text{when } x \in [0, 1] \\ \frac{1}{nL^n}, & \text{when } x \in (1, 1 + L^n] \\ 0, & \text{elsewhere} \end{cases}$$

where L is a positive number not equal to 1. Then $f_n(x)$ converges to $f(x)$ pointwise where $f(x)$ is the uniform distribution over the in-

terval $[0, 1]$. However, the differential entropy from $f_n(x)$, called \mathcal{H}_n , is given by

$$\begin{aligned}\mathcal{H}_n &= -\left(1 - \frac{1}{n}\right) \log\left(1 - \frac{1}{n}\right) - \frac{1}{nL^n} \left[\log \frac{1}{nL^n}\right] L^n \\ &= -\left(1 - \frac{1}{n}\right) \log\left(1 - \frac{1}{n}\right) + \frac{1}{n} \log n + \log L\end{aligned}$$

and, therefore, $\lim_{n \rightarrow \infty} \mathcal{H}_n = \log L \neq 0 = \mathcal{H}_f$.

Example 3: Consider the sequence of pdfs $f_n(x)$ defined over the real line as follows:

$$f_n(x) = \begin{cases} 1 - \frac{1}{n}, & \text{when } x \in [0, 1] \\ \frac{L^n}{n}, & \text{when } x \in (1, 1 + \frac{1}{L^n}] \\ 0, & \text{elsewhere} \end{cases}$$

where L is a positive number not equal to 1. Then $f_n(x)$ converges to $f(x)$ almost everywhere where $f(x)$ is the uniform distribution over the interval $[0, 1]$. The differential entropy from $f_n(x)$, \mathcal{H}_n is given by

$$\begin{aligned}\mathcal{H}_n &= -\left(1 - \frac{1}{n}\right) \log\left(1 - \frac{1}{n}\right) - \frac{L^n}{n} \log \frac{L^n}{n} \\ &= -\left(1 - \frac{1}{n}\right) \log\left(1 - \frac{1}{n}\right) + \frac{1}{n} \log n - \log L\end{aligned}$$

and, therefore, $\lim_{n \rightarrow \infty} \mathcal{H}_n = -\log L \neq 0 = \mathcal{H}_f$.

In both the examples given above we see that convergence of pdfs does not lead to the convergence of the corresponding differential entropies. In Example 2, we see that the second moment $\int |x|^2 f_n(x) dx$ is unbounded whereas in Example 3, we see that the pdf $f_n(x)$ itself is unbounded. It is possible to ask the question if we ensure that the above two quantities are bounded then do we obtain convergence of the differential entropies? The answer is indeed yes and is proved in Section II. For the two examples, we also note that $\int f_n(x) |\log f_n(x)|^\kappa dx$ for $\kappa > 1$ is unbounded. We will see that limiting this quantity also ensures the convergence of differential entropies.

II. MAIN RESULTS

Let $\chi_P(x)$ denote the characteristic function over a set P defined as $\chi_P(x) = 0$ if $x \notin P$ and $\chi_P(x) = 1$ if $x \in P$.

Lemma 1: Let $g : \mathbb{R}^P \rightarrow \mathbb{R}$ be a positive bounded function whose region of support S_g is compact. If there exists a constant L such that $\int g(x) dx \leq L < 1/e$ then

$$\left| \int g(x) \log g(x) dx \right| \leq \max\{|L \log L| + |L \log \text{vol}(S_g)| G_m\}, |L \log\}$$

where $G_m = \sup g(x)$.

Proof: First

$$\int g(x) \log g(x) dx \leq \int g(x) \log G_m dx \leq L \log G_m.$$

Let $\int g(x) dx = I_g$. Consider the pdf $g(x)/I_g$. We know that

$$\int \frac{g(x)}{I_g} \log \frac{g(x)}{I_g f(x)} dx \geq 0$$

for all pdfs $f(x)$. If

$$f(x) = \frac{\chi_{S_g}}{\text{vol}(S_g)}$$

then

$$\begin{aligned}\int g(x) \log g(x) dx &\geq \int g(x) \log(I_g f(x)) \\ &= I_g \log \frac{I_g}{\text{vol}(S_g)}.\end{aligned}$$

This implies

$$\begin{aligned}\left| \int g(x) \log g(x) \right| &\leq \max\left\{|L \log G_m|, \left|I_g \log \frac{I_g}{\text{vol}(S_g)}\right|\right\} \\ &\leq \max\left\{|L \log G_m|, |I_g \log I_g| \right. \\ &\quad \left. + \left|I_g \log \text{vol}(S_g)\right|\right\} \\ &\leq \max\left\{|L \log G_m|, |L \log L| \right. \\ &\quad \left. + |L \log \text{vol}(S_g)|\right\}.\end{aligned}$$

The last inequality follows from the fact that for $x < 1/e$, $|x \log x|$ is an increasing function of x . \square

Theorem 1: Let $\{X_i \in \mathbb{R}^P\}$ be a sequence of continuous random variables with pdfs $\{f_i\}$ and $X \in \mathbb{R}^P$ be a continuous random variable with pdf f such that $f_i \rightarrow f$ pointwise. Let $\|x\| = \sqrt{x^\top x}$ denote the Euclidean norm of $x \in \mathbb{R}^P$. If 1) $\max\{\sup_x f_i(x), \sup_x f(x)\} \leq F_m < \infty$ for all i and 2) $\max\{\int \|x\|^\kappa f_i(x) dx, \int \|x\|^\kappa f(x) dx\} \leq L < \infty$ for some $\kappa > 1$ and all i then $\mathcal{H}(X_i) \rightarrow \mathcal{H}(X)$.

Proof: First, we need to show that $\mathcal{H}(X)$ exists and is finite. Since $f(x) \leq F_m$ we have

$$\mathcal{H}(X) = -\int f(x) \log f(x) dx \geq -\log F_m.$$

We will see in the course of the proof that $\mathcal{H}(X)$ is bounded from above as well. The proof is based on showing that given an $\epsilon > 0$ there exists a set A_ϵ with finite volume such that for all i , $|\int_{A_\epsilon^c} f_i(x) \log f_i(x) dx| < \epsilon$ where S^c denotes the complement of S . This A_ϵ also works for $f(x)$. We claim that $\{x : \|x\| \leq R\}$ for sufficiently large R is the A_ϵ we are looking for.

Since $y \log y \rightarrow 0$ as $y \rightarrow 0$ we have

$$\max_{f(x) \leq F_m} |f(x) \log f(x)| \leq \max\{F_m \log F_m, \epsilon\} \stackrel{\text{def}}{=} K.$$

Therefore, $f_i(x) \log f_i(x) \chi_{\|x\| \leq R}$ is bounded above by a nonnegative L^1 function ($g = K \chi_{\|x\| \leq R}$) and by the dominated convergence theorem we have

$$-\int_{\|x\| \leq R} f_i(x) \log f_i(x) dx \rightarrow -\int_{\|x\| \leq R} f(x) \log f(x) dx.$$

Now, we show that the integral on A_ϵ^c is uniformly bounded by a sufficiently small quantity for all f_i and f . Let g denote either f_i or f . We have $\int \|x\|^\kappa g(x) dx \leq L$. Therefore, by Markov's inequality

$$\int_{R < \|x\| \leq R+1} g(x) dx = I^R \leq L/R^\kappa.$$

Choose R large enough so that for all $l > R$: $I^l < 1/e$. Now

$$\begin{aligned}\left| \int_{\|x\| > R} g(x) \log g(x) dx \right| &\leq \int_{\|x\| > R} |g(x) \log g(x)| dx \\ &= \sum_{l=R}^{\infty} \int_{B_l} |g(x) \log g(x)| dx\end{aligned}$$

where $B_l = \{x : l < \|x\| \leq l+1\}$.

Consider the term $\int_{B_l} |g(x) \log g(x)| dx = G_l$. Also, define $A_+ = \{x : -\log g(x) > 0\}$ and $A_- = \{x : -\log g(x) < 0\}$ Now

$$\begin{aligned} G_l &= \int_{A_+ \cap B_l} |g(x) \log g(x)| dx + \int_{A_- \cap B_l} |g(x) \log g(x)| dx \\ &= \left| \int_{A_+ \cap B_l} g(x) \log g(x) dx \right| + \left| \int_{A_- \cap B_l} g(x) \log g(x) dx \right|. \end{aligned}$$

From Lemma 1, we have

$$G_l \leq 2 \max\{|I^l \log I^l| + |I^l \log \text{vol}(\{B_l\})|, |I^l \log F_m|\}.$$

We know $\text{vol}(\{x : B_l\}) = o(l^{2P})$. Therefore,

$$\int_{B_l} |g(x) \log g(x)| dx \leq \frac{Q}{l^\kappa} \log l$$

where Q is some sufficiently large constant. Therefore, we have

$$\int_{\|x\| > R} |g(x) \log g(x)| dx \leq \sum_{l=R}^{\infty} \frac{Q}{l^\kappa} \log l = O\left(\frac{\log R}{R^{\kappa-1}}\right).$$

As $\kappa > 1$ we can choose R sufficiently large to have

$$\left| \int_{\|x\| > R} g(x) \log g(x) dx \right| < \epsilon.$$

Finally, we show that $\mathcal{H}(X)$ is bounded from above

$$\begin{aligned} \mathcal{H}(X) &\leq \left| \int_{\|x\| \leq R} f(x) \log f(x) dx \right| + \left| \int_{\|x\| > R} f(x) \log f(x) dx \right| \\ &\leq K \text{vol}(\|x\| \leq R) + \epsilon \end{aligned}$$

where K is as defined earlier. \square

Theorem 2: Let $\{X_i \in \mathbb{C}^P\}$ be a sequence of continuous random variables with pdfs $\{f_i\}$ and $X \in \mathbb{C}^P$ be a continuous random variable with pdf f such that $f_i \rightarrow f$ pointwise. If 1) $f(x)$ is bounded, 2) $\int \|x\|^\kappa f_n(x) dx \leq L$ and $\int \|x\|^\kappa f(x) dx \leq L$ for some $\kappa > 1$ and $L < \infty$ then

$$\limsup_{i \rightarrow \infty} \mathcal{H}(X_i) \leq \mathcal{H}(X).$$

Proof: First, we note from the proof of Theorem 1 that $\mathcal{H}(X)$ is finite.

We prove this theorem by showing that for every ϵ there exists a set A_ϵ and a positive integer K such that

$$\left| \int_{A_\epsilon} f_i(x) \log f_i(x) dx - \int_{A_\epsilon} f(x) \log f(x) dx \right| < \epsilon$$

and

$$- \int_{A_\epsilon^c} f_i(x) \log f_i(x) dx \leq - \int_{A_\epsilon^c} f(x) \log f(x) dx$$

for all $i > K$. Let $F_m = \sup f(x)$ and for every i consider the set $S_i = \{x : f_i(x) > N\}$ for a sufficiently large $N > F_m$. Note that $\bigcap_{k=1}^{\infty} \bigcup_{i=k}^{\infty} S_i = \emptyset$ since $f_i \rightarrow f$ pointwise and f is bounded.

Let $S_M = \bigcup_{i=M}^{\infty} S_i$. Note that for all $i > M$, $S_i \subset S_M$. Also, note that S_i is a subset of $S_f = \{x : f(x) > \delta\}$ for sufficiently small δ and for all sufficiently large i . Since S_f has a finite volume $\lim_{i \rightarrow \infty} \text{vol}(S_i) = 0$.

Now, choose M large enough so that

$$\text{vol}(S_M) < \epsilon / (\max\{N \log N\}).$$

We claim that S_M^c for sufficiently large M is the A_ϵ we are looking for. On $S_M^c = A_\epsilon$, $f_i(x)$ and $f(x)$ are bounded and by Theorem 1 we have

$$\int_{S_M^c} f_i(x) \log f_i(x) dx \rightarrow \int_{S_M^c} f(x) \log f(x) dx.$$

On $S_M = A_\epsilon^c$, we have

$$\begin{aligned} - \int_{S_M} f_i(x) \log f_i(x) dx &= - \int_{S_M \cap \{x: f_i(x) \leq N\}} f_i(x) \log f_i(x) dx \\ &\quad - \int_{S_M \cap \{x: f_i(x) > N\}} f_i(x) \log f_i(x) dx. \end{aligned}$$

Since

$$\left| - \int_{S_M \cap \{x: f_i(x) \leq N\}} f_i(x) \log f_i(x) dx \right| < \epsilon$$

and

$$- \int_{S_M \cap \{x: f_i(x) > N\}} f_i(x) \log f_i(x) dx < 0$$

we have for all sufficiently large i

$$\limsup_{M \rightarrow \infty} \int_{S_M} f_i(x) \log f_i(x) dx < 0.$$

On the other hand, we have $\lim_{M \rightarrow \infty} \int_{S_M} f_i(x) \log f_i(x) = 0$ for all sufficiently large i . Therefore, our claim has been established and the proof is complete. \square

Theorem 3: Let $\{X_i \in \mathbb{C}^P\}$ be a sequence of continuous random variables with pdfs $\{f_i\}$ and $X \in \mathbb{C}^P$ be a continuous random variable with pdf f such that $f_i \rightarrow f$ pointwise. If 1) $\max\{\sup_x f(x), \sup_x f_i(x)\} \leq F_m < \infty$ and 2) $\int \|x\|^\kappa f(x) dx$ is bounded for some $\kappa > 1$, then

$$\liminf_{i \rightarrow \infty} \mathcal{H}(X_i) \geq \mathcal{H}(X).$$

Proof: First, for every $\epsilon > 0$ there exists $R > 0$ such that

$$- \int_{|x| < R} g(x) \log g(x) dx \geq \mathcal{H}(X) - \epsilon$$

where $g(x)$ is defined as

$$g(x) = f(x) \chi_{|x| < R}(x) + F_m \chi_{R \leq |x| < R + \Delta R}(x)$$

where ΔR is such that

$$\int_{|x| \geq R} f(x) dx = F_m \text{vol}(\{x : R \leq |x| < R + \Delta R\}).$$

Similarly, define $g_i(x)$ as

$$g_i(x) = f_i(x) \chi_{|x| < R}(x) + F_m \chi_{R \leq |x| < R + \Delta_i R}(x)$$

where $\Delta_i R$ is such that

$$\int_{|x| \geq R} f_i(x) dx = F_m \text{vol}(\{x : R \leq |x| < R + \Delta_i R\}).$$

Then from Theorem 1 we have

$$\liminf_{i \rightarrow \infty} - \int g_i(x) \log g_i(x) dx = - \int g(x) \log g(x) dx.$$

$$\text{Since } - \int_{|x| \geq R} f_i(x) \log f_i(x) dx \geq - \int_{|x| \geq R} g_i(x) \log g_i(x) dx$$

$$\begin{aligned} \liminf_{i \rightarrow \infty} - \int f_i(x) \log f_i(x) dx &\geq \liminf_{i \rightarrow \infty} - \int g_i(x) \log g_i(x) dx \\ &= - \int g(x) \log g(x) dx \\ &\geq \mathcal{H}(X) - \epsilon. \end{aligned}$$

Since ϵ is arbitrary we are done if we can show that $\mathcal{H}(X)$ is finite. But that follows from the proof of Theorem 1. \square

Now, we will relax the boundedness condition on the densities $f_n(x)$ and f . Note that in Examples 2 and 3, $\int f_n(x)(\log f_n(x))^2 dx \rightarrow \infty$ as $n \rightarrow \infty$. We will show that limiting that quantity alone guarantees convergence.

Theorem 4: Let $\{X_i \in \mathbb{C}^P\}$ be a sequence of continuous random variables with pdfs $\{f_i\}$ and $X \in \mathbb{C}^P$ be a continuous random variable with pdf f such that $f_i \rightarrow f$ pointwise. If

$$\max\left\{\int f_i(x)|\log f_i(x)|^\kappa dx, \int f(x)|\log f(x)|^\kappa dx\right\} \leq L < \infty$$

for some $\kappa > 1$ and all i and then $\mathcal{H}(X_i) \rightarrow \mathcal{H}(X)$.

Proof: The boundedness of $\mathcal{H}(X)$ follow from the fact that

$$\begin{aligned} -\int f(x)\log f(x)dx &= -\int_{|\log f(x)| < \log N} f(x)\log f(x)dx \\ &\quad -\int_{|\log f(x)| \geq \log N} f(x)\log f(x)dx \end{aligned}$$

where N is a sufficiently large number. The first term is bounded since f is bounded from above and below and the second term is bounded since

$$\left|\int_{|\log f(x)| \geq \log N} f(x)\log f(x)\right| \leq \int_{|\log f(x)| \geq \log N} f(x)|\log f(x)|^\kappa \leq L.$$

The essence of the proof is in showing that for every ϵ there exists a set A_ϵ such that $\int f_i(x)\log f_i(x)dx$ and $\int f(x)\log f(x)dx$ evaluated on A_ϵ^c are bounded by ϵ and

$$\left|\int_{A_\epsilon} f_i(x)\log f_i(x)dx - \int_{A_\epsilon} f(x)\log f(x)dx\right| < \epsilon$$

for all sufficiently large i .

For sufficiently large N , define

$$S_i = \{x : 1/N \leq f_i(x) \leq N\} \text{ and } S = \{x : 1/N \leq f(x) \leq N\}.$$

Then $\bigcap_{k=1}^\infty \bigcup_{i=k}^\infty S_i = S$ as $f_i \rightarrow f$ pointwise. Let $S_M = \bigcup_{i=M}^\infty S_i \cup S$ then on S_M^c , $|\log f_i(x)| > \log N$ for all $i > M$ and also $|\log f(x)| > \log N$. Therefore, on S_M^c

$$\begin{aligned} \left|\int_{S_M^c} f_i(x)\log f_i(x)dx\right| &\leq \int_{S_M^c} f_i(x)|\log f_i(x)|dx \\ &\leq (\log N)^{\kappa-1} \int_{S_M^c} f_i(x)|\log f_i(x)|^\kappa dx \\ &\leq \int_{S_M^c} f_i(x)|\log f_i(x)|^\kappa dx \leq L. \end{aligned}$$

By similar analysis for f , we see that the integrals evaluated on S_M^c are uniformly bounded by a negligible quantity for all large i .

We will show that S_M is the A_ϵ we are looking for by proving that

$$\lim_{n \rightarrow \infty} \int_{S_M} f_i(x)\log f_i(x)dx = \int_{S_M} f(x)\log f(x)dx.$$

For that, given a sufficient large positive integer K , we divide S_M into two regions \mathcal{D}_1 and \mathcal{D}_2 . \mathcal{D}_1 is the set of all points in S_M such $f(x) > K$ and \mathcal{D}_2 consists of $x \in S_M$ such that $f(x) \leq K$. Since S_M has finite volume, by Egoroff's theorem, [6, Theorem 2.33, p. 60] there exist regions $\mathcal{D}_{1,\epsilon}$ and $\mathcal{D}_{2,\epsilon}$ such that $\max\{\text{vol}(\mathcal{D}_{1,\epsilon}), \text{vol}(\mathcal{D}_{2,\epsilon})\} \leq \epsilon$ and f_i converge to f uniformly on $\mathcal{B}_1 = \mathcal{D}_1 \setminus \mathcal{D}_{1,\epsilon}$ and $\mathcal{B}_2 = \mathcal{D}_2 \setminus \mathcal{D}_{2,\epsilon}$. Let $\mathcal{D}_\epsilon = \mathcal{D}_{1,\epsilon} \cup \mathcal{D}_{2,\epsilon}$.

On \mathcal{B}_2 , all $f_i(x)$ are bounded from above by $K + \delta$ for sufficient large i , therefore, by dominated convergence theorem

$$\lim_{i \rightarrow \infty} \int_{\mathcal{B}_2} f_i(x)\log f_i(x)dx = \int_{\mathcal{B}_2} f(x)\log f(x)dx.$$

On \mathcal{B}_1 , all $f_i(x)$ are bounded from below by a sufficiently large $K' = K + \delta$. Then

$$\begin{aligned} \left|\int_{\mathcal{B}_1} f_i(x)\log f_i(x)dx\right| &< |\log K'|^{\kappa-1} \left|\int_{\mathcal{B}_1} f_i(x)\log f_i(x)dx\right| \\ &\leq \int_{\mathcal{B}_1} f_i(x)|\log f_i(x)|^\kappa dx \\ &\leq \int_{\mathcal{B}_1} f_i(x)|\log f_i(x)|^\kappa dx < L. \end{aligned}$$

Therefore,

$$\left|\int_{\mathcal{B}_1} f_i(x)\log f_i(x)dx\right| < L/|\log K'|^{\kappa-1}$$

and it can be made sufficiently small by choosing K large enough.

Now we will show that on \mathcal{D}_ϵ , $|\int f_i(x)\log f_i(x)dx|$ is negligibly small for all sufficiently large i . For each i , divide \mathcal{D}_ϵ into two regions C_1^i and C_2^i such that $f_i(x)$ is bounded from below by K on C_1^i and bounded from above by K on C_2^i . As before, on C_1^i , we can show

$$\left|\int_{C_1^i} f_n(x)\log f_i(x)dx\right| < L/|\log K|^{\kappa-1}.$$

Finally, on C_2^i we note that

$$|f_i(x)\log f_i(x)| \leq \max\{K \log K, e\}$$

and since, $\text{vol}(C_2^i)$ is small enough $|\int_{C_2^i} f_i(x)\log f_i(x)dx|$ is sufficiently small and we are done. \square

It is of interest to find out if the conditions in Theorem 4 are necessary as well. To disprove the necessity all we need to do is find an example such that

$$\int f_n(x)|\log f_n(x)|dx \rightarrow \int f(x)|\log f(x)|dx$$

but

$$\liminf_{n \rightarrow \infty} \int f_n(x)|\log f_n(x)|^\kappa dx = \infty$$

for all $\kappa > 1$.

Example 4: First, let

$$a_n = \sum_{m=N}^{\infty} \frac{1}{m^{1+1/n}} = \frac{K}{N^{1/n}}$$

and

$$a'_n = \sum_{m=N}^{\infty} \frac{1}{m^{2+1/n}} = \frac{K'}{N^{1+1/n}}$$

where $K > 0$ and $K' > 0$ are constants. Next, define $f_n(x)$ as follows:

$$f_n(x) = \begin{cases} 1 - \frac{K'}{nN^{1+1/n}}, & \text{when } x \in [0, 1] \\ \frac{\exp\{-m\}}{n}, & \text{when } x \in (b_m, b_{m+1}] \end{cases}$$

where N has been defined such that f_n is a proper density function and b_m is defined as follows:

$$b_m = \begin{cases} 0, & \text{when } m < N \\ 1, & \text{when } m = N \\ b_{m-1} + \exp\{m\}/m^{2+1/n}, & \text{when } m > N. \end{cases}$$

Then $f_n(x)$ converges to $f(x)$ pointwise where $f(x)$ is the uniform distribution over the interval $[0, 1]$. The differential entropy from $f_n(x)$, \mathcal{H}_n is given by

$$\begin{aligned} \mathcal{H}_n &= - \left(1 - \frac{K'}{nN^{1+1/n}} \right) \log \left(1 - \frac{K'}{nN^{1+1/n}} \right) \\ &\quad - \sum_{m=N}^{\infty} \left[\frac{\exp\{-m\}}{n} \log \left(\frac{\exp\{-m\}}{n} \right) \frac{\exp\{m\}}{m^{2+1/n}} \right] \\ &= - \left(1 - \frac{K'}{nN^{1+1/n}} \right) \log \left(1 - \frac{K'}{nN^{1+1/n}} \right) \\ &\quad + \frac{1}{n} \sum_{m=N}^{\infty} \frac{1}{m^{1+1/n}} + \frac{\log n}{n} \sum_{m=N}^{\infty} \frac{1}{m^{2+1/n}} \\ &= - \left(1 - \frac{K'}{nN^{1+1/n}} \right) \log \left(1 - \frac{K'}{nN^{1+1/n}} \right) \\ &\quad + \frac{a_n}{n} + \frac{a'_n \log n}{n} \end{aligned}$$

and, therefore, $\lim_{n \rightarrow \infty} \mathcal{H}_n = 0 = \mathcal{H}_f$. But for any $\kappa > 1$ if we choose $n > 1/(\kappa - 1)$ then

$$\begin{aligned} &\int f_n(x) |\log f_n(x)|^\kappa dx \\ &= \left(1 - \frac{K'}{nN^{1+1/n}} \right) \left(\log \left(1 - \frac{K'}{nN^{1+1/n}} \right) \right)^\kappa \\ &\quad + \sum_{m=N}^{\infty} \left[\frac{\exp\{-m\}}{n} (m + \log n)^\kappa \frac{\exp\{m\}}{m^{2+1/n}} \right] \\ &> \left(1 - \frac{K'}{nN^{1+1/n}} \right) \left(\log \left(1 - \frac{K'}{nN^{1+1/n}} \right) \right)^\kappa \\ &\quad + \frac{1}{n} \sum_{m=N}^{\infty} \frac{m^\kappa}{m^{2+1/n}} \\ &= \infty. \end{aligned}$$

III. DISCUSSION AND CONCLUSION

It is useful to make the following simple observations on the results derived in Section II. First, it can be easily seen that the pointwise convergence condition on the density functions in all theorems can be relaxed to almost everywhere convergence with minimal changes to the corresponding proofs. Second, we note that Theorems 1–3 can be applied to the case where density functions are defined on any normed space as long as $\log \mu(\{x : \|x\| \leq R\}) = O(\log R)$, where μ is the measure defined on the underlying space. Third, we note from the proof of Theorem 4 even the requirement of a norm on the underlying space can be relaxed for the applicability of Theorem 4. Fourth, we can generalize Theorem 4 in the same way that Theorem 1 was generalized to Theorems 2 and 3. In that regard, we have the following.

- If the condition in Theorem 4 is relaxed to

$$\max_i \left\{ \int_{f_i > N} f_i(x) |\log f_i(x)|^\kappa dx \right\} \leq L < \infty$$

for some $\kappa > 0$ and some $N > 0$; and $\int f(x) |\log f(x)|^\kappa dx \leq L$ then $\limsup \mathcal{H}(X_i) \leq \mathcal{H}(X)$.

- And if it is relaxed to

$$\max_i \left\{ \int_{f_i < N} f_i(x) |\log f_i(x)|^\kappa dx \right\} \leq L < \infty$$

for some $\kappa > 0$ and some $N > 0$ and; $\int f(x) |\log f(x)|^\kappa dx$ then $\liminf \mathcal{H}(X_i) \geq \mathcal{H}(X)$

We have derived general sufficient conditions for the convergence of differential entropies. The first set of conditions (Theorem 1) is as follows:

- 1) $\sup_x \max\{\sup_n f_n(x), f(x)\} < \infty$.
- 2) $\max\{\sup_n \int |x|^\kappa f_n(x) dx, \int |x|^\kappa f(x) dx\} < \infty$ for some $\kappa > 0$.

The second set of conditions ensuring convergence (Theorem 4) is as follows:

- $\max\{\sup_n \int f_n(x) |\log f_n(x)|^\kappa dx, \int f(x) |\log f(x)|^\kappa dx\} < \infty$ for some $\kappa > 1$.

For future work, it can be investigated whether a weaker form of the preceding conditions is necessary and sufficient for the convergence of entropies.

- There exists a decreasing sequence $\{\kappa_n\}$ such that $\kappa_n \rightarrow 1$ and

$$\max \left\{ \sup_n \int f_n(x) |\log f_n(x)|^{\kappa_n} dx, \int f(x) |\log f(x)|^{\kappa_n} dx \right\} < \infty.$$

Convergence of differential entropies is an interesting mathematical problem and our results give a preliminary solution. Our results also find application in the asymptotic analysis of communication systems. Examples include the capacity calculation of multiple-antenna systems for high signal-to-noise ratio [9], [20]. The results in this correspondence were directly applied in [9]. In [20], because of the special nature of the density functions considered, the convergence of the differential entropies could be proved directly. However, the same analysis could have been performed with the results derived in this correspondence.

REFERENCES

- [1] I. A. Ahmad and P.-E. Lin, "A nonparametric estimation of the entropy for absolutely continuous distributions," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 688–692, May 1976.
- [2] S. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," in *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, 1996, vol. 8, pp. 757–763.
- [3] J. Beirlant, E. J. Dudewicz, L. Györfi, and E. van der Meulen, "Nonparametric entropy estimation: An overview," *Int. J. Math. Stat. Sci.*, vol. 6, no. 1, pp. 17–39, June 1997.
- [4] C.-N. Chuah, D.N. C. Tse, J. M. Kahn, and R. A. Valenzuela, "Capacity scaling in mimo wireless systems under correlated fading," *IEEE Trans. Inform. Theory*, vol. 48, pp. 637–650, Mar. 2002.
- [5] I. Csiszár, private communication, ISIT, Washington, DC 2001.
- [6] G. Folland, *Real Analysis—Modern Techniques and Their Applications*. New York: Wiley-Interscience, 1984.
- [7] A. Gersho, "Asymptotically optimal block quantization," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 373–380, July 1979.
- [8] H. Gish and J. N. Pierce, "Asymptotically efficient quantizing," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 676–683, Sept. 1968.
- [9] M. Godavarti and A. O. Hero III, "Multiple antenna capacity in a deterministic Rician fading channel," *IEEE Trans. Inform. Theory*, submitted for publication.
- [10] P. Hall and S. C. Morton, "On the estimation of entropy," *Ann. Inst. Statist. Math.*, vol. 45, pp. 69–88, 1993.
- [11] A. O. Hero, B. Ma, O. Michel, and J. Gorman, "Applications of entropic spanning graphs," *IEEE Signal Processing Mag.*, vol. 19, pp. 85–95, Oct. 2002.
- [12] A. Hyvärinen, "New approximations of differential entropy for independent component analysis and projection pursuit," Helsinki Univ. Technol., Rep. A47, 1997.
- [13] H. Joe, "On the estimation of entropy and other functionals of a multivariate density," *Ann. Inst. Statist. Math.*, vol. 41, pp. 683–697, 1989.
- [14] A. Kuh and B. W. Dickinson, "Information capacity of associative memories," *IEEE Trans. Inform. Theory*, vol. 35, pp. 59–68, Jan. 1989.

- [15] A. Lozano and A. M. Tulino, "Capacity of multiple-transmit multiple-receive antenna architectures," *IEEE Trans. Inform. Theory*, vol. 48, pp. 3117–3128, Dec. 2002.
- [16] X. Mestre, J. R. Fonollosa, and A. Pages-Zamora, "Capacity of MIMO channels: Asymptotic evaluation under correlated fading," *IEEE J. Select. Areas Commun.*, vol. 21, pp. 829–838, June 2003.
- [17] A. Molkadem, "Estimation of the entropy and information of absolutely continuous random variables," *IEEE Trans. Inform. Theory*, vol. 35, pp. 193–196, Jan. 1989.
- [18] V. V. Prelov and E. C. van der Meulen, "An asymptotic expression for the information and capacity of a multidimensional channel with weak input signals," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1728–1735, Sept. 1993.
- [19] A. Scaglione and U. Sakoglu, "Asymptotic capacity of space-time coding for arbitrary fading: A closed form expression using Girko's law," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 4, May 2001, pp. 2509–2511.
- [20] L. Zheng and D. N. C. Tse, "Packing spheres in the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inform. Theory*, vol. 48, pp. 359–383, Feb. 2002.

Mahesh Godavarti (S'95–A'95–M'01) was born in Jaipur, India, in 1972. He received the B.Tech. degree in electrical and electronics engineering from the Indian Institute of Technology, Madras, India, in 1993 and the M.S. degree in electrical and computer engineering from the University of Arizona, Tucson, in 1995. He was at the University of Michigan, Ann Arbor, from 1997 to 2001 from where he received the M.S. degree in applied mathematics and the Ph.D. degree in electrical engineering.

Currently, he is employed as a Staff DSP Engineer with Ditech Communications, Mountain View, CA, where he is researching new algorithms for speech enhancement. His research interests include topics in speech and signal processing, communications, and information theory.

Alfred O. Hero (S'79–M'80–SM'96–F'98) was born in Boston, MA, in 1955. He received the B.S. degree in electrical engineering (*summa cum laude*) from Boston University in 1980 and the Ph.D. degree from Princeton University, Princeton, NJ, in 1984, both in electrical engineering. While at Princeton, he held the G.V.N. Lothrop Fellowship in Engineering.

Since 1984, he has been a Professor with the University of Michigan, Ann Arbor, where he has appointments in the Department of Electrical Engineering and Computer Science, the Department of Biomedical Engineering, and the Department of Statistics. He has held visiting positions at I3S University of Nice, Sophia-Antipolis, France, in 2001, Ecole Normale Supérieure de Lyon, Lyon, France, in 1999, Ecole Nationale Supérieure des Télécommunications, Paris, France, in 1999, Scientific Research Labs of the Ford Motor Company, Dearborn, MI, in 1993, Ecole Nationale Supérieure des Techniques Avancées (ENSTA) and Ecole Supérieure d'Electricité, Paris, in 1990, and Lincoln Laboratory, Massachusetts Institute of Technology, Cambridge, from 1987 to 1989. His research has been supported by NIH, NSF, AFOSR, NSA, ARO, ONR, DARPA, and by private industry in the areas of estimation and detection, statistical communications, bioinformatics, signal processing, and image processing.

Dr. Hero served as Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY from 1995 to 1998 and again in 1999 and the IEEE TRANSACTIONS ON SIGNAL PROCESSING since 2002. He was Chairman of the Statistical Signal and Array Processing (SSAP) Technical Committee from 1997 to 1998 and Treasurer of the Conference Board of the IEEE Signal Processing Society. He was Chairman for Publicity of the 1986 IEEE International Symposium on Information Theory (Ann Arbor, MI) and General Chairman of the 1995 IEEE International Conference on Acoustics, Speech, and Signal Processing (Detroit, MI). He was Co-Chair of the 1999 IEEE Information Theory Workshop on Detection, Estimation, Classification, and Filtering (Santa Fe, NM) and the 1999 IEEE Workshop on Higher Order Statistics (Caesaria, Israel). He chaired the 2002 NSF Workshop on Challenges in Pattern Recognition. He co-chaired the 2002 Workshop on Genomic Signal Processing and Statistics (GENSIPS). He was Vice President (Finance) of the IEEE Signal Processing Society from 1999 to 2002. He was Chair of Commission C (Signals and Systems) of the U.S. National Commission of the International Union of Radio Science (URSI) from 1999 to 2002. He has been a member of the Signal Processing Theory and Methods (SPTM) Technical Committee of the IEEE Signal Processing Society since 1999. He is also a member of the SEDD Review Panel of the U.S. National Research Council. He will be President-Elect of the IEEE Signal Processing Society from 2004 to 2005. He is a member of Tau Beta Pi, the American Statistical Association (ASA), the Society for Industrial and Applied Mathematics (SIAM), and the U.S. National Commission (Commission C) of URSI. He received the 1998 IEEE Signal Processing Society Meritorious Service Award, the 1998 IEEE Signal Processing Society Best Paper Award, and the IEEE Third Millennium Medal. In 2002, he was appointed IEEE Signal Processing Society Distinguished Lecturer.