A short information theoretic proof of CLT

Peter Harremoës, *Member IEEE* and Christophe Vignat

***Abstract*—A short proof of the Central Limit Theorem is given using a lemma about continuity of differential entropy.**

***Index Terms*— Central Limit Theorem, differential entropy, Fisher information, information divergence, Stam's inequality.**

I. Introduction

The first paper relating the Central Limit Theorem and information theory was [1]. Later [2] and [3] used Fisher information to prove an information theoretic version of the Central Limit Theorem. These proofs involve technical bounds on the tails of the distributions. Results on the rate of convergence can be found in [3], [4], [5] and [6]. Some recent references are [7] and [8].

Let X be a random variable with $E(X) = 0$, $\sigma(X) = 1$ and density f. The *differential entropy* is $h(X) = -\int f(x)\log(f(x))\ dx$. The *information divergence* is $D(X) = \int f(x)\log(f(x)/\phi(x))\ dx$ where ϕ is the normal density. The *Fisher information* of X is $I(X) = \int (f'(x))^2 / f(x)\ dx$. When X and Y are independent they satisfy *Stam's Inequality*

$$I^{-1}(X) + I^{-1}(Y) \leq I^{-1}(X+Y)$$

with equality if and only if X and Y are normal [9]. The *normalized Fisher information* $J(X) = \sigma^2(X) I(X) - 1$ is non-negative according to the Cramér-Rao Inequality. If $X^t = e^{-t}X + \left(1-e^{-2t}\right)^{\frac{1}{2}} N$ with N normal and independent of X then, according to [7],

$$D(X) = \int_0^\infty J\left(X^t\right)\ dt.$$

II. The Central Limit Theorem

Lemma 1: For any sequence $X_1, X_2, \ldots$ of random variables with $E(X_i) = 0$, $\sigma(X_i) = 1$ and densities f_i uniformly bounded and pointwise converging to f, $h(f_n) \to h(f)$ and $D(f_n) \to D(f)$ for $n \to \infty$.

Proof: First $h(f) = h(\phi) - D(f)$ implies that $\limsup h(f_n) \geq h(f)$ by lower semi continuity of information divergence. Put $C = \sup_{n,x} f_n(x)$ so that $-\frac{f_n(x)}{C}\log\frac{f_n(x)}{C} \geq 0$. Then

$$h(f_n) = C\int\left(-\frac{f_n(x)}{C}\log\frac{f_n(x)}{C}\right)dx - \log C,$$

and $\liminf h(f_n) \geq h(f)$ by Fatou's Lemma. ∎

Theorem 2: Let $X_1, X_2, \ldots$ be independent identically distributed random variables with $E(X_i) = 0$ and $\sigma(X_i) = 1$. Put $S_n = \sum_{i=1}^n X_i / n^{1/2}$. Then $D(S_n) \to 0$ for $n \to \infty$ if and only if $D(S_n)$ is finite eventually.

Proof: Assume that $D(S_m) < \infty$. For any $t \in]0;\infty[$ the sequence $J(S_n^t)$ is decreasing according to [7], and thus $J(S_n^t)$ converges to some function $t \curvearrowright g(t)$[1]. Using dominated convergence we only have to prove that $g(t) = 0$ for all $t > 0$. Now,

$$D(S_n^v) = \int_v^\infty J\left(S_n^t\right)\ dt \to \int_v^\infty g(t)\ dt \text{ for } n \to \infty$$

so it suffices to prove that $D(S_n^v) \to 0$ for $n \to \infty$.

The set $\{P \mid D(P) \leq K\}$ is compact so there exists a subsequence S_{n_m} whose distribution converges to the distribution of some random variable Y. Then for $v > 0$ the density of $S_{n_m}^v$ converges to the density of Y^v. Further, the density of $S_{2n_m}^v$ converges to the density of $\left(Y^v + \tilde{Y}^v\right)/2^{1/2}$ where Y and $\tilde{Y}$ are independent and identically distributed. As $m \to \infty$

$$D\left(S_{n_m}^v\right) \to D(Y^v) \text{ and } D\left(S_{2n_m}^v\right) \to D\left(\frac{Y^v + \tilde{Y}^v}{2^{1/2}}\right),$$

which implies that $D(Y^v) = D\left(\frac{Y^v+\tilde{Y}^v}{2^{1/2}}\right)$. In particular, $J(Y^w) = J\left(\frac{Y^w+\tilde{Y}^w}{2^{1/2}}\right)$ for $w > v$ which implies that Y^w is a normal random variable and that $g(w) = J(Y^w) = 0$. ∎

References

[1] Y. V. Linnik, "An information-theoretic proof of the central limit theorem with Lindeberg condition," *Theory Probab. Appl.*, vol. 4, pp. 288–299, 1959.

[2] A. R. Barron, "Entropy and the Central Limit Theorem," *Annals Probab. Theory*, vol. 14, no. 1, pp. 336 – 342, 1986.

[3] O. Johnson and A. R. Barron, "Fisher information inequalities and the central limit theorem." To appear.

[4] K. Ball, F. Barthe, and A. Naor, "Entropy jumps in the presence of a spectral gap," *Duke Math. J*, vol. 119, no. 1, pp. 41–63, 2003.

[5] P. Harremoës, "Lower bound on rate of convergence in information theoretic central limit theorem," in *Book of Abstracts for the Seventh International Symposium on Orthogonal polynomials, Special functions and Applications, Copenhagen*, pp. 53–54, 18-22/8 2003.

[6] P. Harremoës, "Lower bounds on divergence in central limit theorem." To appear, 2004.

[7] S. Artstein, K. Ball, F. Barthe, and A. Naor, "Solution of Shannon's problem on the monotonicity of entropy," *J. Amer. Math. Soc.*, 2004. To appear.

[8] O. Johnson, "Information theory and central limit theorem." To appear.

[9] N. M. Blachman, "The convolution inequality for entropy powers," *IEEE Trans. Inform. Theory*, vol. IT-11, pp. 267 – 271, April 1965.

P. Harremoës is with Dept. of Math., Univ. of Copenhagen, Denmark. E-mail: moes@math.ku.dk .

C. Vignat is with Laboratoire Systèmes de Communications, Université de Marne la Vallée, France, vignat@univ-mlv.fr . This work was done during a visit at Dept. of Math., Univ. of Copenhagen, April 2004.

The first author is supported by a post-doc fellowship from the Villum Kann Rasmussen Foundation and by INTAS (project 00-738) and by the Danish Natural Science Research Council.

[1] The result that $J\left(S_n^t\right)$ is decreasing is not really needed because one can use Stam's Inequality to prove subadditivity of $J\left(S_n^t\right)$ [2]. This implies convergence of $J\left(S_n^t\right)$ and decrease of the sequence $J\left(S_{2^n}^t\right)$.