Available online at www.sciencedirect.com
SCIENCE DIRECT®

ELSEVIER

Electronic Notes in Discrete Mathematics 21 (2005) 309–313

Electronic Notes in DISCRETE MATHEMATICS

www.elsevier.com/locate/endm

Lower Bounds for Divergence in Central Limit Theorem

Peter Harremoës [1]

Department of Mathematics, University of Copenhagen

Abstract

A method for finding asymptotic lower bounds on information divergence is developed and used to determine the rate of convergence in the Central Limit Theorem.

Keywords: Central Limit Theorem, cumulant, Hermite polynomial, information divergence, kurtosis, maximum entropy, rate of convergence, skewness.

1 Introduction

Recently Oliver Johnson and Andrew Barron [JB01] proved that the rate of convergence in the information theoretic Central Limit Theorem is upper bounded by $\frac{c}{n}$ under suitable conditions for some constant c. In general if $r_0 > 2$ is the smallest number such that the r'th moment does not vanish then a lower bound on total variation is $\frac{c}{n^{\frac{r_0}{2}-1}}$ for some constant c. Using Pinsker's inequality this gives a lower bound on information divergence of order $\frac{1}{n^{r_0-2}}$. In this paper more explicit lower bounds are computed. The idea is simple and follows general ideas related to the maximum entropy principle as described

[1] Supported by a Post. Doc. fellowship by the Villum Kann Rasmussen Foundation and by grants from Danish Natural Science Counsil and INTAS (project 00-738). This work was mainly done during a stay at ZIF, Bielefeld.

doi:10.1016/j.endm.2005.07.076

by Jaynes [Jay57]. If some of the higher moments of a random variable X are known the higher moments of the centered and normalized sum of independent copies of X can be calculated. Now, maximize the entropy given these moments. This is equivalent to minimize the divergence to the normal distribution. The distribution maximizing entropy with given moment constraints can not be calculated exactly but letting n go to infinity asymptotic results are obtained.

2 Existence of maximum entropy distributions

Let X be a random variable for which the moments of order $1, 2, ...R$ exists. Without loss of generality we will assume that $E(X) = 0$ and $Var(X) = 1$. The r'th central moment is denoted $\mu_r(X) = E(X^r)$. The *Hermite polynomials* $H_r(X)$ are the orthogonal polynomials with respect to the normal distribution. One easily translate between moments and the Hermite moments $E(H_r(X))$. Let r_0 denote the smallest number greater than 1 such that $E(H_r(X)) \neq 0$. Put $\gamma_0 = E(H_{r_0}(X))$.

It is well known that the normal distribution is the maximum entropy distribution for a random variable with specified first and second moment. It is also known that there exists no maximum entropy distribution if the first 3 moments are specified and the skewness is required to be non-zero [CT91].

Using that a polynomial of even order with positive leading coefficient dominates all polynomials of lower order for arguments outside a bounded set we get the following important result.

Theorem 2.1 *Let K be the convex set of distributions for which the first R moments are defined and satisfies the following equations and inequality*

$$E(H_r(X)) = h_r \text{ for } r < R$$
$$E(H_R(X)) \leq h_R\,.$$

If R is even then the maximum entropy distribution exists.

Corollary 2.2 *Let C be the convex set of distributions for which the first R moments are defined and satisfies the following equations*

$$E(H_r(X)) = h_r \text{ for } r \leq R\,.$$

If any of the following conditions are fulfilled

- *r_0 is odd and $R = r_0 + 1$*