Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications 126 (2016) 3733–3759

stochastic processes and their applications

www.elsevier.com/locate/spa

Asymptotic normality of quadratic estimators

James M. Robins[a,*], Lingling Li[a], Eric Tchetgen Tchetgen[a], Aad van der Vaart[b]

[a] Departments of Biostatistics and Epidemiology, School of Public Health, Harvard University, United States
[b] Mathematical Institute, Leiden University, Netherlands

Received 2 June 2015; accepted 17 February 2016
Available online 18 May 2016

Abstract

We prove conditional asymptotic normality of a class of quadratic U-statistics that are dominated by their degenerate second order part and have kernels that change with the number of observations. These statistics arise in the construction of estimators in high-dimensional semi- and non-parametric models, and in the construction of nonparametric confidence sets. This is illustrated by estimation of the integral of a square of a density or regression function, and estimation of the mean response with missing data. We show that estimators are asymptotically normal even in the case that the rate is slower than the square root of the observations.

MSC: 62G05; 62G20

Keywords: Quadratic functional; Projection estimator; Rate of convergence; U-statistic

1. Introduction

Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be i.i.d. random vectors, taking values in sets $\mathcal{X} \times \mathbb{R}$, for an arbitrary measurable space $(\mathcal{X}, \mathcal{A})$ and $\mathbb{R}$ equipped with the Borel sets. For given symmetric,

* Corresponding author.
E-mail addresses: robins@hsph.harvard.edu (J.M. Robins), lili@hsph.harvard.edu (L. Li), etchetge@hsph.harvard.edu (E.T. Tchetgen), avdvaart@math.leidenuniv.nl (A. van der Vaart).

http://dx.doi.org/10.1016/j.spa.2016.04.005
0304-4149/

measurable functions $K_n : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ consider the U-statistics

$$U_n = \frac{1}{n(n-1)} \sum\sum_{1 \le r \ne s \le n} K_n(X_r, X_s) Y_r Y_s. \tag{1}$$

Would the kernel $(x_1, y_1, x_2, y_2) \mapsto K_n(x_1, x_2) y_1 y_2$ of the U-statistic be independent of n and have a finite second moment, then either the sequence $\sqrt{n}(U_n - \mathrm{E}U_n)$ would be asymptotically normal or the sequence $n(U_n - \mathrm{E}U_n)$ would converge in distribution to Gaussian chaos. The two cases can be described in terms of the Hoeffding decomposition $U_n = \mathrm{E}U_n + U_n^{(1)} + U_n^{(2)}$ of U_n, where $U_n^{(1)}$ is the best approximation of $U_n - \mathrm{E}U_n$ by a sum of the type $\sum_{i=1}^n h(X_r, Y_r)$ and $U_n^{(2)}$ is the remainder, a degenerate U-statistic (compare (28) in Section 5). For a fixed kernel K_n the linear term $U_n^{(1)}$ dominates as soon as it is nonzero, in which case asymptotic normality pertains; in the other case $U_n^{(1)} = 0$ and the U-statistic possesses a nonnormal limit distribution.

If the kernel depends on n, then the separation between the linear and quadratic cases blurs. In this paper we are interested in this situation and specifically in kernels K_n that concentrate as $n \to \infty$ more and more near the diagonal of $\mathcal{X} \times \mathcal{X}$. In our situation the variance of the U-statistics is dominated by the quadratic term $U_n^{(2)}$. However, we show that the sequence $(U_n - \mathrm{E}U_n)/\sigma(U_n)$ is typically still asymptotically normal. The intuitive explanation is that the U-statistics behave asymptotically as "sums across the diagonal $r = s$" and thus behave as sums of independent variables. Our formal proof is based on establishing conditional asymptotic normality given a binning of the variables X_r in a partition of the set $\mathcal{X}$.

Statistics of the type (1) arise in many problems of estimating a functional on a semiparametric model, with K_n the kernel of a projection operator (see [18]). As illustrations we consider in this paper the problems of estimating $\int g^2(x)\,dx$ or $\int f^2(x)\,dG(x)$, where g is a density and f a regression function, and of estimating the mean treatment effect in missing data models. Rate-optimal estimators in the first of these three problems were considered by [2,3,12–14], among others. In Section 3 we prove asymptotic normality of the estimators in [12,13], also in the case that the rate of convergence is slower than $\sqrt{n}$, usually considered to be the "nonnormal domain". For the second and third problems estimators of the form (1) were derived in [18,19,24,22] using the theory of second-order estimating equations. Again we show that these are asymptotically normal, also in the case that the rate is slower than $\sqrt{n}$.

Statistics of the type (1) also arise in the construction of adaptive confidence sets, as in [21], where the asymptotic normality can be used to set precise confidence limits.

Previous work on U-statistics with kernels that depend on n includes [25,1,10,5,6]. These authors prove unconditional asymptotic normality using the martingale central limit theorem, under somewhat different conditions. Our proof uses a Lyapunov central limit theorem (with moment $2 + \varepsilon$) combined with a conditioning argument, and an inequality for moments of U-statistics due to E. Giné. Our conditions relate directly to the contraction of the kernel, and can be verified for a variety of kernels. The conditional form of our limit result should be useful to separate different roles for the observations, such as for constructing preliminary estimators and for constructing estimators of functionals. Another line of research (as in [15]) is concerned with U-statistics that are well approximated by their projection on the initial part of the eigenfunction expansion. This has no relation to the present work, as here the kernels explode and the U-statistic is asymptotically determined by the (eigen) directions "added" to the kernel as the number of observations increases. By making special choices of kernel and variables Y_i, the statistics (1) can reduce to certain chisquare statistics, studied in [16,8].