International Workshop on Empirical Processes and Asymptotic Statistics
Rennes, June 18-20 2007

Central Limit Theorem for multiple integrals with respect to the empirical process

Hélène Boistard

www.boistard.fr

joint work with Eustasio del Barrio

University of Toulouse and University of Valladolid

Introduction and motivation

Multiple integrals w.r.t. the empirical process (Cf. Major [4], [5].)

$$J_{n,m}(h) = \int' h(x_1, \ldots, x_m)\, d\mathbb{G}_n(x_1) \cdots d\mathbb{G}_n(x_m), \qquad (1)$$

where

- $h(x_1, ..., x_m)$ a real valued, symmetric function on $\mathcal{X}^m$,
- $X_1, \ldots, X_n$ i.i.d. $(X, \mathcal{X})$-valued r.v. with distribution P,
- $\int'$ denotes the integral in which diagonals $x_r = x_s$, $1 \le r < s \le m$ are omitted from the domain of integration,
- $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P)$ is the normalized empirical process.

U-statistics

$$U_n(h) = \frac{1}{\binom{n}{m}} \sum_{1 \le i_1 < \cdots < i_m \le n} h(X_{i_1}, \ldots, X_{i_m}). \tag{2}$$

For completely degenerate h and non atomic P :

$$J_{n,m}(h) = \frac{n!}{(n-m)!n^{m/2}} U_n(h).$$

Rubin and Vitale [7] : CLT for completely degenerate U-statistics.
For non degenerate statistics : CLT obtained via the Hoeffding decomposition.
Problem : the limit can be complicate (cf. Arcones and Giné [1] or de la Peña and Giné [3]).

Existing CLT for multiple integrals

Result of Major ([4], [5]) :

$$J_{n,m}(h) = \sum_{j=0}^{m} K_{n,j,m} J_{n,j}(\pi_j h), \tag{3}$$

with coefficients

$$K_{n,j,m} \to K_{j,m}, n \to \infty$$

and $\pi_j h$ Hoeffding projection of h, defined as :

$$\pi_j h(x_1, \dots, x_j) = \begin{cases} (\delta_{x_1} - P) \times \cdots \times (\delta_{x_j} - P) \times P^{m-j} h & \text{for } \quad j = 1, \dots, m \\ P^m h & \text{when } j = 0. \end{cases}$$

Consequence : $J_{n,m}(h)$ converges in distribution to

$$\sum_{j=0}^{m} K_{j,m} I_j(\pi_j h),$$

where I_j denotes the multiple Wiener integral of order j (cf. [6]).
Problem : unknown coefficients $K_{j,m}$.

Our method

Precise decomposition of type (3) for symmetric kernel h : cf. (5) below. Sort of Hoeffding decomposition for multiple integrals.

Convergence of the coefficients to explicit values (related to Hermite polynomials) : cf. (6).

Expression of the limit as a **stochastic multiple integral with respect to the Brownian bridge**.

Multiple integrals with respect to the Brownian bridge

Formal idea

$$\int'_{[0,1]^m} h(x_1,\dots,x_m)d\mathbb{G}_n(x_1)\cdots d\mathbb{G}_n(x_m) \underset{w}{\to} \int_{[0,1]^m} h(x_1,\dots,x_m)dB(x_1)\cdots dB(x_m), \quad (4)$$

with $\{B(x)\}_{x\in(0,1)}$ a Brownian bridge. Expansion of (4) using

$B(x) = W(x) - xW(1)$ with $\{W(x), x \in (0,1)\}$ a Brownian motion :

$$\begin{aligned}\int_{[0,1]^m} h(x_1,\dots,x_m)dB(x_1)\cdots dB(x_m) &= \int_{[0,1]^m} h(x_1,\dots,x_m)dW(x_1)\cdots dW(x_m)\\ &-m\int_{[0,1]^{m-1}} \rho_{m-1}h(x_1,\dots,x_{m-1})dW(x_1)\cdots dW(x_{m-1})W(1)\\ &-\binom{m}{2}\int_{[0,1]^{m-2}} \rho_{m-2}h(x_1,\dots,x_{m-2})dW(x_1)\cdots dW(x_{m-2})W(1)^2\\ &+\cdots+(-1)^m\rho_0 hW(1)^m,\end{aligned}$$

where $\rho_j h(x_1,\dots,x_j) = \begin{cases}\delta_{x_1}\times\cdots\times\delta_{x_j}\times P^{m-j}h & \text{for } j=1,\dots,m,\\ P^m h & \text{for } j=0\ .\end{cases}$

Definition of the integral with respect to the Brownian bridge

Definition 1. *Given a symmetric function $h \in L^2(P^m)$, we define*

$$J_m(h) = \sum_{j=0}^{m} (-1)^{m-j} \binom{m}{m-j} I_j(\rho_j h)(I_1(1))^{m-j}.$$

Remark : h is completely degenerate if $\rho_j h = 0$ for $j < m$. Then, $J_m(h) = I_m(h)$.

Proposition 1. *For symmetric $h \in L^2(P^m)$*

$$J_m(h) = \sum_{j=0}^{m} \frac{m!}{j!} H_{m-j}(0) I_j(\pi_j h),$$

where H_{m-j} is the Hermite polynomial of degree $m-j$ (cf. [6]).

CLT for multiple integrals

Theorem 1. *Let* $(X_1, \ldots, X_n)$ *be an i.i.d. sample with non atomic distribution* P *and* $h_1, \ldots, h_k$ *square integrable kernels in* $m_1, \ldots, m_k$ *variables. Then*

$$(J_{n,m_1}(h_1), \ldots, J_{n,m_k}(h_k)) \xrightarrow{w} (J_{m_1}(h_1), \ldots, J_{m_k}(h_k)) \,.$$

Proof : for the completely degenerate case, the same as for U-statistics. For the non degenerate case : based on Theorem 2.

Hoeffing decomposition for the multiple integral

Theorem 2. *Let $(X_1, \ldots, X_n)$ be an i.i.d. sample with non-atomic distribution P and h a L_2 symmetric kernel of m variables. We have the following decomposition of $J_{n,m}(h)$ as a sum of multiple integrals based on degenerate kernels :*

$$J_{n,m}(h) = \sum_{j=0}^{m} K_{n,j,m} J_{n,j}(\pi_j h) \tag{5}$$

where the coefficients are for $j \leq m$:

$$K_{n,j,m} = n^{-(m-j)/2} \sum_{k=0}^{m-j} (-1)^k n^k \binom{m}{k} \binom{m-k}{m-k-j} \frac{(n-j)!}{(n-m+k)!}.$$

Moreover, the coefficients satisfy the following convergence :

$$\lim_{n \to \infty} K_{n,j,m} = \frac{m!}{j!} H_{m-j}(0), \tag{6}$$

where for all $k \geq 0$, H_k is the Hermite polynomial of order k.

Sketch of the proof when $m = 2$

Let $\tilde{h}(x,y) = \begin{cases} h(x,y), & x \neq y \\ 0 & \text{else.} \end{cases}$

$$J_{n,2}(h) = \frac{1}{n} \sum_{1 \le i,j \le n} \int \int \tilde{h}(x,y) d(\delta_{X_i} - P)(x) d(\delta_{X_j} - P)(y)$$

$$\begin{aligned}
&= \frac{1}{n} \sum_{i \neq j} h(X_i, X_j) - 2 \sum_{i=1}^{n} \int h(X_i, x) dP(x) + n \int \int h(x,y) dP(x) dP(y) \\
&= \frac{1}{n} \sum_{i \neq j} (\pi_2 h(X_i, X_j) + \pi_1 h(X_i) + \pi_1 h(X_j) + \pi_0 h) - 2 \sum_{i=1}^{n} (\pi_1 h(X_i) + \pi_0 h) + n \pi_0 h \\
&= J_{n,2}(\pi_2 h) - \frac{2}{\sqrt{n}} J_{n,1}(\pi_1 h) - I_0(\pi_0 h) \\
&\xrightarrow{w} I_2(\pi_2 h) - I_0(\pi_0 h) = J_2(h).
\end{aligned}$$

Complete proof in Boistard and del Barrio [2].

References

[1] M.A. Arcones and E. Giné. On the bootstrap of U and V-statistics. *Ann. Statist.*, **20**(2) : 655–674, 1992.

[2] H. Boistard and E. del Barrio. Central Limit Theorem for multiple integrals with respect to the empirical process. *Submitted.*

[3] V.H. de la Peña and E. Giné. *Decoupling*. Probability and its applications. Springer-Verlag, New York, 1999.

[4] P. Major. On the tail behaviour of multiple random integrals and degenerate U-statistics. *Preprint.*

[5] P. Major. An estimate on the supremum of a nice class of stochastic integrals and U-statistics. *Probab. Theory Related Fields*, **134**(3) : 489–537, 2006.

[6] D. Nualart. *Malliavin calculus and related topics.* Probability and its applications. Springer-Verlag, New York, 1995.

[7] H.Rubin and R.A. Vitale. Asymptotic distribution of symmetric statistics. *Ann. Statist.*, **8**(1) : 165–170, 1980.