A dichotomy for CLT in total variation

Aihua Xia (夏爱华)

School of Mathematics and Statistics

The University of Melbourne, VIC 3010

16 July, 2016

Lebesgue decomposition

- Every distribution F on $\mathbb{R}$ can be written as

$$F = pF_s + (1-p)F_a,$$

  where $p \in [0,1]$, F_s is singular and F_a is absolutely continuous with respect to the Lebesgue measure.

  - Our reference measure is the Lebesgue measure because our limit is the normal distribution.

Some elementary observations

- Let X_i's be iid taking values 0 and 1 with equal probability, then
  - $X = \sum_{k=1}^{\infty} 2^{-k} X_k$ has uniform distribution on $(0, 1)$: this is a binary expansion.
  - $Y = 3 \sum_{r=1}^{\infty} 4^{-r} X_r$ has singular distribution on $(0, 1)$ and is called *Cantor-type* distribution.
  - $U = \sum_{k=1}^{\infty} 2^{-2k} X_{2k}$, $V = \sum_{k=1}^{\infty} 2^{-(2k-1)} X_{2k-1}$, then U and V are independent, $U \stackrel{\mathrm{d}}{=} 2V$, both U and V have singular distributions.
    * The sum of independent singular rvs may be absolutely continuous!
  - $V \stackrel{\mathrm{d}}{=} \frac{2}{3} Y$.

Berry-Esseen thm

Let η_i's be iid with $\mathbb{E}\eta_i = 0$, $\mathrm{Var}(\eta_i) = 1$ and finite third moment, $Y_n = \frac{\sum_{i=1}^n \eta_i}{\sqrt{n}}$, then

$$d_K(Y_n, Z) := \sup_{x\in\mathbb{R}} |\mathbb{P}(Y_n \leq x) - \mathbb{P}(Z \leq x)| \leq \frac{c\mathbb{E}|\eta_1|^3}{\sqrt{n}},$$

where $Z \sim N(0, 1)$.

Questions

1. What is the speed of convergence in the total variation:

$$d_{TV}(Y_n, Z) := \sup_{A \in \mathscr{B}(\mathbb{R})} |\mathbb{P}(Y_n \in A) - \mathbb{P}(Z \in A)|?$$

2. How about a sequence with dependence?

Known results: all under independence

- Prohorov (1952): $d_{TV}(Y_n, Z) \equiv 1$ for all n or $d_{TV}(Y_n, Z) = o(1)$.
  - If $\mathbb{E}|\eta_1|^3 < \infty$, $d_{TV}(Y_n, Z) = o(n^{-1/2}(\ln n)^{1/2})$.
- Bally and Caramellino (2016): generalisation to higher dimension with mixture distribution, in particular, $d_{TV}(Y_n, Z) = O(n^{-1/2})$.

How to get there?

1. Characteristic functions.
2. Coupling method.
3. Stein's method.

A proof using Stein's method

- *Integration by parts and Stein's method*: for differentiable bounded function f on $\mathbb{R}$ with bounded derivative f', let ϕ be the pdf of Z, then

$$\mathbb{E}f'(Z) = \int f'(z)\phi(z)dz = \int zf(z)\phi(z)dz = \mathbb{E}Zf(Z).$$

- $Z \sim N(0,1)$ iff

$$\mathbb{E}[f'(Z) - Zf(Z)] = 0$$

for a sufficiently rich class of functions f.

- For an $A \in \mathscr{B}(\mathbb{R})$, we want to estimate $\mathbb{E}\mathbf{1}_A(Y_n) - \mathbb{E}\mathbf{1}_A(Z)$, so we consider

$$f'(w) - wf(w) = \mathbf{1}_A(w) - \mathbb{E}\mathbf{1}_A(Z).$$

  – This is called *Stein's equation*.

  – One can solve the differential equation to get

$$\begin{aligned} f_A(w) &= e^{w^2/2}\int_{-\infty}^{w}(\mathbf{1}_A(x) - \mathbb{E}\mathbf{1}_A(Z))e^{-x^2/2}dx \\ &= -e^{-w^2/2}\int_{w}^{\infty}(\mathbf{1}_A(x) - \mathbb{E}\mathbf{1}_A(Z))e^{-x^2/2}dx. \end{aligned}$$

  – The properties of f_A:

$$\|f'_A\| := \sup_{x\in\mathbb{R}}\left|f'_A(x)\right| \le 2\|\mathbf{1}_A(\cdot) - \mathbb{E}\mathbf{1}_A(Z)\| \le 2.$$

- If there exists an m_0 such that $d_{TV}(Y_{m_0}, Z) < 1$, we can define $\eta_i' = \sum_{j=(i-1)m_0+1}^{im_0} \eta_j$, so without loss, we assume $d_{TV}(\eta_1, Z) < 1$.
- Write $Y_n' = Y_n - \eta_1/\sqrt{n}$, then

$$
\begin{aligned}
&\mathbb{E}[f'(Y_n) - Y_n f(Y_n)] = \dots \\
&= \mathbb{E}\left\{\mathbb{E}\left[f'(Y_n' + \eta_1/\sqrt{n}) - f'(Y_n')|\eta_1\right]\right\} \\
&- \int_0^1 \mathbb{E}\left\{\eta_1^2 \mathbb{E}\left[f'(Y_n' + u\eta_1/\sqrt{n}) - f'(Y_n')|\eta_1\right]\right\} du.
\end{aligned}
$$

- The estimate:

$$
\begin{aligned}
&|\mathbb{P}(Y_n \in A) - \mathbb{P}(Z \in A)| \\
&= \Big|\mathbb{E}[f'(Y_n) - Y_n f(Y_n)]\Big| \\
&\le 2\|f'\| \int_{\mathbb{R}} d_{TV}(Y_n', Y_n' + r/\sqrt{n}) dF_{\eta_1}(r) \\
&\quad + 2\|f'\| \int_0^1 \int_{\mathbb{R}} r^2 d_{TV}(Y_n', Y_n' + ur/\sqrt{n}) dF_{\eta_1}(r) du \\
&\le 4 \int_{\mathbb{R}} d_{TV}(Y_n', Y_n' + r/\sqrt{n}) dF_{\eta_1}(r) \\
&\quad + 4 \int_0^1 \int_{\mathbb{R}} r^2 d_{TV}(Y_n', Y_n' + ur/\sqrt{n}) dF_{\eta_1}(r) du.
\end{aligned}
$$

What's crucial?

- $d_{TV}(Y_n, Y_n + v/\sqrt{n}) = d_{TV}(S_n, S_n + v)$, where $S_n = \sum_{i=1}^n \eta_i$.

- Assume $\xi_1, \ldots, \xi_n$ are iid random variables having the triangular density function

$$\kappa_a(x) = \begin{cases} \frac{1}{a}\left(1 - \frac{|x|}{a}\right), & \text{for } |x| \le a, \\ 0, & \text{for } |x| > a, \end{cases} \tag{1}$$

where $a > 0$. Let $T_n = \sum_{i=1}^n \xi_i$. Then for any $\gamma > 0$,

$$d_{TV}(T_n, T_n + \gamma) \le \frac{\gamma}{a}\left\{\sqrt{\frac{3}{\pi n}} + \frac{2}{(2n-1)\pi^{2n}}\right\}. \tag{2}$$

Why?

- The pdf of T_n is symmetric and unimodel.
  - **NB** The convolution of two unimodal pdfs is generally not unimodal
- Let G_n and g_n be the cdf and pdf of T_n, then

$$d_{TV}(T_n, T_n{+}r) = \sup_x |G_n(x) - G_n(x-r)| = \int_{-r/2}^{r/2} g_n(x)dx.$$

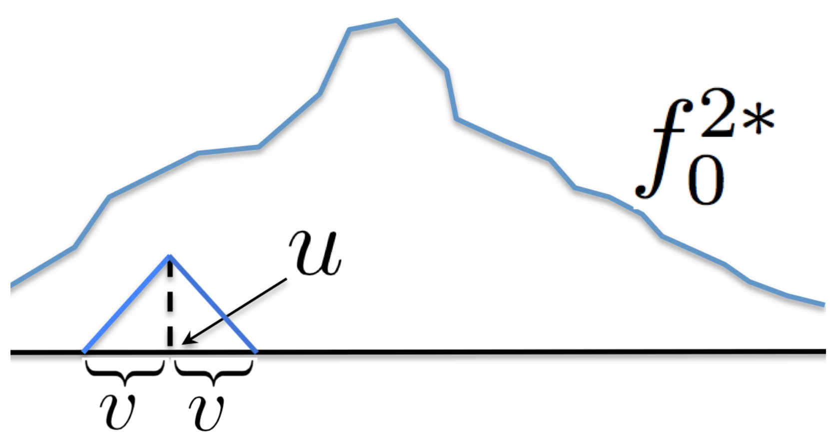
- $g_n(0) \le \frac{1}{a}\left\{\sqrt{\frac{3}{\pi n}} + \frac{2}{(2n{-}1)\pi^{2n}}\right\}$. $\square$

Mixing distribution

If F is non-singular, then there exist $a > 0$, $u \in \mathbb{R}$ and $\theta \in (0, 1]$ such that, with H_1 being the distribution of κ_a,

$$F^{2*} = (1 - \theta)H_2 + \theta H_1 * \delta_u.$$

Why? F is non-singular so there exists a bounded function with bounded support $f_0 \not\equiv 0$ such that $F(A) \geq \int_A f_0(x)dx$. Then f_0^{2*} is continuous.



The bound

For $v > 0$, we have

$$d_{TV}(S_n, S_n + v) \leq (v \vee 1) O(n^{-1/2}),$$

where $O(n^{-1/2})$ does not depend on v.

Why?

- Let $\xi_1 = \eta_1 + \eta_2$, $\xi_2 = \eta_3 + \eta_4$, ..., then ξ_i has a mixture distribution with one component of being triangular.
- $m = \lfloor n/2 \rfloor$, there exists $X_{1j} \sim H_1$, $X_{2j} \sim H_2$ and $X_{3j} \sim$Bernoulli(θ) such that

$$S'_m := \sum_{j=1}^{m} [(X_{1j} + u)X_{3j} + X_{2j}(1 - X_{3j})] \stackrel{\mathrm{d}}{=} \sum_{i=1}^{m} \xi_i.$$

- Let $I = \sum_{j=1}^{m} X_{3j} \sim$Bi$(m, \theta)$, then

$$S'_m \sim \sum_{k=0}^{m} \mathbb{P}(I = k)(H_1 + \delta_u)^{k*} * H_2^{(m-k)*}.$$

- $\mathbb{P}(I \le \lfloor 0.5m\theta \rfloor - 1) = O(m^{-1})$.
- For $k \ge \lfloor 0.5m\theta \rfloor$, it gives bound $\gamma(m^{-1/2})$. $\Box$

Remarks

- ChFs can work easily from identical distribution to non-identical distributions, with some complexity of formulation.
- ChFs fail completely when dependence is present.
- From independence to dependence: yes, a mixing condition is needed and the variance of S_n must become large when $n \to \infty$.

A warning example

- Recall that we can define independent U and V such that both are singular but $U + V$ $\sim$uniform.
- Let $U_i \overset{\mathrm{d}}{=} U - \mathbb{E}U$ and $V_i \overset{\mathrm{d}}{=} V - \mathbb{E}V$ be all independent.
- Consider

$$(U_0 + V_1) + (-V_1 - U_1) + (U_1 + V_2) + (-V_2 - U_2) + \ldots$$

  - This sequence is 1-dependent.
  - $(U_0 + V_1)$, $(-V_1 - U_1)$, $\ldots$ all follow the uniform distribution on $(-0.5, 0.5)$.
  - No CLT for the sum.
  - Hence, mixing condition is not enough: more is needed.

A puzzling fact

- By ChFs, when η_i's are iid with $k \geq 3$ moments, $d_{TV}(Y_n, Z) = O(n^{(k-2)/2})$.
- For $k > 3$, Stein's method has never achieved such a result.

Thank you!