Échange autour de la méthode de Stein

Omar Kassi

Séminaire des doctorants, 14 janvier

CREST

ENSAI École nationale de la statistique et de l'analyse de l'information

- Z_n is said to converge in distribution to a limiting random variable Z if

$$\lim_{n\to\infty} \mathbb{P}(Z_n \leq t) = \mathbb{P}(Z \leq t)$$

- Z_n is said to converge in distribution to a limiting random variable Z if

$$\lim_{n\to\infty} \mathbb{P}(Z_n \leq t) = \mathbb{P}(Z \leq t)$$

- Equivalent to saying that for all bounded continuous functions $g : \mathbb{R} \to \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{E}[g(Z_n)] = \mathbb{E}[g(Z)].$$

- Z_n is said to converge in distribution to a limiting random variable Z if

$$\lim_{n\to\infty} \mathbb{P}(Z_n \leq t) = \mathbb{P}(Z \leq t)$$

- Equivalent to saying that for all bounded continuous functions $g : \mathbb{R} \to \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{E}[g(Z_n)] = \mathbb{E}[g(Z)].$$

- It is not necessary to consider all bounded continuous g, but only g belonging to a smaller class such as $g(x) = e^{itx}$ with $t \in \mathbb{R}$ is arbitrary

There are three classical approaches to proving central limit theorems

- The method of characteristic functions, one simply has to show that for each $t \in \mathbb{R}$

$$\lim_{n\to\infty} \mathbb{E}e^{itZ_n} = e^{i\mu t-\sigma^2 t^2/2}.$$

There are three classical approaches to proving central limit theorems

- The method of characteristic functions, one simply has to show that for each $t \in \mathbb{R}$

$$\lim_{n\to\infty} \mathbb{E}e^{itZ_n} = e^{i\mu t - \sigma^2 t^2/2}.$$

- The method of moments, which involves showing that $\forall k \in \mathbb{N}$

$$\lim_{n\to\infty} \mathbb{E}Z_n^k = \mathbb{E}Z^k.$$

There are three classical approaches to proving central limit theorems

- The method of characteristic functions, one simply has to show that for each $t \in \mathbb{R}$

$$\lim_{n\to\infty} \mathbb{E}e^{itZ_n} = e^{i\mu t-\sigma^2 t^2/2}.$$

- The method of moments, which involves showing that $\forall k \in \mathbb{N}$

$$\lim_{n\to\infty} \mathbb{E}Z_n^k = \mathbb{E}Z^k.$$

- There is an old technique of Lindeberg

- Ross, N. (2011). Fundamentals of Stein's method. Probab. Surv., 8 :210–293
- Chatterjee, S. (2014). A short survey of Stein's method. International Congress of Mathematicians (ICM), pages 1–24.
- A. D. Barbour and L. H. Y. Chen, An introduction to Stein's method. World Scientific, 2005, vol. 4.
- L. H. Y. Chen, L. Goldstein and Q.-M. Shao (2011), Normal approximation by Stein's method, Probability and its Applications (New York), Springer, Heidelberg.

Probability metrics

For two probability measures μ and ν, the probability metrics we use have the form

$$d_{\mathcal{H}}(\mu,\nu) = \sup_{h\in\mathcal{H}} \left| \int h(x)\mathrm{d}\mu(\mathrm{x}) - \int \mathrm{h}(\mathrm{x})\mathrm{d}\nu(\mathrm{x}) \right|,$$

where $\mathcal{H}$ is some family of *test* function

Probability metrics

For two probability measures μ and ν, the probability metrics we use have the form

$$d_{\mathcal{H}}(\mu,\nu) = \sup_{h\in\mathcal{H}} \left| \int h(x)\mathrm{d}\mu(\mathrm{x}) - \int \mathrm{h}(\mathrm{x})\mathrm{d}\nu(\mathrm{x}) \right|,$$

where $\mathcal{H}$ is some family of *test* function

- If $\mathcal{H} = \{\mathbb{I}\{\cdot \leq x\}; x \in \mathbb{R}\}$, we obtain the *Kolmogorov metric* d_K

Probability metrics

For two probability measures μ and ν, the probability metrics we use have the form

$$d_{\mathcal{H}}(\mu,\nu) = \sup_{h\in\mathcal{H}} \left| \int h(x)\mathrm{d}\mu(\mathrm{x}) - \int \mathrm{h}(\mathrm{x})\mathrm{d}\nu(\mathrm{x}) \right|,$$

where $\mathcal{H}$ is some family of *test* function

- If $\mathcal{H} = \{\mathbb{I}\{\cdot \le x\}; x \in \mathbb{R}\}$, we obtain the *Kolmogorov metric* d_K
- If $\mathcal{H} = \{\mathbb{I}\{\cdot \in A\}; A \in \mathrm{Borel}(\mathbb{R})\}$, we obtain the *total variation metric* d_{TV}

Probability metrics

For two probability measures μ and ν, the probability metrics we use have the form

$$d_{\mathcal{H}}(\mu,\nu) = \sup_{h\in\mathcal{H}} \left| \int h(x)\mathrm{d}\mu(\mathrm{x}) - \int \mathrm{h}(\mathrm{x})\mathrm{d}\nu(\mathrm{x}) \right|,$$

where $\mathcal{H}$ is some family of *test* function

- If $\mathcal{H} = \{\mathbb{I}\{\cdot \leq x\}; x \in \mathbb{R}\}$, we obtain the *Kolmogorov metric* d_K
- If $\mathcal{H} = \{\mathbb{I}\{\cdot \in A\}; A \in \text{Borel}(\mathbb{R})\}$, we obtain the *total variation metric* d_{TV}
- If $\mathcal{H} = \{h : \mathbb{R} \to \mathbb{R}; |h(x) - h(y)| \leq |x - y|\}$, we obtain the *Wasserstein metric* d_W

Stein's lemma

The standard normal distribution is the only probability distribution that satisfies the equation

$$\mathbb{E}[Zf(Z)] = \mathbb{E}[f'(Z)] \tag{1}$$

for all continuous f with derivative f' such that $\mathbb{E}[f'(Z)] < \infty$.

The basic idea

- Let $Z \sim \mathcal{N}(0,1)$. Take any bounded measurable function g

The basic idea

- Let $Z \sim \mathcal{N}(0,1)$. Take any bounded measurable function g
- Let f be a bounded solution of the ODE

$$f'(x) - xf(x) = g(x) - \mathbb{E}[g(Z)]$$

Stein showed that a bounded solution always exists

The basic idea

- Let $Z \sim \mathcal{N}(0,1)$. Take any bounded measurable function g
- Let f be a bounded solution of the ODE

$$f'(x) - xf(x) = g(x) - \mathbb{E}[g(Z)]$$

  Stein showed that a bounded solution always exists
- We have for any random variable W :

$$\mathbb{E}[g(W)] - \mathbb{E}[g(Z)] = \mathbb{E}\left[f'(W) - Wf(W)\right]. \qquad (2)$$

- Let $a \in \mathbb{R}$ and $g(\cdot) = \mathbb{I}\{\cdot \leq a\}$

- Let $a \in \mathbb{R}$ and $g(\cdot) = \mathbb{I}\{\cdot \leq a\}$
- The unique bounded solution f_a of the ODE

$$f_a'(x) - xf_a(x) = \mathbb{I}\{x \leq a\} - \Phi(a)$$

is given by

$$f_a(x) = e^{x^2/2} \int_x^{\infty} e^{-t^2/2}(\Phi(a) - \mathbb{I}\{t \leq a\})\mathrm{d}t$$

- Let $a \in \mathbb{R}$ and $g(\cdot) = \mathbb{I}\{\cdot \leq a\}$
- The unique bounded solution f_a of the ODE

$$f_a'(x) - xf_a(x) = \mathbb{I}\{x \leq a\} - \Phi(a)$$

is given by

$$f_a(x) = e^{x^2/2} \int_x^{\infty} e^{-t^2/2}(\Phi(a) - \mathbb{I}\{t \leq a\})\mathrm{d}t$$

- As a result we have for any random variable W

$$|\mathbb{P}(W \leq a) - \Phi(a)| = \left|\mathbb{E}\left[f_a'(W) - Wf_a(W)\right]\right|$$

The general setup

- For two random variables X and Y and some family of functions $\mathcal{H}$, recall the metric

$$d_{\mathcal{H}}(X, Y) = \sup_{h \in \mathcal{H}} |\mathbb{E}[h(X)] - \mathbb{E}[h(Y)]|$$

The general setup

- For two random variables X and Y and some family of functions $\mathcal{H}$, recall the metric

$$d_{\mathcal{H}}(X,Y) = \sup_{h\in\mathcal{H}} |\mathbb{E}[h(X)] - \mathbb{E}[h(Y)]|$$

- For $h \in \mathcal{H}$, let f_h solve

$$f_h'(x) - xf_h(x) = h(x) - \underbrace{\Phi(h)}_{=\mathbb{E}[h(Z)]}$$

The general setup

- For two random variables X and Y and some family of functions $\mathcal{H}$, recall the metric

$$d_{\mathcal{H}}(X, Y) = \sup_{h \in \mathcal{H}} |\mathbb{E}[h(X)] - \mathbb{E}[h(Y)]|$$

- For $h \in \mathcal{H}$, let f_h solve

$$f_h'(x) - x f_h(x) = h(x) - \underbrace{\Phi(h)}_{=\mathbb{E}[h(Z)]}$$

- We have Therefore

$$d_{\mathcal{H}}(X, Z) = \sup_{h \in \mathcal{H}} \left| \mathbb{E}\left[f_h'(W) - W f_h(W) \right] \right|$$

Let f_h be the solution of the differential equation

$$f_h'(x) - xf_h(x) = h(x) - \Phi(h)$$

which is given by

$$f_h(x) = e^{x^2/2}\int_x^\infty e^{-t^2/2}(\Phi(h) - h(t))\mathrm{d}t.$$

Let f_h be the solution of the differential equation

$$f_h'(x) - xf_h(x) = h(x) - \Phi(h)$$

which is given by

$$f_h(x) = e^{x^2/2} \int_x^\infty e^{-t^2/2}(\Phi(h) - h(t))\mathrm{d}t.$$

1. If h is bounded, then

$$\|f_h\|_\infty \leq \sqrt{\frac{\pi}{2}}\|h(\cdot)-\Phi(h)\|_\infty, \quad \text{and } \|f_h'\|_\infty \leq 2\|h(\cdot)-\Phi(h)\|_\infty.$$

Let f_h be the solution of the differential equation

$$f_h'(x) - x f_h(x) = h(x) - \Phi(h)$$

which is given by

$$f_h(x) = e^{x^2/2} \int_x^\infty e^{-t^2/2} (\Phi(h) - h(t)) \mathrm{d}t.$$

1. If h is bounded, then

$$\|f_h\|_\infty \leq \sqrt{\frac{\pi}{2}} \|h(\cdot) - \Phi(h)\|_\infty, \quad \text{and } \|f_h'\|_\infty \leq 2\|h(\cdot) - \Phi(h)\|_\infty.$$

2. If h is differentiable, then

$$\|f_h\| \leq 2\|h'\|_\infty, \quad \|f_h'\| \leq \sqrt{\frac{2}{\pi}} \|h'\|_\infty, \quad \text{and } \|f_h''\| \leq 2\|h'\|_\infty.$$

- Herein, the main focus will be on the Wasserstein metric d_W
- If $Z \sim \mathcal{N}(0,1)$ and X is a random variable, we have

$$d_K(X,Z) \leq (2/\pi)^{1/4}\sqrt{d_W(X,Z)}$$

- The class $\mathcal{H}$ used for the Wasserstein distance is the set of functions with Lipschitz constant equal to one. If $h \in \mathcal{H}$, then $\|h'\|_\infty \leq 1$ so the Item 2 in the previous statement is true.

- Herein, the main focus will be on the Wasserstein metric d_W
- If $Z \sim \mathcal{N}(0,1)$ and X is a random variable, we have

$$d_K(X,Z) \leq (2/\pi)^{1/4}\sqrt{d_W(X,Z)}$$

- The class $\mathcal{H}$ used for the Wasserstein distance is the set of functions with Lipschitz constant equal to one. If $h \in \mathcal{H}$, then $\|h'\|_\infty \leq 1$ so the Item 2 in the previous statement is true.

Theorem

If W is a random variable and Z has the standard normal distribution, and we define the family of functions $\mathcal{F} = \left\{f; \|f\|, \|f''\| \leq 2, \|f'\| \leq \sqrt{2/\pi}\right\}$, *then*

$$d_W(W,Z) \leq \sup_{f\in\mathcal{F}} \left|\mathbb{E}\left[f'(W) - Wf(W)\right]\right|.$$

Definition

The ordered pair (W, W') of random variables is called an *exchangeable pair* if $(W, W') \stackrel{d}{=} (W', W)$. if for some $0 < \lambda \leq 1$, the echangeable pair (W, W') satisfies the relation

$$\mathbb{E}[W' \mid W] = (1 - \lambda)W,$$

the we call (W, W') an λ-Stein pair.

Easy facts : Let (W, W') an exchangeable pair.

1. If $F : \mathbb{R}^2 \to \mathbb{R}$ is anti-symmetric function ; that is $F(x, y) = -F(y, x)$, then $\mathbb{E}[F(W, W')] = 0$.
2. If (W, W') is an λ-Stein pair with $\mathrm{Var}(W) = \sigma^2$, then $\mathbb{E}[W] = 0$ and $\mathbb{E}[(W - W')^2] = 2\lambda\sigma^2$

Theorem

If (W, W') is a λ-Stein pair with $\mathbb{E}W^2 = 1$ and Z has the standard normal distribution, then

$$d_W(W, Z) \leq \frac{1}{\sqrt{2\pi}\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(W' - W)^2 \mid W])} + \frac{1}{3\lambda}\mathbb{E}|W' - W|^3.$$

Theorem

If (W, W') is a λ-Stein pair with $\mathbb{E}W^2 = 1$ and Z has the standard normal distribution, then

$$d_W(W, Z) \leq \frac{1}{\sqrt{2\pi}\lambda}\sqrt{\mathrm{Var}(\mathbb{E}[(W' - W)^2 \mid W])} + \frac{1}{3\lambda}\mathbb{E}|W' - W|^3.$$

Example : Let $X_1, \ldots, X_n$ independent with $\mathbb{E}X_i^4 < \infty, \mathbb{E}X_i = 0, \mathrm{Var}(X_i) = 1$ and $W = n^{-1/2}\sum_{i=1}^n X_i$. Then

$$d_W(W, Z) \leq \sqrt{\frac{2}{\pi}}\frac{\sqrt{\sum_{i=1}^n \mathbb{E}[X_i^4]}}{2n} + \frac{2}{3n^{3/2}}\sum_{i=1}^n \mathbb{E}|X_i|^3.$$

Proof.

Let f be bounded with bounded first and second derivative and let $F(w) = \int_0^w f(t)\mathrm{d}t$

$$\begin{aligned} 0 &= \mathbb{E}[F(W') - F(W)] \\ &= \mathbb{E}\left[(W' - W)f(W) + \frac{1}{2}(W' - W)^2 f'(W) + \frac{1}{6}(W' - W)^3 f''(W^*)\right]. \end{aligned}$$

Proof.

Let f be bounded with bounded first and second derivative and let $F(w) = \int_0^w f(t)\mathrm{d}t$

$$0 = \mathbb{E}[F(W') - F(W)]$$
$$= \mathbb{E}\left[(W' - W)f(W) + \frac{1}{2}(W' - W)^2 f'(W) + \frac{1}{6}(W' - W)^3 f''(W^*)\right].$$

The condition on the Stein pair yields

$$\mathbb{E}[(W' - W)f(W)] = \mathbb{E}[\mathbb{E}(W' - W) \mid W]f(W)] = -\lambda\mathbb{E}[Wf(W)].$$

Proof.

Let f be bounded with bounded first and second derivative and let $F(w) = \int_0^w f(t)\mathrm{d}t$

$$0 = \mathbb{E}[F(W') - F(W)]$$
$$= \mathbb{E}\left[(W' - W)f(W) + \frac{1}{2}(W' - W)^2 f'(W) + \frac{1}{6}(W' - W)^3 f''(W^*)\right].$$

The condition on the Stein pair yields

$$\mathbb{E}[(W' - W)f(W)] = \mathbb{E}[\mathbb{E}(W' - W) \mid W]f(W)] = -\lambda\mathbb{E}[Wf(W)].$$

Gathering facts

$$\mathbb{E}[Wf(W)] = \mathbb{E}\left[\frac{1}{2\lambda}(W' - W)^2 f'(W) + \frac{1}{6\lambda}(W' - W)^3 f''(W^*)\right].$$

□

Definition

For a random variable $X \geq 0$ with $\mathbb{E}[X] = \mu < \infty$, we say that the random variable X^s has the size-bias distribution with respect to X if for all f such that $\mathbb{E}[|Xf(X)|] < \infty$ we have

$$\mathbb{E}[Xf(X)] = \mu \mathbb{E}[f(X^s)].$$

Definition

For a random variable $X \geq 0$ with $\mathbb{E}[X] = \mu < \infty$, we say that the random variable X^s has the size-bias distribution with respect to X if for all f such that $\mathbb{E}[|Xf(X)|] < \infty$ we have

$$\mathbb{E}[Xf(X)] = \mu\mathbb{E}[f(X^s)].$$

Fact : If $X \geq 0$ is a random variable with $\mathbb{E}[X] = \mu < \infty$ and distribution function F, then the size-bias distribution of X is absolutely continuous with respect to the measure of X with density read form

$$dF^s(x) = \frac{x}{\mu}dF(x).$$

Theorem

Let $X \geq 0$ *be a random variable with* $\mathbb{E}[X] = \mu < \infty$ *and* $\mathrm{Var}(X) = \sigma^2$*. Let* X^s *be defined on the same space as* X *and have the size-bias distribution with respect to* X*. If* $W = (X - \mu)/\sigma$ *and* $Z \sim \mathcal{N}(0,1)$*, then*

$$d_W(W, Z) \leq \frac{\mu}{\sigma^2}\sqrt{\frac{2}{\pi}}\sqrt{\mathrm{Var}(\mathbb{E}[X^s - X \mid X])} + \frac{\mu}{\sigma^3}\mathbb{E}[(X^s - X)^2].$$

Proof.

Taylor expansion Yields

$$\mathbb{E}[Wf(W)] = \frac{\mu}{\sigma}\mathbb{E}\left[\frac{X^s - X}{\sigma} f'\left(\frac{X-\mu}{\sigma}\right) + \frac{(X^s - X)^2}{2\sigma^2} f''\left(\frac{X^* - \mu}{\sigma}\right)\right].$$

We obtain

$$|\mathbb{E}[f'(W) - Wf(W)]| \leq \left|\mathbb{E}\left[f'(W)\left(1 - \frac{\mu}{\sigma^2}(X^s - X)\right)\right]\right| + \frac{\mu}{2\sigma^3}\left|\mathbb{E}\left[f''\left(\frac{f^* - \mu}{\sigma}\right)(X^s - X)^2\right]\right|. \quad (3)$$

□

Coupling Construction

We have the following recipe to construct a size-bias version of X in the case that $X = \sum_{i=1}^{n} X_i$, where $X_i \geq 0$ and $\mathbb{E}[X_i] = \mu_i$:

Coupling Construction

We have the following recipe to construct a size-bias version of X in the case that $X = \sum_{i=1}^{n} X_i$, where $X_i \geq 0$ and $\mathbb{E}[X_i] = \mu_i$:

1. For each $i = 1, \ldots, n$, let X_i^s have the size-bias distribution of X_i independent of $(X_j)_{j\neq i}$ and $(X_j^s)_{j\neq i}$.Given $X_i^s = x$, define the vector $(X_j^{(i)})_{j\neq i}$ to have the distribution of $(X_j)_{j\neq i}$ conditional on $X_i = x$.

Coupling Construction

We have the following recipe to construct a size-bias version of X in the case that $X = \sum_{i=1}^{n} X_i$, where $X_i \geq 0$ and $\mathbb{E}[X_i] = \mu_i$:

1. For each $i = 1, \ldots, n$, let X_i^s have the size-bias distribution of X_i independent of $(X_j)_{j \neq i}$ and $(X_j^s)_{j \neq i}$.Given $X_i^s = x$, define the vector $(X_j^{(i)})_{j \neq i}$ to have the distribution of $(X_j)_{j \neq i}$ conditional on $X_i = x$.
2. Choose a random summand X_I, where the index I has $\mathbb{P}(I = i) = \mu_i / \mathbb{E}X$. and independent of all else.

Coupling Construction

We have the following recipe to construct a size-bias version of X in the case that $X = \sum_{i=1}^{n} X_i$, where $X_i \geq 0$ and $\mathbb{E}[X_i] = \mu_i$:

1. For each $i = 1, \ldots, n$, let X_i^s have the size-bias distribution of X_i independent of $(X_j)_{j \neq i}$ and $(X_j^s)_{j \neq i}$.Given $X_i^s = x$, define the vector $(X_j^{(i)})_{j \neq i}$ to have the distribution of $(X_j)_{j \neq i}$ conditional on $X_i = x$.
2. Choose a random summand X_I, where the index I has $\mathbb{P}(I = i) = \mu_i / \mathbb{E}X$. and independent of all else.
3. Define $X^s = \sum_{j \neq I} X_j^{(I)} + X_I^s$.

Proposition : Let $X = \sum_{i=1}^{n} X_i$, with $X_i \geq 0$ and $\mathbb{E}[X_i] = \mu_i$ and also $\mu = \sum_i \mu_i$. If X^s is constructed by Items 1-3, then X^s has the size-bias distribution of X.

Corollary

Let $X_1, \ldots X_n$ be non-negative independent random variables with $\mathbb{E}[X_i] = \mu_i$, and for each $i = 1, \ldots, n$, let X_i^s have the size-bias distribution of X_i independent of $(X_j)_{j \neq i}$ and $(X_j^s)_{j \neq i}$. If $X = \sum_{i=1}^n X_i, \mathbb{E}[X] = \mu$, and I independent of all else with $\mathbb{P}(I = i) = \mu_i/\mu$, then $X^s = X - X_I + X_I^s$ has the size-bias distribution of X.

Corollary

Let $X_1, \dots X_n$ be non-negative independent random variables with $\mathbb{E}[X_i] = \mu_i$, and for each $i = 1, \dots, n$, let X_i^s have the size-bias distribution of X_i independent of $(X_j)_{j \neq i}$ and $(X_j^s)_{j \neq i}$. If $X = \sum_{i=1}^n X_i, \mathbb{E}[X] = \mu$, and I independent of all else with $\mathbb{P}(I = i) = \mu_i / \mu$, then $X^s = X - X_I + X_I^s$ has the size-bias distribution of X.

Exercise : Let us bound the Wasserstein distance between the normalized sum of independent variables with finite third moment and the normal distribution.

Definition

For a centred random variable W with variance σ^2, we say that the random variable W^z has the zero-bias distribution with respect to W if for all f such that $\mathbb{E}[|Wf(W)|] < \infty$ we have

$$\mathbb{E}[Wf(W)] = \sigma^2 \mathbb{E}[f(W^z)].$$

Theorem

Let W be a mean zero, variance one random variable and let W^z be defined on the same space as W and have the zero-bias distribution with respect to W. If $Z \sim \mathcal{N}(0,1)$, then

$$d_W(W,Z) \leq 2\mathbb{E}[|W^z - W|].$$

Coupling construction

Let $X_1, \ldots, X_n$ independent random variables having zero mean and such that, $\mathrm{Var}(X_i) = \sigma_i^2$, $\sum_{i=1}^n \sigma_i^2 = 1$, and define $W = \sum_{i=1}^n X_i$.

Coupling construction

Let $X_1, \ldots, X_n$ independent random variables having zero mean and such that, $\mathrm{Var}(X_i) = \sigma_i^2$, $\sum_{i=1}^n \sigma_i^2 = 1$, and define $W = \sum_{i=1}^n X_i$.

1. For each $i = 1, \ldots, n$, let X_i^Z have the zero-bias distribution of X_i independent of $(X_j)_{j \neq i}$ and $(X_j^z)_{j \neq i}$.

Coupling construction

Let $X_1, \ldots, X_n$ independent random variables having zero mean and such that, $\mathrm{Var}(X_i) = \sigma_i^2$, $\sum_{i=1}^n \sigma_i^2 = 1$, and define $W = \sum_{i=1}^n X_i$.

1. For each $i = 1, \ldots, n$, let X_i^Z have the zero-bias distribution of X_i independent of $(X_j)_{j \neq i}$ and $(X_j^z)_{j \neq i}$.
2. Choose a random summand X_I, where the index I has $\mathbb{P}(I = i) = \sigma_i^2$. and independent of all else.

Coupling construction

Let $X_1, \ldots, X_n$ independent random variables having zero mean and such that, $\mathrm{Var}(X_i) = \sigma_i^2$, $\sum_{i=1}^n \sigma_i^2 = 1$, and define $W = \sum_{i=1}^n X_i$.

1. For each $i = 1, \ldots, n$, let X_i^Z have the zero-bias distribution of X_i independent of $(X_j)_{j \neq i}$ and $(X_j^z)_{j \neq i}$.
2. Choose a random summand X_I, where the index I has $\mathbb{P}(I = i) = \sigma_i^2$. and independent of all else.
3. Define $W^z = \sum_{j \neq I} X_i + X_I^z$.

- For multivariate normal approximation : Reinert, G. and Röllin, A. (2009). Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. Ann. Probab., 37(6) :2150– 2173.

- For multivariate normal approximation : Reinert, G. and Röllin, A. (2009). Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. Ann. Probab., 37(6) :2150– 2173.
- For poisson approximation : A. D. Barbour, L. Holst, and S. Janson. Poisson approximation, volume 2 of Oxford Studies in Probability. The Clarendon Press Oxford Uni- versity Press, New York, 1992. Oxford Science Publications.

- For multivariate normal approximation : Reinert, G. and Röllin, A. (2009). Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. Ann. Probab., 37(6) :2150– 2173.
- For poisson approximation : A. D. Barbour, L. Holst, and S. Janson. Poisson approximation, volume 2 of Oxford Studies in Probability. The Clarendon Press Oxford Uni- versity Press, New York, 1992. Oxford Science Publications.
- For exponential approximation : E. A. Peköz and A. Röllin. New rates for exponential approximation and the theorems of Rényi and Yaglom. Ann. Probab., 39(2) :587–608, 2011.

- For multivariate normal approximation : Reinert, G. and Röllin, A. (2009). Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. Ann. Probab., 37(6) :2150– 2173.
- For poisson approximation : A. D. Barbour, L. Holst, and S. Janson. Poisson approximation, volume 2 of Oxford Studies in Probability. The Clarendon Press Oxford Uni- versity Press, New York, 1992. Oxford Science Publications.
- For exponential approximation : E. A. Peköz and A. Röllin. New rates for exponential approximation and the theorems of Rényi and Yaglom. Ann. Probab., 39(2) :587–608, 2011.
- For geometric approximation : E. Peköz, A. Röllin, and N. Ross. Total variation and local limit error bounds for geometric approximation. Bernoulli, 2010.

- for concentration inequality : S. Ghosh and L. Goldstein. Concentration of measures via size- biased couplings. Probability Theory and Related Fields, 149 :271–278, 2011. 10.1007/s00440-009-0253-3.
- S. Chatterjee, "Stein's method for concentration inequalities," Probab. Theory Related Fields, vol. 138, pp. 305–321, 2007

Thank You