Stein's method for Conditional CLT

Grigory Terlov
(joint with Partha S. Dey)

University of Illinois at Urbana-Champaign

Frontier Probability Days, Dec 3-5th, 2021

Central Limit Theorem

Theorem (CLT)

Suppose $\omega_1, \omega_2, \ldots$ *are* *i.i.d.* *random variables with* $\mathbb{E}\omega_i = 0$ *and* $\mathbb{E}\omega_i^2 = 1$, *then*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \omega_i \Rightarrow Z,$$

where Z *is a standard normal random variable.*

Central Limit Theorem

Theorem (CLT)

Suppose $\omega_1, \omega_2, \ldots$ *are i.i.d. random variables with* $\mathbb{E}\omega_i = 0$ *and* $\mathbb{E}\omega_i^2 = 1$, *then*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \omega_i \Rightarrow Z,$$

where Z *is a standard normal random variable.*

Theorem (Berry-Esseen)

Suppose $\omega_1, \omega_2, \ldots$ *are i.i.d. random variables with* $\mathbb{E}\omega_i = 0$, $\mathbb{E}\omega_i^2 = 1$, *and* $\mathbb{E}|\omega_i|^3 = \gamma \in (1, \infty)$ *then for all* n

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \omega_i \leq x \right) - \mathbb{P}(Z \leq x) \right| \leq \frac{C\gamma}{\sqrt{n}}.$$

Conditional CLTs

Often in probability theory we study sums of random variables conditioned on another such sum.

Conditional CLTs

Often in probability theory we study sums of random variables conditioned on another such sum.

- Conditionally independent random variables [Bulinskii'16, Yang – Wei'14 among others],

Conditional CLTs

Often in probability theory we study sums of random variables conditioned on another such sum.

- Conditionally independent random variables [Bulinskii'16, Yang – Wei'14 among others],
- Under assumption of sufficient statistics [Holst'79],

Conditional CLTs

Often in probability theory we study sums of random variables conditioned on another such sum.

- Conditionally independent random variables [Bulinskii'16, Yang – Wei'14 among others],
- Under assumption of sufficient statistics [Holst'79],
- Stationary sequence $\sum_{i=1}^{n}(X_0 \circ T^i)$, $n \geq 1$ conditioned on $\mathcal{M}_i := T^{-i}(\mathcal{M}_0)$. [Dedecker – Merlevède '02]

Conditional CLTs with explicit rates of convergence

Also known as semi-local Berry-Esseen theorems.

- A positively recurrent Markov chain with finite absolute third moment conditioned on the time of the n^{th} return to 0. [Bolthausen'80],

Conditional CLTs with explicit rates of convergence

Also known as semi-local Berry-Esseen theorems.

- A positively recurrent Markov chain with finite absolute third moment conditioned on the time of the n^{th} return to 0. [Bolthausen'80],
- $\sum_{i=1}^{n}(X_i, Y_i)$ of i.i.d. random vectors with $\mathbb{E}(|X_1| + |Y_1|)^{2+\delta} < \infty$ conditioned on $\sum Y_i = k$. [Guo – Peterson'18]

Stein's method

Lemma (Stein'72)

A random variable W has standard normal distribution if and only if for every piecewise continuously differentiable function $f : \mathbb{R} \to \mathbb{R}$ with $\mathbb{E}|f'(Z)| < \infty$, $Z \sim N(0,1)$, we have

$$\mathbb{E}\left(Wf(W) - f'(W)\right) = 0.$$

Stein's method

Lemma (Stein'72)

A random variable W has standard normal distribution if and only if for every piecewise continuously differentiable function $f : \mathbb{R} \to \mathbb{R}$ with $\mathbb{E}|f'(Z)| < \infty$, $Z \sim N(0,1)$, we have

$$\mathbb{E}\left(Wf(W) - f'(W)\right) = 0.$$

$$d_{\mathcal{W}}(W, Z) = \sup_{g:1-\text{Lip.}} |\mathbb{E}g(W) - \mathbb{E}g(Z)|$$

Stein's method

Lemma (Stein'72)

A random variable W has standard normal distribution if and only if for every piecewise continuously differentiable function $f : \mathbb{R} \to \mathbb{R}$ with $\mathbb{E}|f'(Z)| < \infty$, $Z \sim N(0,1)$, we have

$$\mathbb{E}\left(Wf(W) - f'(W)\right) = 0.$$

$$d_{\mathcal{W}}(W, Z) = \sup_{g:1-\text{Lip.}} |\mathbb{E}g(W) - \mathbb{E}g(Z)| \leq \sup_{f \in \mathcal{A}} \left|\mathbb{E}Wf(W) - f'(W)\right|,$$

Stein's method

Lemma (Stein'72)

A random variable W has standard normal distribution if and only if for every piecewise continuously differentiable function $f : \mathbb{R} \to \mathbb{R}$ with $\mathbb{E}|f'(Z)| < \infty$, $Z \sim N(0,1)$, we have

$$\mathbb{E}\left(Wf(W) - f'(W)\right) = 0.$$

$$d_{\mathcal{W}}(W,Z) = \sup_{g:1-\text{Lip.}} |\mathbb{E}g(W) - \mathbb{E}g(Z)| \le \sup_{f\in\mathcal{A}} \left|\mathbb{E}Wf(W) - f'(W)\right|,$$

where $\mathcal{A} := \{f : ||f||_\infty < 1, ||f'||_\infty \le \sqrt{\frac{2}{\pi}}, \text{ and } ||f''||_\infty < 2\}$.

Exchangeable pairs

Definition

Random variables W and W' are called exchangeable if (W, W') and (W', W) follow the same distribution.

Exchangeable pairs

Definition

Random variables W and W' are called exchangeable if (W, W') and (W', W) follow the same distribution.

Theorem (Stein'72)

Let (W, W') be an exchangeable pair of random variables and $\Delta W := W' - W$. Suppose $\mathbb{E}W = 0$, $\mathbb{E}W^2 = 1$ and $\mathbb{E}(\Delta W|W) = -\lambda(W + R_1)$ and $\mathbb{E}\left(\Delta W^2 \mid W\right) = 2\lambda(1 + R_2)$ for some $\lambda \in (0, 1)$.

Exchangeable pairs

Definition

Random variables W and W' are called exchangeable if (W, W') and (W', W) follow the same distribution.

Theorem (Stein'72)

Let (W, W') be an exchangeable pair of random variables and $\Delta W := W' - W$. Suppose $\mathbb{E}W = 0$, $\mathbb{E}W^2 = 1$ and $\mathbb{E}(\Delta W|W) = -\lambda(W + R_1)$ and $\mathbb{E}\left(\Delta W^2 \mid W\right) = 2\lambda(1 + R_2)$ for some $\lambda \in (0,1)$. Then we have

$$d_{\mathcal{W}}\left(W, Z\right) \le \mathbb{E}|R_1| + \sqrt{\frac{2}{\pi}}\mathbb{E}|R_2| + \frac{1}{3\lambda}\mathbb{E}|\Delta W|^3,$$

where $Z \sim N(0,1)$ and $d_{\mathcal{W}}$ denotes Wasserstein distance.

#01 in a binary sequence

Example

Let X_n be the number of times #01 appears in the random binary sequence $(\omega_1, \omega_2, \ldots, \omega_n, \omega_1)$, where ω_i are i.i.d. Bernoulli$\left(\frac{1}{2}\right)$. and define $W_n := \frac{1}{\sigma_{X_n}}(X_n - \mathbb{E}X_n)$.

#01 in a binary sequence

Example

Let X_n be the number of times #01 appears in the random binary sequence $(\omega_1, \omega_2, \ldots, \omega_n, \omega_1)$, where ω_i are i.i.d. Bernoulli$\left(\frac{1}{2}\right)$. and define $W_n := \frac{1}{\sigma_{X_n}}(X_n - \mathbb{E}X_n)$.

It is easy to check that

$$\mathbb{E}(\Delta W \mid W) = -\frac{2}{n}W \text{ and } \mathbb{E}\left((\Delta W)^2 \mid W\right) = 2\frac{2}{n}(1 + R_2),$$

where $\mathbb{E}|R_2| \leq O(\sqrt{n})$. Moreover $n\mathbb{E}|\Delta W|^3 \leq O(1/\sqrt{n})$, thus by the theorem above

$$d_{\mathcal{W}}(W_n, Z) \leq \frac{C}{\sqrt{n}}.$$

Conditional version of the Stein's method

Consider is a random variable $(W \mid Y = k)$ where $|k - \mathbb{E}Y| \ll \sigma_Y$.

Conditional version of the Stein's method

Consider is a random variable $(W \mid Y = k)$ where $|k - \mathbb{E}Y| \ll \sigma_Y$.

- If $\mathbb{P}(W' \neq W \mid Y' = Y) > 0$ then classical technique is applicable.

Conditional version of the Stein's method

Consider is a random variable $(W \mid Y = k)$ where $|k - \mathbb{E}Y| \ll \sigma_Y$.

- If $\mathbb{P}(W' \neq W \mid Y' = Y) > 0$ then classical technique is applicable.
- If $\mathbb{P}(W' \neq W \mid Y' = Y) = 0$ a new approach is needed.

Assumptions

Assume that

I.1 Random variable W has mean 0 and variance to be 1,

I.2 Y is a mean 0 random variable with variance σ_Y^2,

I.3 W and Y are uncorrelated,

I.4 The random vectors (W, Y) and (W', Y') are exchangeable.

Assumptions cont.

Further assume that

II Y takes values in $\zeta + \mathbb{Z}$ for some $\zeta \in [0,1)$, $\Delta Y \in \{-1,0,1\}$ and

$$\mathbb{P}(\Delta Y = \pm 1 \mid W, Y) = \lambda \sigma_Y^2 + R_{0,\pm}.$$

III

$$\mathbb{E}(\Delta W \mathbb{1}_{\Delta Y = \pm 1} \mid W, Y) = -\lambda \left(\frac{1}{2} \psi W + R_{1,\pm} \right).$$

IV

$$\mathbb{E}\left((\Delta W)^2 \mathbb{1}_{\Delta Y = \pm 1} \mid W, Y \right) = \lambda \left(\psi + R_{2,\pm} \right).$$

Conditional version of the Stein's method

Theorem (Dey, T. 2021+)

Suppose W and Y satisfy assumptions I-IV and $\mathbb{E}|W|^3 < \infty$. Let k be such that $\mathbb{P}(Y = k) > 0$ and $|k| \ll \sigma_Y$. Then

$$d_{\mathcal{W}}((W \mid Y = k), Z) \leq \frac{2}{\psi} A_k + \sqrt{\frac{2}{\pi\psi^2}} B_k + \frac{1}{\lambda\sigma_Y^2} C_k + \frac{1}{\lambda\sigma_Y^2} D_k + \frac{2}{3\lambda\psi} E_k + \frac{ck}{\sigma_Y^2},$$

where

$$A_k = \mathbb{E}\left(|R_{1,-}| \,\middle|\, Y = k\right) + \mathbb{E}\left(|R_{1,+}| \,\middle|\, Y = k-1\right),$$
$$B_k = \mathbb{E}\left(|R_{2,-}| \,\middle|\, Y = k\right) + \mathbb{E}\left(|R_{2,+}| \,\middle|\, Y = k-1\right),$$
$$C_k = \mathbb{E}\left(|W|(|R_{0,+}| + |R_{0,-}|) \,\middle|\, Y \in \{k-1, k\}\right),$$
$$D_k = \mathbb{E}\left(|\Delta W| \,\middle|\, Y \in \{k-1, k\}\right),$$
$$E_k = \mathbb{E}(|\Delta W|^3 \mid Y \in \{k-1, k\}).$$

Conditional version of the Stein's method

Usually in applications after making X and Y uncorrelated those linearity conditions would look as follows:

$$\mathbb{E}\left(\Delta X \mathbb{1}_{\Delta Y = \pm 1} \mid X, Y\right) = -\lambda(a_{\pm} X + b_{\pm} Y + R_{1,\pm})$$

Conditional version of the Stein's method

Usually in applications after making X and Y uncorrelated those linearity conditions would look as follows:

$$\mathbb{E}\left(\Delta X \mathbb{1}_{\Delta Y=\pm 1} \mid X, Y\right) = -\lambda(a_{\pm} X + b_{\pm} Y + R_{1,\pm})$$

Proposition (Univariate change of variable)

Suppose X and Y are random variables satisfying Assumptions I, II, III' and IV. Assume that $R_{0,\pm} = \mp\lambda a_{\pm} Y$ and define the change of variable

$$W^0 := X + \lambda\,\psi\,\alpha XY + \frac{\lambda\,\theta}{2}\left(Y^2 - \mathbb{E}Y^2\right) + \frac{\lambda^2(\psi+1)\alpha\theta}{3}Y^3,$$

where $\alpha := (a_+ - a_-)/\lambda\sigma_Y^2$ and $\theta := b_+/\lambda\sigma_Y^2$. Let $W = W^0/\sigma_{W^0}$. Then (W, Y) satisfies Assumptions I–IV and $\widetilde{R}_{i,\pm} \approx R_{i,\pm}$ for $i = \{1, 2\}$.

Back to our example

Example

Let V_n be the number of $\#1$'s in the random binary sequence $(\omega_1, \omega_2, \ldots, \omega_n, \omega_1)$ and define $Y_n := V_n - \mathbb{E}V_n$ and W_n be the modified number of $\#01$;

Then by the theorem above we get that

$$d_{\mathcal{W}}\left(\left(\frac{W}{\sigma_W}\middle|Y = k\right), Z\right) \leq \frac{C}{n^{1/2-\varepsilon}},$$

for some constants C and $\varepsilon > 0$.

Other applications

- In homogeneous binary sequence: $\omega_i \sim \mathrm{Bernoulli(p)}$ if i is even and $\mathrm{Bernoulli(1\text{-}p)}$ if i is odd.

Other applications

- In homogeneous binary sequence: $\omega_i \sim \mathrm{Bernoulli(p)}$ if i is even and $\mathrm{Bernoulli(1\text{-}p)}$ if i is odd. Let $X = \sum_{i=1}^{n} (-1)^n \omega_i \omega_{i+1}$, $W = X/\sigma_x$,

Other applications

- In homogeneous binary sequence: $\omega_i \sim \mathrm{Bernoulli(p)}$ if i is even and $\mathrm{Bernoulli(1\text{-}p)}$ if i is odd. Let $X = \sum_{i=1}^{n} (-1)^n \omega_i \omega_{i+1}$, $W = X/\sigma_x$, then
$$d_{\mathcal{W}}\left((W \mid Y = k), Z\right) \leq C \cdot n^{-1/2+\varepsilon}.$$

Other applications

- In homogeneous binary sequence: $\omega_i \sim \mathrm{Bernoulli(p)}$ if i is even and $\mathrm{Bernoulli(1\text{-}p)}$ if i is odd. Let $X = \sum_{i=1}^{n} (-1)^n \omega_i \omega_{i+1}$, $W = X/\sigma_x$, then
$$d_{\mathcal{W}}\left((W \mid Y = k), Z\right) \leq C \cdot n^{-1/2+\varepsilon}.$$

Subgraph counts in Erdös-Rényi random graph:

- number of wedges given the number of edges with rate $n^{-1/2+\varepsilon}$,

Other applications

- In homogeneous binary sequence: $\omega_i \sim \mathrm{Bernoulli(p)}$ if i is even and $\mathrm{Bernoulli(1\text{-}p)}$ if i is odd. Let $X = \sum_{i=1}^{n} (-1)^n \omega_i \omega_{i+1}$, $W = X/\sigma_x$, then
$$d_{\mathcal{W}}\left((W \mid Y = k), Z\right) \leq C \cdot n^{-1/2+\varepsilon}.$$

Subgraph counts in Erdös-Rényi random graph:

- number of wedges given the number of edges with rate $n^{-1/2+\varepsilon}$,
- (triangle, wedge) counts given the number of edges with rate $n^{-1/2+\varepsilon}$,

Other applications

- In homogeneous binary sequence: $\omega_i \sim \mathrm{Bernoulli(p)}$ if i is even and $\mathrm{Bernoulli(1\text{-}p)}$ if i is odd. Let $X = \sum_{i=1}^{n}(-1)^n \omega_i \omega_{i+1}$, $W = X/\sigma_x$, then
$$d_{\mathcal{W}}\left((W \mid Y = k), Z\right) \le C \cdot n^{-1/2+\varepsilon}.$$

Subgraph counts in Erdös-Rényi random graph:

- number of wedges given the number of edges with rate $n^{-1/2+\varepsilon}$,
- (triangle, wedge) counts given the number of edges with rate $n^{-1/2+\varepsilon}$,
- (general subgraph, triangle, wedge) counts given the number of edges with rate $n^{-1/2+\varepsilon}$.

Thank You!