Central limit theorem: variants and applications

Anindya De

University of Pennsylvania

Introduction

One of the most cornerstone results in probability theory and statistics.

Introduction

One of the most cornerstone results in probability theory and statistics.

Informally, sum of independent random variables converges to a Gaussian.

Introduction

One of the most cornerstone results in probability theory and statistics.

Informally, sum of independent random variables converges to a Gaussian.

Applications in computer science

Many extensions discovered in the context of algorithmic problems.

1. Invariance principle [Mossel-O'Donnell-Oleskiewicz] – numerous applications in hardness of approximation, derandomization and social choice.

Applications in computer science

Many extensions discovered in the context of algorithmic problems.

1. Invariance principle [Mossel-O'Donnell-Oleskiewicz] – numerous applications in hardness of approximation, derandomization and social choice.

2. Multidimensional central limit theorems [Daskalakis-Papadimtriou, Daskalakis-Kamath-Tzamos, Valiant-Valiant] – many extensions and applications in algorithmic game theory and lower bounds in statistics.

Applications in computer science

Many extensions discovered in the context of algorithmic problems.

3. Moment matching theorems [P. Valiant, Gopalan-Klivans-Meka] –lower bounds in statistics, learning theory.

Applications in computer science

Many extensions discovered in the context of algorithmic problems.

3. Moment matching theorems [P. Valiant, Gopalan-Klivans-Meka] –lower bounds in statistics, learning theory.

4. Central limit theorems for low-degree polynomials / polytopes [Harsha-Klivans-Meka, De-Servedio] – Derandomization.

Applications in computer science

Many extensions discovered in the context of algorithmic problems.

3. Moment matching theorems [P. Valiant, Gopalan-Klivans-Meka] –lower bounds in statistics, learning theory.
4. Central limit theorems for low-degree polynomials / polytopes [Harsha-Klivans-Meka, De-Servedio] – Derandomization.
5. Discrete central limit theorems [Chen-Goldstein-Shao] – computational learning theory.

Why is central limit theorem useful?

Central limit theorem: Even if $X_1, \ldots, X_n$ are unwieldy random variables, their sum $X_1 + \ldots + X_n$ is *nice*.

Why is central limit theorem useful?

Central limit theorem: Even if $X_1, \dots, X_n$ are unwieldy random variables, their sum $X_1 + \dots + X_n$ is *nice*.

In some applications, the precise convergence to the Gaussian distribution is important.

Why is central limit theorem useful?

Central limit theorem: Even if $X_1, \ldots, X_n$ are unwieldy random variables, their sum $X_1 + \ldots + X_n$ is *nice*.

In some applications, the precise convergence to the Gaussian distribution is important.

In others, the fact that a Gaussian can be parameterized by two parameters is sufficient.

Berry-Esséen theorem

Theorem

Let $X_1, \dots, X_n$ be n independent centered random variables such that $Var(X_i) = \sigma_i^2$ and $\mathbf{E}[|X_i|^3] = \beta_{3,i}$. Define $S = \sum X_i$, $\sigma^2 = Var(S)$ and $\beta_3 = \sum \beta_{3,i}$. Then,

$$d_K(S, \mathcal{N}(0, \sigma^2)) = O(1) \cdot \frac{\beta_3}{\sigma^3}.$$

$$d_K(X, Y) = \sup_{z \in \mathbb{R}} \Big| \Pr[X \leq z] - \Pr[Y \leq z] \Big|.$$

Corollary of the Berry-Esséen theorem

Corollary

Let $X_1, \ldots, X_n$ be n independent identical centered random variables such that $Var(X_i) = \sigma_*^2$ and $\mathbf{E}[|X_i|^3] = \beta_{3,*}$ (for all $1 \leq i \leq n$). Define $S = \sum_i X_i$. Then,

$$d_K(S, \mathcal{N}(0, n\sigma_*^2)) = O\left(\frac{1}{\sqrt{n}}\right) \cdot \frac{\beta_{3,*}}{\sigma_*^3}.$$

Corollary of the Berry-Esséen theorem

Let us assume that the random variable X_i is *hypercontractive* – in other words, there is a constant $C > 0$ such that

$$\mathbf{E}[|X_i|^3] \leq C \cdot \left(\mathbf{E}[|X_i|^2]\right)^{3/2}.$$

Corollary of the Berry-Esséen theorem

Let us assume that the random variable X_i is *hypercontractive* – in other words, there is a constant $C > 0$ such that

$$\mathbf{E}[|X_i|^3] \leq C \cdot \left(\mathbf{E}[|X_i|^2]\right)^{3/2}.$$

This implies that $\beta_{3,*}/\sigma_*^3 \leq C$. Thus, the error term in Berry-Esséen theorem is now

$$d_K(S, \mathcal{N}(0, n\sigma_*^2)) = O\left(\frac{C}{\sqrt{n}}\right).$$

Berry–Esséen for non identical random variables

Continue to assume that $X_1, \ldots, X_n$ are all C-hypercontractive.

Suppose $\max_i \sigma_i^2 \leq \epsilon^2 \cdot (\sum_{j=1}^n \sigma_j^2)$.

Then, the error term in Berry–Esséen becomes

$$\frac{\beta_3}{\sigma^3} \leq C \cdot \frac{\sum_j \sigma_j^3}{(\sum_j \sigma_j^2)^{1.5}} \leq C \cdot \epsilon.$$

Berry–Esséen for non identical random variables

Continue to assume that $X_1, \ldots, X_n$ are all C-hypercontractive.

Suppose $\max_i \sigma_i^2 \leq \epsilon^2 \cdot (\sum_{j=1}^n \sigma_j^2)$.

Then, the error term in Berry–Esséen becomes

$$\frac{\beta_3}{\sigma^3} \leq C \cdot \frac{\sum_j \sigma_j^3}{(\sum_j \sigma_j^2)^{1.5}} \leq C \cdot \epsilon.$$

Thus, as long as none of the individual variances are too large, the sum $\sum X_i$ converges to a Gaussian.

How do you prove Berry–Esséen?

How do you prove Berry–Esséen?

There are many known techniques used to prove "central limit theorems".

How do you prove Berry–Esséen?

There are many known techniques used to prove "central limit theorems".

1. Lindeberg exchange method (hybrid method) – used by MOO in their proof of the invariance principle.

2. Stein's method – based on constructing an operator of which the Gaussian is a fixed point.

3. Characteristic functions – aka Fourier analysis, the original method of Esséen.

We will only prove (at a high level) this for i.i.d. random variables.

Assume that $X_1, \ldots, X_n$ are i.i.d. with common distribution $\mathbf{X}$. Further, $\mathbf{E}[\mathbf{X}] = 0$, $\mathbf{E}[\mathbf{X}^2] = 1$ and $\mathbf{E}[\mathbf{X}^4] \leq 10$. In fact, for simplicity, assume that all the moments of $\mathbf{X}$ exist.

Goal: Show that $\mathbf{S} = \frac{X_1 + \ldots + X_n}{\sqrt{n}}$ satisfies

$$d_K(S, \mathcal{N}(0,1)) = O\left(\frac{1}{\sqrt{n}}\right).$$

Characteristic functions

For any $\xi \in \mathbb{R}$ and real-valued random variable $\mathbf{W}$, we define

$$\widehat{\mathbf{W}}(\xi) = \mathbf{E}[e^{i\xi\mathbf{W}}].$$

Characteristic functions

For any $\xi \in \mathbb{R}$ and real-valued random variable $\mathbf{W}$, we define

$$\widehat{\mathbf{W}}(\xi) = \mathbf{E}[e^{i\xi\mathbf{W}}].$$

Observe that $\widehat{\mathbf{W}}(0) = 1$ for any $\mathbf{W}$. When $X_1, \ldots, X_n$ are independent, then

$$\widehat{\mathbf{S}}(\xi) = \prod_{i=1}^{n} \widehat{X}_i(\xi/\sqrt{n}) = (\widehat{\mathbf{X}}(\xi/\sqrt{n}))^n.$$

Characteristic functions

For any $\xi \in \mathbb{R}$ and real-valued random variable $\mathbf{W}$, we define

$$\widehat{\mathbf{W}}(\xi) = \mathbf{E}[e^{i\xi\mathbf{W}}].$$

Observe that $\widehat{\mathbf{W}}(0) = 1$ for any $\mathbf{W}$. When $X_1, \ldots, X_n$ are independent, then

$$\widehat{\mathbf{S}}(\xi) = \prod_{i=1}^{n} \widehat{X}_i(\xi/\sqrt{n}) = (\widehat{\mathbf{X}}(\xi/\sqrt{n}))^n.$$

Characteristic functions are nothing but the Fourier transform of random variables.

High level proof idea of Berry-Esséen theorem

1. Let $\mathbf{Z} = \mathcal{N}(0, 1)$.
2. Goal: Show that $\mathbf{S}$ is close to $\mathbf{Z}$ is Kolmogorov distance.

High level proof idea of Berry-Esséen theorem

1. Let $\mathbf{Z} = \mathcal{N}(0,1)$.
2. Goal: Show that $\mathbf{S}$ is close to $\mathbf{Z}$ is Kolmogorov distance.
3. First show that $\widehat{\mathbf{S}}(\xi)$ is close to $\widehat{\mathbf{Z}}(\xi)$ where $\mathbf{Z} = \mathcal{N}(0,1)$.

High level proof idea of Berry-Esséen theorem

1. Let $\mathbf{Z} = \mathcal{N}(0, 1)$.
2. Goal: Show that $\mathbf{S}$ is close to $\mathbf{Z}$ is Kolmogorov distance.
3. First show that $\widehat{\mathbf{S}}(\xi)$ is close to $\widehat{\mathbf{Z}}(\xi)$ where $\mathbf{Z} = \mathcal{N}(0, 1)$.
4. Perform an approximate Fourier inversion to show that $\mathbf{S}$ is close to $\mathbf{Z}$.

Approximate Fourier inversion

What is meant by approximate Fourier inversion?

Approximate Fourier inversion

What is meant by approximate Fourier inversion?

Exact Fourier inversion:

$$\Pr[\mathbf{S} \leq x] - \Pr[\mathbf{Z} \leq x] = \lim_{T \to \infty} \frac{1}{2\pi} \int_{\xi=-T}^{\xi=T} e^{-i\xi x} \frac{\widehat{\mathbf{S}}(\xi) - \widehat{\mathbf{Z}}(\xi)}{i\xi} d\xi$$

Approximate Fourier inversion

What is meant by approximate Fourier inversion?

Exact Fourier inversion:

$$\Pr[\mathbf{S} \le x] - \Pr[\mathbf{Z} \le x] = \lim_{T \to \infty} \frac{1}{2\pi} \int_{\xi=-T}^{\xi=T} e^{-i\xi x} \frac{\widehat{\mathbf{S}}(\xi) - \widehat{\mathbf{Z}}(\xi)}{i\xi} d\xi$$

Approximate Fourier inversion

$$\Pr[\mathbf{S} \le x] - \Pr[\mathbf{Z} \le x] \le \frac{1}{2\pi} \int_{\xi=-T}^{\xi=T} \frac{|\widehat{\mathbf{S}}(\xi) - \widehat{\mathbf{Z}}(\xi)|}{|\xi|} d\xi + O\left(\frac{1}{T}\right).$$

Approximate Fourier inversion

Strategy to prove the Berry-Eséeen theorem

Approximate Fourier inversion

$$\Pr[\mathbf{S} \leq x] - \Pr[\mathbf{Z} \leq x] \leq \frac{1}{2\pi} \int_{\xi=-T}^{\xi=T} \frac{|\widehat{\mathbf{S}}(\xi) - \widehat{\mathbf{Z}}(\xi)|}{|\xi|} d\xi + O\left(\frac{1}{T}\right).$$

Approximate Fourier inversion

Strategy to prove the Berry-Eséeen theorem

Approximate Fourier inversion

$$\Pr[\mathbf{S} \leq x] - \Pr[\mathbf{Z} \leq x] \leq \frac{1}{2\pi} \int_{\xi=-T}^{\xi=T} \frac{|\widehat{\mathbf{S}}(\xi) - \widehat{\mathbf{Z}}(\xi)|}{|\xi|} d\xi + O\left(\frac{1}{T}\right).$$

Choose $T \approx \sqrt{n}$. We will bound $|\widehat{\mathbf{S}}(\xi) - \widehat{\mathbf{Z}}(\xi)|$ for $|\xi| \leq T$.

Showing $\widehat{\mathbf{S}}(\xi)$ is close to $\widehat{\mathbf{Z}}(\xi)$

Let us start with $\widehat{\mathbf{Z}}(\xi)$. Recall $\mathbf{Z} = \mathcal{N}(0,1)$. It easily follows that

$$\widehat{\mathbf{Z}}(\xi) = \int_x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} e^{i\xi x} dx = e^{-\frac{\xi^2}{2}}.$$

Showing $\widehat{\mathbf{S}}(\xi)$ is close to $\widehat{\mathbf{Z}}(\xi)$

Let us start with $\widehat{\mathbf{Z}}(\xi)$. Recall $\mathbf{Z} = \mathcal{N}(0,1)$. It easily follows that

$$\widehat{\mathbf{Z}}(\xi) = \int_x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} e^{i\xi x} dx = e^{-\frac{\xi^2}{2}}.$$

On the other hand,

$$\widehat{\mathbf{S}}(\xi) = (\widehat{\mathbf{X}}(\xi/\sqrt{n}))^n = \left(1 + \sum_{j=1}^{\infty} \frac{i^j \cdot \mathbf{E}[\mathbf{X}^j]}{j!} \frac{\xi^j}{n^{j/2}}\right)^n$$

$$\widehat{\mathbf{S}}(\xi) = \left(1 - \frac{\xi^2}{2n} + o(|\xi|^2/n)\right)^n.$$

Showing $\widehat{\mathbf{S}}(\xi)$ is close to $\widehat{\mathbf{Z}}(\xi)$

Note: Taylor expansion of $\widehat{\mathbf{X}}(\xi/\sqrt{n})$ is valid only if $|\xi|$ is small.

Claim: For $|\xi| \leq \frac{\sqrt{n}}{100}$, we have

$$\left|\widehat{\mathbf{S}}(\xi) - \widehat{\mathbf{Z}}(\xi)\right| = O\left(\frac{1}{\sqrt{n}}|\xi|^3 e^{-\xi^2/3}\right).$$

Showing $\widehat{\mathbf{S}}(\xi)$ is close to $\widehat{\mathbf{Z}}(\xi)$

Note: Taylor expansion of $\widehat{\mathbf{X}}(\xi/\sqrt{n})$ is valid only if $|\xi|$ is small.

Claim: For $|\xi| \leq \frac{\sqrt{n}}{100}$, we have

$$\left|\widehat{\mathbf{S}}(\xi) - \widehat{\mathbf{Z}}(\xi)\right| = O\left(\frac{1}{\sqrt{n}}|\xi|^3 e^{-\xi^2/3}\right).$$

Plugging this back into *approximate Fourier inversion* (with $T = \sqrt{n}/100$),

$$\Pr[\mathbf{S} \leq x] - \Pr[\mathbf{Z} \leq x] \leq \frac{1}{2\pi}\int_{\xi=-T}^{\xi=T} \frac{|\xi|^2 e^{-\xi^2/3}}{\sqrt{n}} d\xi + O\left(\frac{1}{T}\right).$$

Showing $\widehat{\mathbf{S}}(\xi)$ is close to $\widehat{\mathbf{Z}}(\xi)$

Note: Taylor expansion of $\widehat{\mathbf{X}}(\xi/\sqrt{n})$ is valid only if $|\xi|$ is small.

Claim: For $|\xi| \le \frac{\sqrt{n}}{100}$, we have

$$\left|\widehat{\mathbf{S}}(\xi) - \widehat{\mathbf{Z}}(\xi)\right| = O\left(\frac{1}{\sqrt{n}}|\xi|^3 e^{-\xi^2/3}\right).$$

Showing $\widehat{\mathbf{S}}(\xi)$ is close to $\widehat{\mathbf{Z}}(\xi)$

Note: Taylor expansion of $\widehat{\mathbf{X}}(\xi/\sqrt{n})$ is valid only if $|\xi|$ is small.

Claim: For $|\xi| \leq \frac{\sqrt{n}}{100}$, we have

$$\left|\widehat{\mathbf{S}}(\xi) - \widehat{\mathbf{Z}}(\xi)\right| = O\left(\frac{1}{\sqrt{n}}|\xi|^3 e^{-\xi^2/3}\right).$$

Proof technique: Split $|\xi|$ into the high $|\xi|$ regime and the low $|\xi|$ regime. In particular, define

$$\Gamma_{\text{low}} = \{\xi : |\xi| \leq n^{\frac{1}{6}}\} \text{ and } \Gamma_{\text{high}} = \{\xi : n^{\frac{1}{2}}/100 \geq |\xi| > n^{\frac{1}{6}}\}.$$

Showing $\widehat{\mathbf{S}}(\xi)$ is close to $\widehat{\mathbf{Z}}(\xi)$

Note: Taylor expansion of $\widehat{\mathbf{X}}(\xi/\sqrt{n})$ is valid only if $|\xi|$ is small.

Claim: For $|\xi| \leq \frac{\sqrt{n}}{100}$, we have

$$\left|\widehat{\mathbf{S}}(\xi) - \widehat{\mathbf{Z}}(\xi)\right| = O\left(\frac{1}{\sqrt{n}}|\xi|^3 e^{-\xi^2/3}\right).$$

Showing $\widehat{\mathbf{S}}(\xi)$ is close to $\widehat{\mathbf{Z}}(\xi)$

Note: Taylor expansion of $\widehat{\mathbf{X}}(\xi/\sqrt{n})$ is valid only if $|\xi|$ is small.

Claim: For $|\xi| \leq \frac{\sqrt{n}}{100}$, we have

$$\left|\widehat{\mathbf{S}}(\xi) - \widehat{\mathbf{Z}}(\xi)\right| = O\left(\frac{1}{\sqrt{n}}|\xi|^3 e^{-\xi^2/3}\right).$$

Proof technique: When $\xi \in \Gamma_{\text{low}}$, then we apply Taylor's expansion – recall

$$\widehat{\mathbf{S}}(\xi) = \left(1 - \frac{\xi^2}{2n} + o(|\xi|^2/n)\right)^n.$$

Showing $\widehat{\mathbf{S}}(\xi)$ is close to $\widehat{\mathbf{Z}}(\xi)$

Note: Taylor expansion of $\widehat{\mathbf{X}}(\xi/\sqrt{n})$ is valid only if $|\xi|$ is small.

Claim: For $|\xi| \le \frac{\sqrt{n}}{100}$, we have

$$\left|\widehat{\mathbf{S}}(\xi) - \widehat{\mathbf{Z}}(\xi)\right| = O\left(\frac{1}{\sqrt{n}}|\xi|^3 e^{-\xi^2/3}\right).$$

Showing $\widehat{\mathbf{S}}(\xi)$ is close to $\widehat{\mathbf{Z}}(\xi)$

Note: Taylor expansion of $\widehat{\mathbf{X}}(\xi/\sqrt{n})$ is valid only if $|\xi|$ is small.

Claim: For $|\xi| \leq \frac{\sqrt{n}}{100}$, we have

$$\left|\widehat{\mathbf{S}}(\xi) - \widehat{\mathbf{Z}}(\xi)\right| = O\left(\frac{1}{\sqrt{n}}|\xi|^3 e^{-\xi^2/3}\right).$$

Proof technique: On the other hand, it is not difficult to show that $|\widehat{\mathbf{S}}(\xi)| \leq e^{-\frac{2\xi^2}{3}}$. Using the fact that $|\widehat{\mathbf{Z}}(\xi)| = e^{-\frac{\xi^2}{2}}$. When $\xi \in \Gamma_{\mathrm{high}}$, this is enough

Finishing the proof of Berry-Esséen

For all $|\xi| \leq \frac{\sqrt{n}}{100}$, we have

$$\left|\widehat{\mathbf{S}}(\xi) - \widehat{\mathbf{Z}}(\xi)\right| = O\left(\frac{1}{\sqrt{n}}|\xi|^3 e^{-\xi^2/3}\right).$$

Finishing the proof of Berry-Esséen

For all $|\xi| \le \frac{\sqrt{n}}{100}$, we have

$$\left|\widehat{\mathbf{S}}(\xi) - \widehat{\mathbf{Z}}(\xi)\right| = O\left(\frac{1}{\sqrt{n}}|\xi|^3 e^{-\xi^2/3}\right).$$

Plugging this back into the approximate Fourier inversion formula (which is)

$$|\Pr[\mathbf{S} \le x] - \Pr[\mathbf{Z} \le x]| \le \frac{1}{2\pi}\int_{\xi=-T}^{\xi=T} \frac{|\widehat{\mathbf{S}}(\xi) - \widehat{\mathbf{Z}}(\xi)|}{|\xi|} d\xi + O\left(\frac{1}{T}\right),$$

we get $\Pr[\mathbf{S} \le x] - \Pr[\mathbf{Z} \le x]| = O(n^{-1/2})$.

Application of the Berry-Esséen theorem

Stochastic knapsack

Suppose you have n items, each with a profit c_i and a stochastic weight $\mathbf{X}_i$ where each $\mathbf{X}_i$ is a positive valued random variable.

Application of the Berry-Esséen theorem

Stochastic knapsack

Suppose you have n items, each with a profit c_i and a stochastic weight $\mathbf{X}_i$ where each $\mathbf{X}_i$ is a positive valued random variable.

Goal: Given a knapsack with capacity θ and error tolerance probability p, pack a subset S of items such that

$$\Pr[\sum_{j \in S} \mathbf{X}_j \leq \theta] \geq 1 - p,$$

so that the *profit* $\sum_{j \in S} c_j$ is maximized.

Berry-Esséen theorem for stochastic knapsack

Stochastic knapsack

Suppose you have n items, each with a profit c_i and a stochastic weight $\mathbf{X}_i$ where each $\mathbf{X}_i$ is of the form

$$\mathbf{X}_i = \begin{cases} w_{\ell,i} & \text{w.p. } \frac{1}{2} \\ w_{h,i} & \text{w.p. } \frac{1}{2} \end{cases}$$

Here all $w_{\ell,i} \in [1, \ldots, M/4]$ and $w_{h,i} \in [3M/4, \ldots, M]$ where $M = \text{poly}(n)$. Further, all profits $c_i \in [1, \ldots, M]$.

Algorithmic result for stochastic knapsack

Result: There is an algorithm which for any error parameter $\epsilon > 0$, runs in time $\mathrm{poly}(M, n^{1/\epsilon^2})$ and outputs a set S_* such that

$$\Pr[\sum_{j \in S_*} \mathbf{X}_j \leq \theta] \geq 1 - p - \epsilon,$$

such that $\sum_{j \in S_*} c_j = \mathrm{OPT}$.

Algorithmic result for stochastic knapsack

Result: There is an algorithm which for any error parameter $\epsilon > 0$, runs in time $\text{poly}(M, n^{1/\epsilon^2})$ and outputs a set S_* such that

$$\Pr[\sum_{j \in S_*} \mathbf{X}_j \leq \theta] \geq 1 - p - \epsilon,$$

such that $\sum_{j \in S_*} c_j = \mathrm{OPT}$.

Key feature: We do not relax the knapsack capacity θ.

Algorithmic result for stochastic knapsack

Result: There is an algorithm which for any error parameter $\epsilon > 0$, runs in time $\text{poly}(M, n^{1/\epsilon^2})$ and outputs a set S_* such that

$$\Pr[\sum_{j \in S_*} \mathbf{X}_j \leq \theta] \geq 1 - p - \epsilon,$$

such that $\sum_{j \in S_*} c_j = \text{OPT}$.

Key feature: We do not relax the knapsack capacity θ. See paper in SODA 2018 for the most general version of results.

Main idea behind the algorithm

Observation I: If we “center” random variable $\mathbf{X}_i$, i.e, $\mathbf{Y}_i = \mathbf{X}_i - \mathbf{E}[\mathbf{X}_i]$, then it satisfies

$$\mathbf{E}[|\mathbf{Y}_i|^3] \leq \left(\mathbf{E}[|\mathbf{Y}_i|^2]\right)^{3/2}.$$

Main idea behind the algorithm

Observation I: If we "center" random variable $\mathbf{X}_i$, i.e, $\mathbf{Y}_i = \mathbf{X}_i - \mathbf{E}[\mathbf{X}_i]$, then it satisfies

$$\mathbf{E}[|\mathbf{Y}_i|^3] \leq \left(\mathbf{E}[|\mathbf{Y}_i|^2]\right)^{3/2}.$$

Thus, we can potentially apply Berry-Esséen to a sum of $\mathbf{X}_i$.

Main idea behind the algorithm

Observation I: If we "center" random variable $\mathbf{X}_i$, i.e, $\mathbf{Y}_i = \mathbf{X}_i - \mathbf{E}[\mathbf{X}_i]$, then it satisfies

$$\mathbf{E}[|\mathbf{Y}_i|^3] \leq \left(\mathbf{E}[|\mathbf{Y}_i|^2]\right)^{3/2}.$$

Thus, we can potentially apply Berry-Esséen to a sum of $\mathbf{X}_i$.

Observation II: Consider any subset of items S with $|S| \geq 100/\epsilon^2$. Then,

$$\max_i \mathrm{Var}(\mathbf{X}_i) \leq \epsilon^2 \cdot (\sum_{j \in S} \mathrm{Var}(\mathbf{X}_j)).$$

Algorithmic idea

Step 1: Either the optimum solution S_{opt} is such that $|S_{\mathrm{opt}}| \leq 100/\epsilon^2$. In this case, we can *brute-force* search for S_{opt}. Running time is $n^{\Theta(1/\epsilon^2)}$.

Algorithmic idea

Step 1: Either the optimum solution S_{opt} is such that $|S_{\text{opt}}| \leq 100/\epsilon^2$. In this case, we can *brute-force* search for S_{opt}. Running time is $n^{\Theta(1/\epsilon^2)}$.

Step 2: Otherwise, $|S_{\text{opt}}| > 100/\epsilon^2$. In this case, define μ_{opt}, σ^2_{opt} and C_{opt} as (i) $\mu_{\text{opt}} = \sum_{j \in S_{\text{opt}}} \mathbf{E}[\mathbf{X}_j]$; (ii) $\sigma^2_{\text{opt}} = \sum_{j \in S_{\text{opt}}} \text{Var}(\mathbf{X}_j)$; (iii) $C_{\text{opt}} = \sum_{j \in S_{\text{opt}}} c_j$.

Algorithmic idea

Observe that μ_{opt}, σ^2_{opt} and C_{opt} are all integral multiples of $1/4$ bounded by M^2.

Algorithmic idea

Observe that μ_{opt}, σ^2_{opt} and C_{opt} are all integral multiples of $1/4$ bounded by M^2.

We use dynamic programming to find S_* such that $C_{\text{opt}} = C_*$, $\mu_{\text{opt}} = \mu_*$ and $\sigma^2_{\text{opt}} = \sigma^2_*$.

Algorithmic idea

Observe that μ_{opt}, σ^2_{opt} and C_{opt} are all integral multiples of $1/4$ bounded by M^2.

We use dynamic programming to find S_* such that $C_{\text{opt}} = C_*$, $\mu_{\text{opt}} = \mu_*$ and $\sigma^2_{\text{opt}} = \sigma^2_*$.

Consequence of Berry-Esséen theorem:

$$\Pr[\sum_{j \in S_*} \mathbf{X}_j \leq \theta] \geq \Pr[\sum_{j \in S_{\text{opt}}} \mathbf{X}_j \leq \theta] - \epsilon.$$

Algorithmic idea

Consequence of Berry-Esséen theorem:

$$\Pr[\sum_{j \in S_*} \mathbf{X}_j \leq \theta] \geq \Pr[\sum_{j \in S_{\text{opt}}} \mathbf{X}_j \leq \theta] - \epsilon.$$

This is because by Berry-Esséen, the distribution of $\sum_{j \in S_*} \mathbf{X}_j$ and $\sum_{j \in S_{\text{opt}}} \mathbf{X}_j$ follows essentially Gaussian (and their means and variances match).

General algorithmic result for stochastic optimization

Suppose the item sizes $\{\mathbf{X}_i\}_{i=1}^n$ are all *hypercontractive* – i.e., $\mathbf{E}[|\mathbf{X}_i|^3] \leq O(1) \cdot (\mathbf{E}[|\mathbf{X}_i|^2])^{3/2}$.

General algorithmic result for stochastic optimization

Suppose the item sizes $\{\mathbf{X}_i\}_{i=1}^n$ are all *hypercontractive* – i.e., $\mathbf{E}[|\mathbf{X}_i|^3] \leq O(1) \cdot (\mathbf{E}[|\mathbf{X}_i|^2])^{3/2}$.

Theorem: When item sizes are hypercontractive, then there is an algorithm running in time $n^{O(1/\epsilon^2)}$ such that the output set S_* satisfies

1. $\sum_{j \in S_*} c_j \geq (1-\epsilon) \cdot (\sum_{j \in S_{\mathrm{opt}}} c_j)$.
2. $\Pr[\sum_{j \in S_*} \mathbf{X}_j \leq \theta] \geq \Pr[\sum_{j \in S_{\mathrm{opt}}} \mathbf{X}_j \leq \theta] - \epsilon$.

General algorithmic result for stochastic optimization

Suppose the item sizes $\{\mathbf{X}_i\}_{i=1}^n$ are all *hypercontractive* – i.e., $\mathbf{E}[|\mathbf{X}_i|^3] \leq O(1) \cdot (\mathbf{E}[|\mathbf{X}_i|^2])^{3/2}$.

Theorem: When item sizes are hypercontractive, then there is an algorithm running in time $n^{O(1/\epsilon^2)}$ such that the output set S_* satisfies

1. $\sum_{j \in S_*} c_j \geq (1-\epsilon) \cdot (\sum_{j \in S_{\mathrm{opt}}} c_j)$.
2. $\Pr[\sum_{j \in S_*} \mathbf{X}_j \leq \theta] \geq \Pr[\sum_{j \in S_{\mathrm{opt}}} \mathbf{X}_j \leq \theta] - \epsilon$.

Read SODA 2018 paper for more details.

Central limit theorems: *Citius*, *Altius*, *Fortius*

Central limit theorems: *Citius*, *Altius*, *Fortius*

Let's do *Altius* – as in higher degree polynomials.

Central limit theorems: *Citius*, *Altius*, *Fortius*

Let's do *Altius* – as in higher degree polynomials.

Berry-Esséen says that sums of independent random variables under mild conditions converges to a Gaussian.

Let's do *Altius* – as in higher degree polynomials.

Berry-Esséen says that sums of independent random variables under mild conditions converges to a Gaussian.

What if we replace the sum by a polynomial? Let us think of the easy case when the degree is 2.

Central limit theorem for low-degree polynomials

Consider $p(x) = \left(\frac{x_1 + \ldots + x_n}{\sqrt{n}}\right)^2$. As $n \to \infty$, $x_1, \ldots, x_n$ are i.i.d. copies of unbiased ± 1 random variables, the distribution of $p(x)$ goes to a

Central limit theorem for low-degree polynomials

Consider $p(x) = \left(\frac{x_1+\ldots+x_n}{\sqrt{n}}\right)^2$. As $n \to \infty$, $x_1, \ldots, x_n$ are i.i.d. copies of unbiased ± 1 random variables, the distribution of $p(x)$ goes to a χ^2 distribution.

Central limit theorem for low-degree polynomials

Consider $p(x) = \left(\frac{x_1+\ldots+x_n}{\sqrt{n}}\right)^2$. As $n \to \infty$, $x_1, \ldots, x_n$ are i.i.d. copies of unbiased ± 1 random variables, the distribution of $p(x)$ goes to a χ^2 distribution.

In fact, suppose $p(x)$ is of degree-2 and of the following form:

$$p(x) = \lambda \cdot \ell^2(x) + q(x)m$$

where $\ell(x)$ is a linear form and $\lambda = \mathbf{E}[p(x) \cdot \ell^2(x)]$. If λ is large, then $p(x)$ is very far from a Gaussian.

Central limit theorem for quadratic polynomials

Theorem

Let $p(x) : \mathbb{R}^n \to \mathbb{R}$ *such that* $\mathrm{Var}(p(x)) = 1$ *and* $\mathbf{E}[p(x)] = \mu$. *Express* $p(x) = x^T Ax + \langle b, x \rangle + c$ *where* $A \in \mathbb{R}^{n \times n}$ *and* $b \in \mathbb{R}^n$. *Let* $\|A\|_{op} \leq \epsilon$ *and* $\|b\|_\infty \leq \epsilon$. *Suppose,* $\mathbf{x} \sim \{-1, 1\}^n$. *Then,*

$$d_K(p(\mathbf{x}), \mathcal{N}(\mu, 1)) = O(\sqrt{\epsilon}).$$

If $p(x)$ is not correlated with product of two linear forms, then it is distributed as a Gaussian.

Central limit theorem for higher degree polynomials

Corresponding to any multilinear polynomial $p : \mathbb{R}^n \to \mathbb{R}$ of degree-d, we have a sequence of tensors $(\mathcal{A}_d, \ldots, \mathcal{A}_0)$ where $\mathcal{A}_i \in \mathbb{R}^{n^{\times i}}$ is a tensor of order i.

Central limit theorem for higher degree polynomials

Corresponding to any multilinear polynomial $p : \mathbb{R}^n \to \mathbb{R}$ of degree-d, we have a sequence of tensors $(\mathcal{A}_d, \ldots, \mathcal{A}_0)$ where $\mathcal{A}_i \in \mathbb{R}^{n^{\times i}}$ is a tensor of order i.

For a tensor $\mathcal{A}_i$ (where $i > 1$), we use $\sigma_{\max}(\mathcal{A}_i)$ to denote the "maximum singular value" obtained by a non-trivial flattening.

Central limit theorem for higher degree polynomials

Corresponding to any multilinear polynomial $p : \mathbb{R}^n \to \mathbb{R}$ of degree-d, we have a sequence of tensors $(\mathcal{A}_d, \ldots, \mathcal{A}_0)$ where $\mathcal{A}_i \in \mathbb{R}^{n^{\times i}}$ is a tensor of order i.

For a tensor $\mathcal{A}_i$ (where $i > 1$), we use $\sigma_{\max}(\mathcal{A}_i)$ to denote the "maximum singular value" obtained by a non-trivial flattening.

Theorem

Let $p : \mathbb{R}^n \to \mathbb{R}$ be a degree-d polynomial with $\mathrm{Var}(p(x)) = 1$ and $\mathbf{E}[p(x)] = \mu$. Let $(\mathcal{A}_d, \ldots, \mathcal{A}_0)$ denote the tensors corresponding to p. Then,

$$d_K(p(\mathbf{x}), \mathcal{N}(\mu, 1)) = O_d(\sqrt{\epsilon}),$$

where $\mathbf{x} \sim \{-1, 1\}^n$. Here $\epsilon \geq \max_{j>1} \sigma_{\max}(\mathcal{A}_j)$ and $\epsilon \geq \|\mathcal{A}_0\|_\infty$.

Features of the central limit theorem

1. Qualitatively tight: in particular, for a polynomial if $\max_{j>1} \sigma_{\max}(\mathcal{A}_j)$ is large, then the distribution of $p(\mathbf{x})$ does not look like a Gaussian.

Features of the central limit theorem

1. Qualitatively tight: in particular, for a polynomial if $\max_{j>1} \sigma_{\max}(\mathcal{A}_j)$ is large, then the distribution of $p(\mathbf{x})$ does not look like a Gaussian.

2. $\max_{j>1} \sigma_{\max}(\mathcal{A}_j)$ is essentially capturing correlation of $p(\mathbf{x})$ with product of two lower degree polynomials.

Features of the central limit theorem

1. Qualitatively tight: in particular, for a polynomial if $\max_{j>1} \sigma_{\max}(\mathcal{A}_j)$ is large, then the distribution of $p(\mathbf{x})$ does not look like a Gaussian.

2. $\max_{j>1} \sigma_{\max}(\mathcal{A}_j)$ is essentially capturing correlation of $p(\mathbf{x})$ with product of two lower degree polynomials.

3. Condition for convergence to normal is efficiently checkable.

Proof of the central limit theorem

- The first step is to go from $\mathbf{x} \sim \{-1,1\}^n$ to $\mathbf{x} \sim \mathcal{N}^n(0,1)$. Accomplished via the *invariance principle*.

Proof of the central limit theorem

- The first step is to go from $\mathbf{x} \sim \{-1, 1\}^n$ to $\mathbf{x} \sim \mathcal{N}^n(0, 1)$. Accomplished via the *invariance principle*.

- Once in the Gaussian domain, the question is *when does a polynomial of a Gaussian look like a Gaussian?*

Proof of the central limit theorem

- The first step is to go from $\mathbf{x} \sim \{-1,1\}^n$ to $\mathbf{x} \sim \mathcal{N}^n(0,1)$. Accomplished via the *invariance principle*.

- Once in the Gaussian domain, the question is *when does a polynomial of a Gaussian look like a Gaussian?*

- Proof technique: Stein's method + Malliavin calculus.

Central limit theorem – application in derandomization

- Derandomization – fertile ground both for applications and discovery of central limit theorems (in computer science)

Central limit theorem – application in derandomization

- Derandomization – fertile ground both for applications and discovery of central limit theorems (in computer science)
- Why is central limit theorem useful for derandomization?

Central limit theorem – application in derandomization

- Derandomization – fertile ground both for applications and discovery of central limit theorems (in computer science)
- Why is central limit theorem useful for derandomization?
- Example: Suppose we are given a halfspace $f : \{-1, 1\}^n \to \{-1, 1\}$ where $f(x) = \text{sign}(\sum_{i=1}^{n} w_i x_i - \theta)$.

Central limit theorem – application in derandomization

- Derandomization – fertile ground both for applications and discovery of central limit theorems (in computer science)
- Why is central limit theorem useful for derandomization?
- Example: Suppose we are given a halfspace $f : \{-1,1\}^n \to \{-1,1\}$ where $f(x) = \text{sign}(\sum_{i=1}^n w_i x_i - \theta)$.

  Deterministically compute $\Pr_{\mathbf{x} \in \{-1,1\}^n}[f(\mathbf{x}) = 1]$.

Derandomizing halfspaces

- For $f(x) = \text{sign}(\sum_{i=1}^{n} w_i x_i - \theta)$, exactly computing $\Pr_{\mathbf{x} \in \{-1,1\}^n}[f(\mathbf{x}) = 1]$ is $\#P$-hard.

Derandomizing halfspaces

- For $f(x) = \text{sign}(\sum_{i=1}^n w_i x_i - \theta)$, exactly computing $\Pr_{\mathbf{x} \in \{-1,1\}^n}[f(\mathbf{x}) = 1]$ is $\#P$-hard.

- Computing $\Pr_{\mathbf{x} \in \{-1,1\}^n}[f(\mathbf{x}) = 1]$ to additive error ϵ is trivial using randomness.

Derandomizing halfspaces

- For $f(x) = \mathrm{sign}(\sum_{i=1}^{n} w_i x_i - \theta)$, exactly computing $\Pr_{\mathbf{x} \in \{-1,1\}^n}[f(\mathbf{x}) = 1]$ is $\#P$-hard.
- Computing $\Pr_{\mathbf{x} \in \{-1,1\}^n}[f(\mathbf{x}) = 1]$ to additive error ϵ is trivial using randomness.
- What can we do deterministically? or how are CLTs going to be useful?

Derandomizing halfspaces

- For $f(x) = \text{sign}(\sum_{i=1}^{n} w_i x_i - \theta)$, exactly computing $\Pr_{\mathbf{x} \in \{-1,1\}^n}[f(\mathbf{x}) = 1]$ is $\#P$-hard.
- Computing $\Pr_{\mathbf{x} \in \{-1,1\}^n}[f(\mathbf{x}) = 1]$ to additive error ϵ is trivial using randomness.
- What can we do deterministically? or how are CLTs going to be useful?
- [Servedio 2007]: Suppose all the $|w_i| \leq \epsilon/100$ (where $\|w\|_2 = 1$).

Berry-Esséen in action

$$\Pr_{\mathbf{x}\in\{-1,1\}^n}\Big[\sum_{i=1}^{n} w_i x_i - \theta \geq 0\Big] \approx_\epsilon \Pr_{\mathbf{g}\sim\mathcal{N}(0,1)}\big[\mathbf{g} - \theta \geq 0\big]$$

Berry-Esséen in action

$$\Pr_{\mathbf{x}\in\{-1,1\}^n}\left[\sum_{i=1}^{n} w_i x_i - \theta \geq 0\right] \approx_\epsilon \Pr_{\mathbf{g}\sim\mathcal{N}(0,1)}\left[\mathbf{g} - \theta \geq 0\right]$$

- However, $\Pr_{\mathbf{g}\sim\mathcal{N}(0,1)}\left[\mathbf{g} - \theta \geq 0\right]$ can be computed in $O_\epsilon(1)$ time.

Berry-Esséen in action

$$\Pr_{\mathbf{x}\in\{-1,1\}^n}\left[\sum_{i=1}^{n} w_i x_i - \theta \geq 0\right] \approx_\epsilon \Pr_{\mathbf{g}\sim\mathcal{N}(0,1)}\left[\mathbf{g} - \theta \geq 0\right]$$

- However, $\Pr_{\mathbf{g}\sim\mathcal{N}(0,1)}\left[\mathbf{g} - \theta \geq 0\right]$ can be computed in $O_\epsilon(1)$ time.
- Thus, when $|w_i| \leq \epsilon/100$, $\Pr_{\mathbf{x}\in\{-1,1\}^n}\left[\sum_{i=1}^{n} w_i x_i - \theta \geq 0\right]$ can be computed to $\pm\epsilon$ in $O_\epsilon(1) \cdot \text{poly}(n)$ time.

What if $\max |w_i| \geq \epsilon/100$?

- Suppose $|w_1| \geq \epsilon/100$. We recurse on the variable x_1.

$$f_{x_1=1} = \mathrm{sign}(\sum_{i=2}^{n} w_i x_i - \theta + w_1); \;\; f_{x_1=-1} = \mathrm{sign}(\sum_{i=2}^{n} w_i x_i - \theta - w_1)$$

What if $\max |w_i| \geq \epsilon/100$?

- Suppose $|w_1| \geq \epsilon/100$. We recurse on the variable x_1.

$$f_{x_1=1} = \mathrm{sign}(\sum_{i=2}^{n} w_i x_i - \theta + w_1); \;\; f_{x_1=-1} = \mathrm{sign}(\sum_{i=2}^{n} w_i x_i - \theta - w_1)$$

- Observe that it suffices to compute

$$\frac{1}{2} \cdot \left(\Pr_{\mathbf{x} \in \{-1,1\}^{n-1}} [f_{x_1=1}(\mathbf{x}) = 1] + \Pr_{\mathbf{x} \in \{-1,1\}^{n-1}} [f_{x_1=-1}(\mathbf{x}) = 1] \right).$$

Berry-Esséen in recursive action

- Either $\max_{j \geq 2} |w_j| \leq (\epsilon/100) \cdot \sqrt{\sum_{i=2}^{n} w_i^2}$.

Berry-Esséen in recursive action

- Either $\max_{j \geq 2} |w_j| \leq (\epsilon/100) \cdot \sqrt{\sum_{i=2}^{n} w_i^2}$.
- If yes, we can apply the Berry-Esséen theorem.

Berry-Esséen in recursive action

- Either $\max_{j\geq 2}|w_j| \leq (\epsilon/100)\cdot\sqrt{\sum_{i=2}^{n} w_i^2}$.
- If yes, we can apply the Berry-Esséen theorem.
- Else, we restrict w_2. Note: every time we restrict a variable, we capture an ϵ-fraction of the remaining ℓ_2 mass.

Berry-Esséen in recursive action

- Either $\max_{j \geq 2} |w_j| \leq (\epsilon/100) \cdot \sqrt{\sum_{i=2}^{n} w_i^2}$.
- If yes, we can apply the Berry-Esséen theorem.
- Else, we restrict w_2. Note: every time we restrict a variable, we capture an ϵ-fraction of the remaining ℓ_2 mass.
- Suppose the process goes on for j iterations. Either $j \leq \epsilon^{-1} \log(1/\epsilon)$ or $j > \epsilon^{-1} \log(1/\epsilon)$.

Berry-Esséen in recursive action

- If $j \le \epsilon^{-1} \log(1/\epsilon)$, then this reduces the problem to $\exp(1/\epsilon)$ subproblems – each of which can be solved using Berry-Esséen.

Berry-Esséen in recursive action

- If $j \leq \epsilon^{-1}\log(1/\epsilon)$, then this reduces the problem to $\exp(1/\epsilon)$ subproblems – each of which can be solved using Berry-Esséen.

- If $j > \epsilon^{-1}\log(1/\epsilon)$, we simply stop at $j = \epsilon^{-1}\log(1/\epsilon)$.

- The top $\epsilon^{-1}\log(1/\epsilon)$ weights capture most of the ℓ_2 mass.

Berry-Esséen in recursive action

- If $j \leq \epsilon^{-1}\log(1/\epsilon)$, then this reduces the problem to $\exp(1/\epsilon)$ subproblems – each of which can be solved using Berry-Esséen.
- If $j > \epsilon^{-1}\log(1/\epsilon)$, we simply stop at $j = \epsilon^{-1}\log(1/\epsilon)$.
- The top $\epsilon^{-1}\log(1/\epsilon)$ weights capture most of the ℓ_2 mass.
- Non-trivial: Since $\epsilon^{-1}\log(1/\epsilon)$ weights capture most of the mass of the vector w, we can just consider the halfspace over these variables.

Finishing the proof

- Thus if $j \geq \epsilon^{-1} \log(1/\epsilon)$, then we have reduced it to a $\epsilon^{-1} \log(1/\epsilon)$-dimensional problem.

Finishing the proof

- Thus if $j \geq \epsilon^{-1} \log(1/\epsilon)$, then we have reduced it to a $\epsilon^{-1} \log(1/\epsilon)$-dimensional problem.
- Meta idea: If the condition of CLT is met, apply.

Finishing the proof

- Thus if $j \geq \epsilon^{-1}\log(1/\epsilon)$, then we have reduced it to a $\epsilon^{-1}\log(1/\epsilon)$-dimensional problem.

- Meta idea: If the condition of CLT is met, apply.

- If it is not met, then restrict a variable and recurse.

Finishing the proof

- Thus if $j \geq \epsilon^{-1} \log(1/\epsilon)$, then we have reduced it to a $\epsilon^{-1} \log(1/\epsilon)$-dimensional problem.

- Meta idea: If the condition of CLT is met, apply.

- If it is not met, then restrict a variable and recurse.

- Meta-idea repeated in several works in derandomization which use limit theorems.

Central limit theorem: *Fortius*

- So far, we have seen central limit theorems which provide convergence in Kolmogorov distance.

Central limit theorem: *Fortius*

- So far, we have seen central limit theorems which provide convergence in Kolmogorov distance.
- In other words, we choose a threshold x and compare $\Pr[\mathbf{S} \leq x]$ with $\Pr[\mathbf{Z} \leq x]$.

Central limit theorem: *Fortius*

- So far, we have seen central limit theorems which provide convergence in Kolmogorov distance.
- In other words, we choose a threshold x and compare $\Pr[\mathbf{S} \leq x]$ with $\Pr[\mathbf{Z} \leq x]$.
- It is possible to sometimes get convergence in stronger metrics.

Central limit theorem: *Fortius*

Theorem (Chen-Goldstein-Shao)

Let X_1, X_2, $\ldots, X_n$ *be independent Bernoulli random variables such that* $\mathbf{S} = \sum X_i$ *has mean* μ *and variance* σ^2*. Then,*

$$\|\mathbf{S} - \mathcal{N}_{\text{disc}}(\mu, \sigma^2)\|_1 = O(\sigma^{-1}).$$

Central limit theorem: *Fortius*

Theorem (Chen-Goldstein-Shao)

Let $X_1, X_2, \ldots, X_n$ be independent Bernoulli random variables such that $\mathbf{S} = \sum X_i$ has mean μ and variance σ^2. Then,

$$\|\mathbf{S} - \mathcal{N}_{\mathrm{disc}}(\mu, \sigma^2)\|_1 = O(\sigma^{-1}).$$

Discrete CLTs have found many applications in derandomization and learning.

Central limit theorem: *Fortius*

Theorem (Chen-Goldstein-Shao)

Let $X_1, X_2, \ldots, X_n$ be independent Bernoulli random variables such that $\mathbf{S} = \sum X_i$ has mean μ and variance σ^2. Then,

$$\|\mathbf{S} - \mathcal{N}_{\mathsf{disc}}(\mu, \sigma^2)\|_1 = O(\sigma^{-1}).$$

Discrete CLTs have found many applications in derandomization and learning.

Also check out the new discrete CLTs proven by Valiant-Valiant, Daskalakis-Kamath-Tzamos and many others.

Central limit theorem: *Citius*

- Recall: Sum of n-i.i.d. random variables converges to a Gaussian at a rate of $O(n^{-1/2})$.

Central limit theorem: *Citius*

- Recall: Sum of n-i.i.d. random variables converges to a Gaussian at a rate of $O(n^{-1/2})$.
- Without more conditions, not possible to beat $O(n^{-1/2})$.
- However, if the limiting distribution can include non-Gaussian distributions, we can get better than $n^{-1/2}$ convergence rate.

Central limit theorem: *Citius*

- Gaussians are parameterized by two parameters (i.e., two moments).

Central limit theorem: *Citius*

- Gaussians are parameterized by two parameters (i.e., two moments).

- By allowing richer families, say parameterized by k moments, one can get convergence rates of $n^{-k/2}$.

Central limit theorem: *Citius*

- Gaussians are parameterized by two parameters (i.e., two moments).
- By allowing richer families, say parameterized by k moments, one can get convergence rates of $n^{-k/2}$.
- However, conditions required are a little more delicate.

Central limit theorem: *Citius*

- Gaussians are parameterized by two parameters (i.e., two moments).
- By allowing richer families, say parameterized by k moments, one can get convergence rates of $n^{-k/2}$.
- However, conditions required are a little more delicate.
- Referred to as "asymptotic expansions" (see FOCS 2015 paper for a 'computer science' introduction).

Summary

- Central limit theorem(s) can be used to summarize statistics such as sums and low-degree polynomial of independent random variables.

Summary

- Central limit theorem(s) can be used to summarize statistics such as sums and low-degree polynomial of independent random variables.
- Many applications in learning theory, game theory, algorithms and complexity.

Summary

- Central limit theorem(s) can be used to summarize statistics such as sums and low-degree polynomial of independent random variables.

- Many applications in learning theory, game theory, algorithms and complexity.

- May be there are even CS inspired CLTs waiting to be discovered?

Thanks