Método de Stein

Arturo Jaramillo Gil

Centro de Investigación en Matemáticas (CIMAT)

Mis lineas de investigación

Principalmente, trabajo en

1. Cálculo de Malliavin (funcionales de procesos Gaussianos, ecuaciones estocásticas parciales, regularidad de distribuciones).
2. Movimiento Browniano fraccionario (autointersecciones, integrales estocásticas, leyes límite de estadísticos).
3. Tiempos locales (estadísticos de alta frecuencia, funcionales aditivos, tiempos de ocupación).
4. Matrices aleatorias (colisión de eigenvalores, ley de Wigner funcional, fluctuaciones de segundo orden del espectro).
5. Teoría de números probabilista (funciones aritméticas aditivas, función ζ de Riemann).
6. Teoremas límite y **Método de Stein**.

Contexto histórico

En 1971, Charles Stein (Stanford) publicó un manuscrito que estudiaba la distribución límite de sumas de variables aleatorias mediante una metodología moderna.

Contexto histórico

En 1971, Charles Stein (Stanford) publicó un manuscrito que estudiaba la distribución límite de sumas de variables aleatorias mediante una metodología moderna.

Consecuencias inmediatas

Cuantificar el error de aproximaciones gaussianas para sumas de variables aleatorias no necesariamente independientes.

Relevancia (parte 1)

Sea S_n una variable aleatoria de interés y sea $\mathcal{N}$ una v.a. gaussiana estándar.

Relevancia (parte 1)

Sea S_n una variable aleatoria de interés y sea $\mathcal{N}$ una v.a. gaussiana estándar. En muchas ocasiones, hay constantes μ_n, σ_n, t.q.

$$S_n \approx \mu_n + \sigma_n \mathcal{N},$$

Relevancia (parte 1)

Sea S_n una variable aleatoria de interés y sea $\mathcal{N}$ una v.a. gaussiana estándar. En muchas ocasiones, hay constantes μ_n, σ_n, t.q.

$$S_n \approx \mu_n + \sigma_n \mathcal{N}, \qquad \frac{S_n - \mu_n}{\sigma_n} \approx \mathcal{N}$$

Relevancia (parte 1)

Sea S_n una variable aleatoria de interés y sea $\mathcal{N}$ una v.a. gaussiana estándar. En muchas ocasiones, hay constantes μ_n, σ_n, t.q.

$$S_n \approx \mu_n + \sigma_n \mathcal{N}, \qquad \frac{S_n - \mu_n}{\sigma_n} \approx \mathcal{N}$$

Relevancia (parte 1)

Sea S_n una variable aleatoria de interés y sea $\mathcal{N}$ una v.a. gaussiana estándar. En muchas ocasiones, hay constantes μ_n, σ_n, t.q.

$$S_n \approx \mu_n + \sigma_n \mathcal{N}, \qquad \frac{S_n - \mu_n}{\sigma_n} \approx \mathcal{N}$$

¿Cómo medimos la calidad de dicha aproximación?

Distancias de probabilidad

Sea $\mathcal{C} := \{\mathbb{1}_{(-\infty,x]} \ ; \ x \in \mathbb{R}\}$. Definimos la distancia de Kolmogorov d_K entre S_n y $\mathcal{N}$ como

$$d_K(S_n, \mathcal{N}) := \sup_{x\in\mathbb{R}} |\mathbb{P}[S_n \leq x] - \mathbb{P}[\mathcal{N} \leq x]|$$

.

Distancias de probabilidad

Sea $\mathcal{C} := \{\mathbb{1}_{(-\infty,x]} \; ; \; x \in \mathbb{R}\}$. Definimos la distancia de Kolmogorov d_K entre S_n y $\mathcal{N}$ como

$$\begin{aligned} d_K(S_n, \mathcal{N}) &:= \sup_{x \in \mathbb{R}} |\mathbb{P}[S_n \leq x] - \mathbb{P}[\mathcal{N} \leq x]| \\ &= \sup_{h \in \mathcal{C}} |\mathbb{E}[h(S_n) - h(\mathcal{N})]|. \end{aligned}$$

Distancias de probabilidad

Sea $\mathcal{C} := \{\mathbb{1}_{(-\infty,x]} \ ; \ x \in \mathbb{R}\}$. Definimos la distancia de Kolmogorov d_K entre S_n y $\mathcal{N}$ como

$$\begin{aligned} d_K(S_n, \mathcal{N}) &:= \sup_{x \in \mathbb{R}} |\mathbb{P}[S_n \leq x] - \mathbb{P}[\mathcal{N} \leq x]| \\ &= \sup_{h \in \mathcal{C}} |\mathbb{E}[h(S_n) - h(\mathcal{N})]|. \end{aligned}$$

Distancias de probabilidad

Sea $\mathcal{C} := \{\mathbb{1}_{(-\infty,x]}\ ;\ x \in \mathbb{R}\}$. Definimos la distancia de Kolmogorov d_K entre S_n y $\mathcal{N}$ como

$$\begin{aligned} d_K(S_n,\mathcal{N}) &:= \sup_{x\in\mathbb{R}} |\mathbb{P}[S_n \leq x] - \mathbb{P}[\mathcal{N} \leq x]| \\ &= \sup_{h\in\mathcal{C}} |\mathbb{E}[h(S_n) - h(\mathcal{N})]|. \end{aligned}$$

El método de Stein es una colección de técnicas probabilistas que, en parte, permiten **estimar $\boldsymbol{d_K(S_n,\mathcal{N})}$.**

La heurística de Stein

La discrepancia de S_n y $\mathcal{N}$ se estudia mediante expresiones del tipo

$$|\mathbb{E}[h(S_n) - h(\mathcal{N})]|. \tag{1}$$

La heurística de Stein

La discrepancia de S_n y $\mathcal{N}$ se estudia mediante expresiones del tipo

$$|\mathbb{E}[h(S_n) - h(\mathcal{N})]|. \tag{1}$$

Notar que $S_n \overset{Law}{=} \mathcal{N}$ si se cumple la **propiedad** de que "(1) es cero para suficientes funciones de prueba h".

La heurística de Stein

La discrepancia de S_n y $\mathcal{N}$ se estudia mediante expresiones del tipo

$$|\mathbb{E}[h(S_n) - h(\mathcal{N})]|. \tag{1}$$

Notar que $S_n \stackrel{Law}{=} \mathcal{N}$ si se cumple la **propiedad** de que "(1) es cero para suficientes funciones de prueba h".

Otra **propiedad** que caracteriza la ley de $\mathcal{N}$...

La heurística de Stein

La discrepancia de S_n y $\mathcal{N}$ se estudia mediante expresiones del tipo

$$|\mathbb{E}[h(S_n) - h(\mathcal{N})]|. \tag{1}$$

Notar que $S_n \overset{Law}{=} \mathcal{N}$ si se cumple la **propiedad** de que "(1) es cero para suficientes funciones de prueba h".

Otra **propiedad** que caracteriza la ley de $\mathcal{N}$...

$$|\mathbb{E}[\mathcal{N} f(\mathcal{N}) - f'(\mathcal{N})]| = 0. \tag{2}$$

La heurística de Stein

La discrepancia de S_n y $\mathcal{N}$ se estudia mediante expresiones del tipo

$$|\mathbb{E}[h(S_n) - h(\mathcal{N})]|. \tag{1}$$

Notar que $S_n \overset{Law}{=} \mathcal{N}$ si se cumple la **propiedad** de que "(1) es cero para suficientes funciones de prueba h".

Otra **propiedad** que caracteriza la ley de $\mathcal{N}$...

$$|\mathbb{E}[\mathcal{N} f(\mathcal{N}) - f'(\mathcal{N})]| = 0. \tag{2}$$

Heurística de Stein

$$|\mathbb{E}[\mathcal{N} f(S_n) - f'(S_n)]| \approx 0 \qquad \Rightarrow \qquad |\mathbb{E}[h(S_n) - h(\mathcal{N})]| \approx 0.$$

Problemática central del método de Stein

La heurística de Stein reduce el estudio de $d_K(S_n, \mathcal{N})$ al estudio de

$$|\mathbb{E}[\mathcal{N} f(S_n) - f'(S_n)]|. \tag{3}$$

Problemática central del método de Stein

La heurística de Stein reduce el estudio de $d_K(S_n, \mathcal{N})$ al estudio de

$$|\mathbb{E}[\mathcal{N} f(S_n) - f'(S_n)]|. \tag{3}$$

Problemática principal del método: ¿cómo estimamos expresiones del tipo (3)?

Problemática central del método de Stein

La heurística de Stein reduce el estudio de $d_K(S_n, \mathcal{N})$ al estudio de

$$|\mathbb{E}[\mathcal{N} f(S_n) - f'(S_n)]|. \tag{3}$$

Problemática principal del método: ¿cómo estimamos expresiones del tipo (3)?

Avances

1. Sumas de variables aleatorias con “buenas propiedades” (Stein, Chen, Diaconis, Barbour, Röllin).

Problemática central del método de Stein

La heurística de Stein reduce el estudio de $d_K(S_n, \mathcal{N})$ al estudio de

$$|\mathbb{E}[\mathcal{N} f(S_n) - f'(S_n)]|. \tag{3}$$

Problemática principal del método: ¿cómo estimamos expresiones del tipo (3)?

Avances

1. Sumas de variables aleatorias con “buenas propiedades” (Stein, Chen, Diaconis, Barbour, Röllin).
2. Funcionales de procesos Gaussianos con “buenas propiedades” (Nualart, Peccati, Nourdin, Podolsjik, Tudor, Jaramillo).

Problemática central del método de Stein

La heurística de Stein reduce el estudio de $d_K(S_n, \mathcal{N})$ al estudio de

$$|\mathbb{E}[\mathcal{N} f(S_n) - f'(S_n)]|. \tag{3}$$

Problemática principal del método: ¿cómo estimamos expresiones del tipo (3)?

Avances

1. Sumas de variables aleatorias con "buenas propiedades" (Stein, Chen, Diaconis, Barbour, Röllin).
2. Funcionales de procesos Gaussianos con "buenas propiedades" (Nualart, Peccati, Nourdin, Podolsjik, Tudor, Jaramillo).
3. **Otros casos:** a veces se puede, pero cada problema se trata ad hoc.

Problemática central del método de Stein

La heurística de Stein reduce el estudio de $d_K(S_n, \mathcal{N})$ al estudio de

$$|\mathbb{E}[\mathcal{N} f(S_n) - f'(S_n)]|. \tag{3}$$

Problemática principal del método: ¿cómo estimamos expresiones del tipo (3)?

Avances

1. Sumas de variables aleatorias con "buenas propiedades" (Stein, Chen, Diaconis, Barbour, Röllin).
2. Funcionales de procesos Gaussianos con "buenas propiedades" (Nualart, Peccati, Nourdin, Podolsjik, Tudor, Jaramillo).
3. **Otros casos:** a veces se puede, pero cada problema se trata ad hoc. ¡Muchas veces es un problema difícil!

Problemática central del método de Stein

La heurística de Stein reduce el estudio de $d_K(S_n, \mathcal{N})$ al estudio de

$$|\mathbb{E}[\mathcal{N} f(S_n) - f'(S_n)]|. \tag{3}$$

Problemática principal del método: ¿cómo estimamos expresiones del tipo (3)?

Avances

1. Sumas de variables aleatorias con “buenas propiedades” (Stein, Chen, Diaconis, Barbour, Röllin).
2. Funcionales de procesos Gaussianos con “buenas propiedades” (Nualart, Peccati, Nourdin, Podolsjik, Tudor, Jaramillo).
3. **Otros casos:** a veces se puede, pero cada problema se trata ad hoc. ¡Muchas veces es un problema difícil!, pero interesante y relevante.

Espectro de la utilidad del método

Aplicaciones en:

1. Sumas de variables aleatorias (Stein, Chen, Barbour).

Espectro de la utilidad del método

Aplicaciones en:

1. Sumas de variables aleatorias (Stein, Chen, Barbour).
2. Análisis estocástico. Teorema de Breuer-Major, relación con caos de Wiener, estadísticos de procesos Gaussianos (Nualart, Peccati, Nourdin, Podolsjik, Tudor, Jaramillo).

Espectro de la utilidad del método

Aplicaciones en:

1. Sumas de variables aleatorias (Stein, Chen, Barbour).
2. Análisis estocástico. Teorema de Breuer-Major, relación con caos de Wiener, estadísticos de procesos Gaussianos (Nualart, Peccati, Nourdin, Podolsjik, Tudor, Jaramillo).
3. Teoría de números probabilista. Funciones aritméticas aditivas, multiplicidades primas de enteros aleatorios (Chen, Jaramillo, Yang, Harper).

Espectro de la utilidad del método

Aplicaciones en:

1. Sumas de variables aleatorias (Stein, Chen, Barbour).
2. Análisis estocástico. Teorema de Breuer-Major, relación con caos de Wiener, estadísticos de procesos Gaussianos (Nualart, Peccati, Nourdin, Podolsjik, Tudor, Jaramillo).
3. Teoría de números probabilista. Funciones aritméticas aditivas, multiplicidades primas de enteros aleatorios (Chen, Jaramillo, Yang, Harper).
4. Matrices aleatorias. Fluctuaciones de segundo orden del espectro de matrices aleatorias gaussianas. (Jaramillo, Pardo, Díaz, Pérez-Abreu).

Espectro de la utilidad del método

Aplicaciones en:

1. Sumas de variables aleatorias (Stein, Chen, Barbour).
2. Análisis estocástico. Teorema de Breuer-Major, relación con caos de Wiener, estadísticos de procesos Gaussianos (Nualart, Peccati, Nourdin, Podolsjik, Tudor, Jaramillo).
3. Teoría de números probabilista. Funciones aritméticas aditivas, multiplicidades primas de enteros aleatorios (Chen, Jaramillo, Yang, Harper).
4. Matrices aleatorias. Fluctuaciones de segundo orden del espectro de matrices aleatorias gaussianas. (Jaramillo, Pardo, Díaz, Pérez-Abreu).
5. Probabilidad no conmutativa. variables aleatorias infinitamente divisibles (Arizmendi, Gaxiola, Jaramillo).

Espectro de la utilidad del método

Más aplicaciones en:

1. Geometría estocástica (Gunter Last, Peccati).
2. Gráficas aleatorias (Röllin, Kaur, Arizmendi, Salazar, Arenas).
3. Eficiencia de algoritmos (Goldstein, Bhattacharjee).
4. Machine learning (Qiang Liu).

Ejemplos muy concretos

Las variables S_n no necesariamente deben ser sumas de otras variables aleatorias más sencillas.

Ejemplos muy concretos

Las variables S_n no necesariamente deben ser sumas de otras variables aleatorias más sencillas.

Theorem (Chen, Jaramillo, Yang)
Si S_n es el número de factores primos en una muestra uniforme en $\{1, \ldots, n\}$,

Ejemplos muy concretos

Las variables S_n no necesariamente deben ser sumas de otras variables aleatorias más sencillas.

Theorem (Chen, Jaramillo, Yang)
Si S_n es el número de factores primos en una muestra uniforme en $\{1, \ldots, n\}$, entonces

$$d_K\left(\frac{S_n - \log\log(n)}{\sqrt{\log\log(n)}}, \mathcal{N}\right) \leq \frac{C}{\sqrt{\log\log(n)}}$$

Ejemplos muy concretos

Theorem (Diaz, Jaramillo, Pardo)
Sea $X^{(n)}(t) \in \mathbb{R}^{n\times n}$ *un proceso matricial simétrico cuyas entradas* $X_{i,j}^{(n)} = \{X_{i,j}^{(n)}(t) \ ; \ t \geq 0\}$*, para* $i \leq j$ *son procesos Gaussianos i.i.d. Si* $\lambda_1^n(t) \leq \cdots \lambda_n^n(t)$ *denotan a los eigenvalores ordenados de* $X_t^{(n)}$ *y* μ_t^n *es la distribución empírica asociada, entonces para toda función de prueba* $f : \mathbb{R} \to \mathbb{R}$*,*

$$n\left(\int_{\mathbb{R}} f(x)\mu_t^n(dx) - \mathbb{E}[\int_{\mathbb{R}} f(x)\mu_t^n(dx)]\right) \ ; \ t \geq 0)$$

converge a un proceso Gaussiano centrado con covarianza explícita.

Ejemplos muy concretos

Theorem (Arizmendi, Jaramillo)
Si S_n son variables infinitamente divisibles, el método de Stein se puede usar para probar que si S_n son estándar,

$$d_K(S_n, \mathcal{N}) \leq C\sqrt{|\mathbb{E}[S_n^4] - \mathbb{E}[\mathcal{N}^4]|}.$$

Espectro de la utilidad del método

Se puede desarrollar método de Stein para muchas otras distribuciones, no solo la gaussiana

1. Poisson
2. Poisson compuesta
3. Exponencial
4. Gamma
5. Uniforme
6. Beta
7. Semicirculo
8. Arco seno
9. Dickman

Gracias!

References

Chen L., Jaramillo A., Yang X. A probabilistic approach to the Erdös-Kac theorem for additive functions. Soon in Arxiv.