Test de pruebas de hipótesis via discrepancia de Stein kernelizada

Arturo Jaramillo

October 16, 2024

Problemática

Contexto:
Proponemos un modelo q para un conjunto de datos en $\mathbb{R}^d$.

Pregunta central
Los datos realmente se ajustan a q?

Dificultad:
Sólo puedo acceder a la función score $\mathbf{s}_q(x) := \nabla V(x)$, donde

$$q(x) = \exp\{V(x)\}.$$

Relevancia:
Útil cuando sólo se conoce a q salvo múltiplos constantes.

Propuesta metodológica

- Discrepancia de Stein kernelizada (KSD).
- Debe permitir buenos ajustes para distribuciones complejas.
- Funciona sin acceso completo a las verosimilitudes.

Ingredientes teóricos: Identidad de Stein

Sea $\mathcal{F}$ una familia de funciones reales en $\mathbb{R}^d$.

Theorem

Dos distribuciones p y q son iguales si para toda $f \in \mathcal{F}$,

$$\mathbb{E}_{x \sim p} \left[\mathcal{A}_q[f](x) \right] = 0, \tag{1}$$

donde $\mathcal{A}_q[f](x) := \boldsymbol{s}_q(x) f(x) + \nabla f(x)$.

Ingredientes teóricos: Identidad de Stein

Sea $\mathcal{F}$ una familia de funciones reales en $\mathbb{R}^d$.

Theorem

Dos distribuciones p y q son iguales si para toda $f \in \mathcal{F}$,

$$\mathbb{E}_{x\sim p}\left[\mathcal{A}_q[f](x)\right] = 0, \tag{1}$$

donde $\mathcal{A}_q[f](x) := \boldsymbol{s}_q(x)f(x) + \nabla f(x)$.

Definition

La **discrepancia de Stein** se define como

$$\mathbb{S}(p, q) := \sup_{f\in\mathcal{F}}(\mathbb{E}_{x\sim p}[\mathcal{A}_q[f](x)])^2$$

Cómo escogemos a $\mathcal{F}$?

Ingrediente inicial: un kernel $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ positivo definido.

Theorem

Existe un espacio de funciones $\mathcal{H}$ con producto interno $\langle \cdot, \cdot \rangle_{\mathcal{H}}$, tal que $k(\cdot, x) \in \mathcal{H}$ para $x \in \mathbb{R}^d$, y

$$f(x) = \langle f(\cdot), k(\cdot, x) \rangle_{\mathcal{H}},$$

para toda $f \in \mathcal{H}$.

Cómo escogemos a $\mathcal{F}$?

Ingrediente inicial: un kernel $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ positivo definido.

Theorem

Existe un espacio de funciones $\mathcal{H}$ con producto interno $\langle \cdot, \cdot \rangle_{\mathcal{H}}$, tal que $k(\cdot, x) \in \mathcal{H}$ para $x \in \mathbb{R}^d$, y

$$f(x) = \langle f(\cdot), k(\cdot, x) \rangle_{\mathcal{H}},$$

para toda $f \in \mathcal{H}$.

Idea

Proponer $\mathcal{F}$ como el disco unitario de $\mathcal{H}$ en la discrepancia de Stein.

Otras formas de la identidad de Stein

Si $\boldsymbol{f}$ es una función con valores en $\mathbb{R}^{d'}$, definimos $\mathcal{A}_q[\boldsymbol{f}](x)$ como la matriz $d \times d'$, dada por

$$\mathcal{A}_q[\boldsymbol{f}](x) := \mathbf{s}_q(x)\mathbf{f}(x)^T + \nabla\mathbf{f}(x).$$

Otras formas de la identidad de Stein

Si $\boldsymbol{f}$ es una función con valores en $\mathbb{R}^{d'}$, definimos $\mathcal{A}_q[\boldsymbol{f}](x)$ como la matriz $d \times d'$, dada por

$$\mathcal{A}_q[\boldsymbol{f}](x) := \mathbf{s}_q(x)\mathbf{f}(x)^T + \nabla \mathbf{f}(x).$$

Lemma (Ley, Swan, 2013)

Sean p y q densidades ciertas condiciones de regularidad. Entonces

$$\mathbb{E}_{x\sim p}[\mathcal{A}_q[f](x)] = \mathbb{E}_p[(\mathbf{s}_q(x) - \mathbf{s}_p(x))\mathbf{f}(x)^T].$$

En particular, si $d' = d$, se tiene

$$\mathbb{E}_{x\sim p}[Tr(\mathcal{A}_q[f](x))] = \mathbb{E}_p[(\mathbf{s}_q(x) - \mathbf{s}_p(x)) \cdot \mathbf{f}(x)].$$

Discrepancia de Stein kernelizada

Bajo las consideraciones previas

Definition (Discrepancia de Stein kernelizada, KSD)

Está definida por

$$\mathbb{S}_K(p, q) := \mathbb{E}_{x,x'\sim p}[(\boldsymbol{s}_q(x) - \boldsymbol{s}_p(x))K(x, x')(\boldsymbol{s}_q(x') - \boldsymbol{s}_p(x'))]$$

Discrepancia de Stein kernelizada

Bajo las consideraciones previas

Definition (Discrepancia de Stein kernelizada, KSD)

Está definida por

$$\mathbb{S}_K(p, q) := \mathbb{E}_{x,x'\sim p}[(\boldsymbol{s}_q(x) - \boldsymbol{s}_p(x))K(x, x')(\boldsymbol{s}_q(x') - \boldsymbol{s}_p(x'))]$$

Ok... definimos algo pero no parece servir de nada..

Representación de KSD

Theorem

Se tiene que

$$\mathbb{S}_K(p, q) := \mathbb{E}_{x, x' \sim p}[u_q(x, x')],$$

donde u_q *es una función completamente explícita que sólo depende de* $\boldsymbol{s}_q$ *y de* K*.*

Representación de KSD

Theorem

Se tiene que

$$\mathbb{S}_K(p, q) := \mathbb{E}_{x,x' \sim p}[u_q(x, x')],$$

donde u_q es una función completamente explícita que sólo depende de $\boldsymbol{s}_q$ y de K.

Estimador

$$\hat{\mathbb{S}}_K(p, q) := \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} u_q(x_i, x_j),$$

donde $x_1, \ldots, x_n$ son muestreadas de p.

Ok, ok, si insisten...

La función u_q está dada por

$$\begin{aligned} u_q(x, x') := {} & \mathbf{s}_q(x)^T k(x, x') \mathbf{s}_q(x') \\ & + \mathbf{s}_q(x)^T \nabla_{x'} k(x, x') \\ & + \nabla_x k(x, x')^T \mathbf{s}_q(x') \\ & + Tr[\nabla_{x,x'} K(x, x')] \end{aligned}$$

Cómo funciona la prueba de hipótesis?

Fundamento teórico

Se puede mostrar que bajo la hipótesis Datos $\sim q$, se tiene que $n\hat{\mathbb{S}}_K(p,q)$ converge a un límite con cuantiles calculables numéricamente en términos de $\mathbf{s}_q$.

Cómo funciona la prueba de hipótesis?

Fundamento teórico
Se puede mostrar que bajo la hipótesis Datos $\sim q$, se tiene que $n\hat{\mathbb{S}}_K(p, q)$ converge a un límite con cuantiles calculables numéricamente en términos de $\mathbf{s}_q$. **Algoritmo**

- Calculamos un $(1-\alpha)$-cuantil de la ley asintótica antes mencionada.
- Calculamos $n\hat{\mathbb{S}}_K(p, q)$ y comparamos con el cuantil. Rechazamos si es mayor.

Comparación con métodos clásicos

- **Alternativas:** Pruebas (χ^2) y Kolmogorov-Smirnov test.
- Limitaciones: la constante normalizadora :-(
- KSD funciona bien en éstas situaciones.

Bibliography

Liu, Q. and Lee, J. D. (2016).
A kernelized Stein discrepancy for goodness-of-fit tests.
Advances in Neural Information Processing Systems (NeurIPS), 29.

Chwialkowski, K., Strathmann, H., and Gretton, A. (2016).
A kernel test of goodness of fit.
In Proceedings of the 33rd International Conference on Machine Learning (ICML).

Gracias!

Preguntas?