Distribuciones de Máxima Entropía en Bolas de Wasserstein

Luis Felipe Vargas Beltrán

Universidad de Los Andes

28 de Mayo, 2019

Contenido

1. Motivación y Enunciado.
   - Distribuciones de Máxima Entropía
   - El Problema
   - Segunda Motivación: Cotas Superiores Para la Entropía
2. Análisis del Problema
   - Distancia de Wasserstein
   - Reducción del Problema
   - Generalización
3. Comentarios Finales

Distribuciones de Máxima Entropía

Si no tenemos ninguna información acerca de una variable aleatoria ¿Cuál debemos considerar?

Distribuciones de Máxima Entropía

Si no tenemos ninguna información acerca de una variable aleatoria ¿Cuál debemos considerar?
La Uniforme

Distribuciones de Máxima Entropía

Si no tenemos ninguna información acerca de una variable aleatoria ¿Cuál debemos considerar?
La Uniforme

Si sabemos que $f \in \mathcal{F}$, ¿Cuál distribución debemos considerar?

Distribuciones de Máxima Entropía

Si no tenemos ninguna información acerca de una variable aleatoria ¿Cuál debemos considerar?
La Uniforme

Si sabemos que $f \in \mathcal{F}$, ¿Cuál distribución debemos considerar?
La Distribución de Máxima Entropía

Distribuciones de Máxima Entropía

Si no tenemos ninguna información acerca de una variable aleatoria ¿Cuál debemos considerar?
La Uniforme

Si sabemos que $f \in \mathcal{F}$, ¿Cuál distribución debemos considerar?
La Distribución de Máxima Entropía

- Ejemplo: Dada la varianza σ^2 y la media μ, la distribución que maximiza la entropía es la normal $N(\mu, \sigma^2)$.

Definición

Para una distribución de probabilidad discreta p de un conjunto enumerable $\{x_1, x_2, \dots\}$ con $p_i = p(x_i)$ la *entropía* está definida por

$$h(p) = -\sum_{i \geq 0} p_i log(p_i)$$

Y para una función de distribución continua $f(x)$ la *entropía* es

$$h(f) = -\iint_R f(x) log(f(x)) dA$$

Donde R es la región donde está definida f.

El Problema

Distribuciones de Máxima Entropía en Bolas de Wasserstein

$$\max_{f} \iint_{R} -f(x)log(f(x))dA \qquad st$$

$$f \in P$$

$$\iint_{R} f(x)dA = 1$$

$$f(x) \geq 0 \quad \forall x \in R$$

Donde P lo vamos a llamar el conjunto de ambigüedad.

El Problema

Distribuciones de Máxima Entropía en Bolas de Wasserstein

Sean $x_1, x_2, \ldots, x_n \in \mathbb{R}^2$. Sea $\tilde{f}$ la distribución empírica dada por los puntos $x_1, \ldots, x_n$, cada punto con probabilidad $\frac{1}{n}$. Vamos a considerar el siguiente problema

$$\max_{f \in L^1(R)} \iint_R -f(x) log(f(x)) dA \qquad st$$

$$\mathcal{D}(f, \tilde{f}) \leq t$$

$$\iint_R f(x) dA = 1$$

$$f(x) \geq 0 \quad \forall x \in R$$

Donde $\mathcal{D}$ denota la distancia de Wasserstein.

Segunda Motivación

Cotas Superiores Para la Entropía

Suponga que $x_1, x_2, \ldots, x_N$ son un muestreo de una distribución f

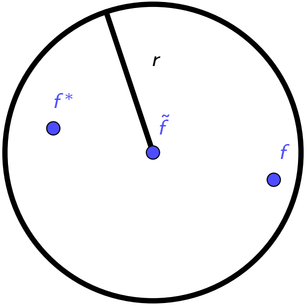
- Se puede hallar $r = r(N)$ tal que $f \in B_r(\tilde{f})$ con probabilidad 1.

Segunda Motivación

Cotas Superiores Para la Entropía

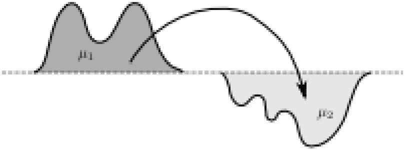
Suponga que $x_1, x_2, \ldots, x_N$ son un muestreo de una distribución f

- Se puede hallar $r = r(N)$ tal que $f \in B_r(\tilde{f})$ con probabilidad 1.
- Sea f^* la solución al problema anterior con $t = r$.
- Por lo tanto $Entr(f^*) \geq Entr(f)$ con probabilidad 1.



Distancia de Wasserstein

Si pensamos en dos distribuciones μ_1 y μ_2 como dos montañas de arena entonces $D(\mu_1, \mu_2)$ es el trabajo necesario para convertir una montaña en la otra



Distancia de Wasserstein Respecto a una Empírica

Teorema

Sean $x_1, x_2, \ldots, x_n \in \mathbb{R}^2$ *y* $\tilde{f}$ *la distribución empírica centrada en estos puntos cada uno con probabilidad* $\frac{1}{n}$ *y sea* f *una función de distribución en un compacto* $R \in \mathbb{R}^2$*. Entonces*

- $\mathcal{D}(f, \tilde{f})$ *es el óptimo al problema*

$$\max_{\lambda \in \mathbb{R}^n} \iint\limits_R f(x) \min_i \{\|x - x_i\| - \lambda_i\} dA \quad s.a$$

$$\mathbf{e}^t \lambda = 0$$

- *Si* $f(x) > 0 \ \forall x \in R$*. Entonces el optimizador* λ^* *es único.*

Distancia de Wasserstein Respecto a una Empírica

Si λ^* es un maximizador al problema, entonces un transporte oṕtimo viene dado por las regiones

$$R_i^* = \{x \in R : \|x - x_i\| - \lambda_i^* < \|x - x_j\| - \lambda_j^* \quad \forall j \neq i\}$$

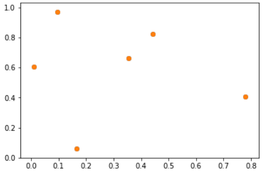


Figura: 6 datos aleatorios

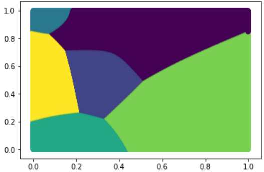


Figura: Regiones dadas por $\lambda = (0{,}15, 0{,}05, -0{,}07, -0{,}15, 0{,}085, 0{,}035)$

Distancia de Wasserstein Respecto a una Empírica

Si λ^* es un maximizador al problema, entonces un transporte óptimo viene dado por las regiones

$$R_i^* = \{x \in R : \|x - x_i\| - \lambda_i^* < \|x - x_j\| - \lambda_j^* \quad \forall j \neq i\}$$

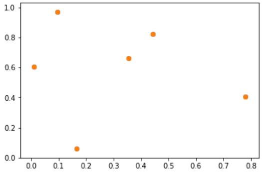


Figura: 6 datos aleatorios

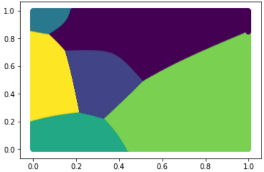
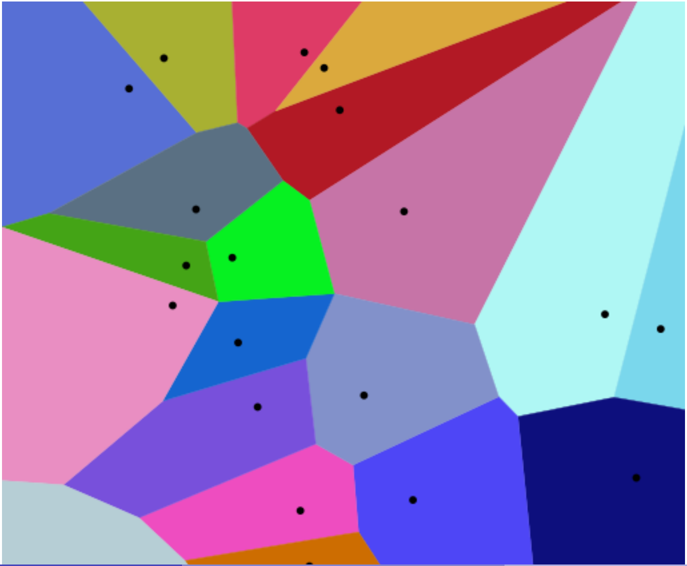


Figura: Regiones dadas por $\lambda = (0{,}15, 0{,}05, -0{,}07, -0{,}15, 0{,}085, 0{,}035)$

Note que si $\lambda = (0, \ldots, 0)$ obtenemos los diagramas de Voronoi

Diagramas de Voronoi

$$R_i = \{x \in R : \|x - x_i\| \leq \|x - x_j\| \, \forall j \neq i\}$$



Reformulación del Problema

$$\max_{f \in L^1(R)} \iint_R -f(x) log(f(x)) dA \qquad st$$

$$\iint_R f(x) \min_i \{\|x - x_i\| - \lambda_i\} dA \leq t \quad \forall \lambda : e^t \lambda = 0$$

$$\iint_R f(x) dA = 1$$

$$f(x) \geq 0 \quad \forall x \in R$$

Reformulación del Problema

$$\max_{f \in L^1(R)} \iint\limits_R -f(x) log(f(x)) dA \qquad st$$

$$\iint\limits_R f(x) \min_i \{\|x - x_i\| - \lambda_i\} dA \leq t \quad \forall \lambda : e^t \lambda = 0$$

$$\iint\limits_R f(x) dA = 1$$

$$f(x) \geq 0 \quad \forall x \in R$$

- Este problema tiene infinitas restricciones
- La meta va ser reducir el problema a sólo una restricción

Análisis del Problema

Teorema

El problema descrito tiene solución única f^* que satisface $f^*(x) > 0$ $\forall x \in R$

Demostración.

- Unicidad: Por la concavidad estricta de la función
- Es posible construir la solución y ver que es de la forma

$$f^*(x) = e^{-1-v_0^*\|x-x_i\|-v_i^*}$$

En cada región R_i. Donde $\cup R_i = R$ es una partición de R.

$\square$

Análisis del Problema

Teorema (Velasco, Vargas)

La distribución de máxima entropía en la bola $B_t(\tilde{f})$ toma la forma

$$f^*(x) = e^{-1-u^* \min_i\{\|x-x_i\|-\lambda_i^*\}-v^*}$$

Demostración.

El óptimo del problema es el mismo que el del problema

$$\max_{f\in L^1(R)} \iint_R -f(x)log(f(x))dA \qquad st$$

$$\iint_R f(x)\min_i\{\|x-x_i\|-\lambda_i^*\}dA \leq t$$

$$\iint_R f(x)dA = 1$$

$$f(x) \geq 0 \quad \forall x \in R$$

Solución del Problema

La solución es

$$f^*(x) = e^{-1-u^* \min_i \|x-x_i\|-\lambda_i^*\}-v^*}$$

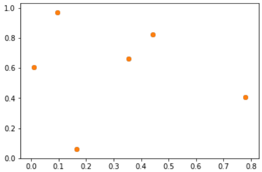
Donde u^* y v^* son la solución al problema

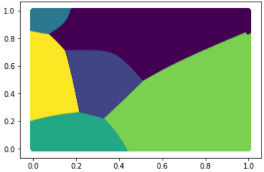
$$\min_{u,v} \iint_R \frac{1}{e^{1+u \min_i\{\|x-x_i\|-\lambda_i^*\}+v}} dA + ut + v \qquad s.a \qquad (**)$$

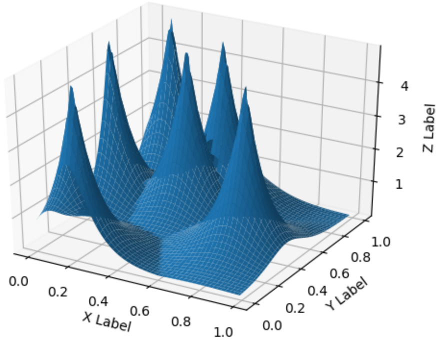
$$u \geq 0$$

- Si se conoce λ^* sabemos cómo hallar u^* y v^*

Resultados


1.0
0.8
0.6
0.4
0.2
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8


1.0
0.8
0.6
0.4
0.2
0.0
0.0 0.2 0.4 0.6 0.8 1.0


4
3
2
1
Z Label
0.0 0.2 0.4 0.6 0.8 1.0
X Label
0.0 0.2 0.4 0.6 0.8 1.0
Y Label

Generalización

Teorema (Velasco, Vargas)

Sea $\mathcal{G}$ una función estrictamente cóncava tal que la función $g(x) = \mathcal{G}(x) - ax$ se maximiza en un único punto $x_a > 0$. El problema

$$\max_f \iint_R \mathcal{G}(f(x))dA \qquad s.a$$

$$D(f, \tilde{f}) \leq t$$

$$\iint_R f(x)dA = 1$$

$$f(x) \geq 0 \quad \forall x \in R$$

Tiene solución única f^* con $f^*(x) > 0$.

- Por lo tanto existe un algoritmo de Cutting-Planes semejante al de máxima entropía para solucionar el problema

Comentarios Finales

- El algoritmo de Fortune Generalizado permite hallar los diagramas descritos en el transporte de Wasserstein en tiempo $o(nlog(n))$
- Los Funcionales Euclidianos Monótonos Subaditivos asintóticamente satisfacen las condiciones de generalización.

¿Preguntas?