Optimización bajo Incertidumbre
B. Revisión

Carlos Testuri – Germán Ferrari

Departamento de Investigación Operativa – Instituto de Computación
Facultad de Ingeniería – Universidad de la República

2003-2022

Contenido

Espacio de probabilidad

Espacio de probabilidad: $(\Omega, \mathcal{A}, P)$

- Espacio muestral: conjunto Ω, de resultados posibles $\omega \in \Omega$ de un experimento aleatorio.
- σ-álgebra en conjunto Ω: clase $\mathcal{A}$ de subconjuntos $A \in \mathcal{A}$ (denominados eventos) del conjunto Ω.
- Medida de probabilidad en $\mathcal{A}$: $P(A) : \mathcal{A} \to [0, 1]$.

Un espacio de probabilidad se dice que es:

- *finito*, si su espacio muestral es finito,
- *discreto*, si su espacio muestral es finito o infinito numerable,
- *general*, si su espacio muestral no es numerable.

Variable aleatoria

Se define *variable aleatoria* en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$ como la función

$$\xi : \Omega \to \mathbb{R}.$$

Dado que el conjunto $\xi^{-1}(r)$ (elementos de Ω cuya imagen por ξ es el real r) es un elemento de $\mathcal{A}$, entonces este tiene una probabilidad asociada $P(\xi^{-1}(r)) = P(\xi = r)$.

Función de [distribución de] probabilidad

Según sea la función de probabilidad:

Variable discreta $\boldsymbol{\xi}$ con realizaciones $\xi^k, k \in K$,

$$f(\xi^k) = P(\boldsymbol{\xi} = \xi^k).$$

Variable continua $\boldsymbol{\xi}$ se define como

$$F_{\xi}(x) = P(\boldsymbol{\xi} \leq x) = P(\{\omega | \boldsymbol{\xi}(\omega) \leq x\}).$$

$$P(\boldsymbol{\xi} \leq x) = \int_{-\infty}^{x} f(\xi) \mathrm{d}\xi = \int_{-\infty}^{x} \mathrm{d}F(\xi),$$

$$\text{donde} f(\xi) = \frac{\mathrm{d}F(\xi)}{\mathrm{d}\xi} \text{ es la función de densidad.}$$

Eventos casi seguros

Sea el espacio de probabilidad $(\Omega, \mathcal{A}, P)$, el evento A ocurre *casi seguramente* (abreviado “a.s.") si $P(A) = 1$.

Casi-seguramente vs. seguramente
Un evento ocurre *seguramente* si siempre ocurre; es decir, ningún otro evento (aún eventos con probabilidad cero) puede ocurrir.
Si un evento ocurre *casi seguramente*, entonces otros eventos con probabilidad cero pueden ocurrir (casi nunca).

Momentos de una variable aleatoria

El r-ésimo *momento* de la variable ξ es

$$\mu_r' = \mathbb{E}[\xi^r] = \int_{-\infty}^{\infty} \xi^r f(\xi) \mathrm{d}\xi$$

donde $\quad \mu_1' = \mu = \mathbb{E}[\xi] \quad$ se denomina *media*.

El r-ésimo *momento central* de la variable ξ es

$$\mu_r = \mathbb{E}[(\xi - \mu)^r] = \int_{-\infty}^{\infty} (\xi - \mu)^r f(\xi) \mathrm{d}\xi$$

donde $\quad \mu_2 = \mathbb{E}[(\xi - \mu)^2] = \mathrm{Var}(\xi) \quad$ se denomina *varianza*.

Propiedades de esperanza y varianza

Sean a, b y c constantes, ξ, ξ_1 y ξ_2 variables aleatorias

- Invariante esperanza: $\mathbb{E}[c] = c$
- Linealidad: $\mathbb{E}[a\xi_1 + b\xi_2] = a\mathbb{E}[\xi_1] + b\mathbb{E}[\xi_2]$
- Invariante varianza: $\mathrm{Var}(a\xi + b) = a^2\mathrm{Var}(\xi)$
- Vínculo esperanza y varianza: $\mathrm{Var}(\xi) = \mathbb{E}[\xi^2] - (\mathbb{E}[\xi])^2$
- No linealidad: $\mathrm{Var}(\xi_1 + \xi_2) = \mathrm{Var}(\xi_1) + \mathrm{Var}(\xi_2) + 2cov(\xi_1, \xi_2)$;
  $cov(\xi_1, \xi_2) = \mathbb{E}[(\xi_1 - \mathbb{E}[\xi_1])(\xi_2 - \mathbb{E}[\xi_2])] = \mathbb{E}[\xi_1\xi_2] - \mathbb{E}[\xi_1]\mathbb{E}[\xi_2]$
- ¿Cuándo se dice que ξ_1 y ξ_2 son independientes?
  En caso de serlo su covarianza es cero y $\mathbb{E}[\xi_1\xi_2] = \mathbb{E}[\xi_1]\mathbb{E}[\xi_2]$

Teorema Central del Límite

Si las variables aleatorias X_1, ..., X_n son independientes y tienen la misma distribución (iid) con esperanza μ finita y varianza σ^2 finita, no nula, entonces la variable aleatoria

$$Z_n = \frac{X_1 + ... + X_n - n\mu}{\sigma\sqrt{n}}$$

es asintóticamente normal, es decir su función de distribución $F_n(z)$ cumple para todo z

$$\lim_{n\to\infty} F_n(z) = \lim_{n\to\infty} P\left(\frac{X_1 + ... + X_n - n\mu}{\sigma\sqrt{n}} \leq z\right) = \mathcal{N}(0, 1).$$

Producto interno [escalar]

El producto interno de los vectores $v, w \in \mathbb{R}^n$ tiene como resultado el escalar

$$v \cdot w = v_1 w_1 + ... + v_n w_n = \sum_{i=1}^{n} v_i w_i = v^\tau w.$$

Para el caso del espacio Euclidiano, el producto interno de los vectores es

$$v^\tau w = \|v\| \|w\| \cos(\theta),$$

donde $\| \cdot \|$ establece la norma de los vectores y θ el ángulo entre ellos.

¿Cuándo vale cero el producto interno de dos vectores?

¿Cuándo es positivo o negativo el producto?

Producto interno y ecuaciones (1/2)

Dado el vector constante $a \in \mathbb{R}^n$.
¿Qué región del espacio describe el vector variable $x \in \mathbb{R}^n$ tal que

1. $a^\tau x = 0$?

2. $a^\tau x \geq 0$?

3. $a^\tau x \leq 0$?

Producto interno y ecuaciones (2/2)

Dados los vectores constantes $a, d \in \mathbb{R}^n$.
¿Qué región del espacio describe el vector variable $x \in \mathbb{R}^n$ tal que

$$a^\tau (x + d) = 0\,?$$

Equivalentemente:

$$a^\tau x = -a^\tau d$$

Sea $b := -a^\tau d$,

$$a^\tau x = b$$

Hiperplano y semiespacio

Dados el vector $a \in \mathbb{R}^n$, el escalar $b \in \mathbb{R}$ y el vector variable $x \in \mathbb{R}^n$ se define el *hiperplano afín*

$$a^\tau x = \sum_{i=1}^{n} a_i x_i = b,$$

y los *semiespacios afines*

$$\begin{aligned} a^\tau x = \textstyle\sum_{i=1}^{n} a_i x_i \leq b, \\ a^\tau x = \textstyle\sum_{i=1}^{n} a_i x_i \geq b. \end{aligned}$$

Conjunto convexo

Sea C un subconjunto de $\mathbb{R}^n$. Se dice que C es *convexo* si para todo $x, y \in C$, y para todo $\lambda \in [0, 1]$, se cumple

$$\lambda x + (1 - \lambda) y \in C.$$

Combinación convexa y Casco convexo

Sean $x_1, ..., x_k$ vectores de $\mathbb{R}^n$, y $\lambda_1, ..., \lambda_k$ escalares no negativos cuya suma es uno,

- se denomina *combinación convexa* de $x_1, ..., x_k$ al vector $\sum_{i=1}^{k} \lambda_i x_i$
- se denomina *casco convexo* de los vectores de $x_1, ..., x_k$, al conjunto de todas sus combinaciones convexas. Dicho conjunto se denota $conv(\{x_1, ..., x_k\})$.

Convexidad: propiedades

- La intersección de conjuntos convexos es convexa.
- El conjunto $P = \{x \in \mathbb{R}^n : Ax \leq b\}$, denominado *poliedro*, es un conjunto convexo.
- La combinación convexa de un número finito de elementos de un conjunto convexo pertenece al conjunto.
- El casco convexo de un número finito de vectores es un conjunto convexo.

Función convexa

Sea C subconjunto convexo de $\mathbb{R}^n$. Se dice que la función $f : C \to \mathbb{R}$ es *convexa* si para todo $x, y \in C$, y para todo $\lambda \in [0, 1]$, se cumple

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).$$

Desigualdad de Jensen

Establece una cota de una función convexa de una integral con la integral de la función convexa.

Ejemplo: la transformación convexa de la media es menor o igual a la media de la transformación convexa.
Sean la variable aleatoria ξ y la función convexa $f(\cdot)$ se cumple

$$f(\mathbb{E}[\xi]) \leq \mathbb{E}[f(\xi)].$$

Optimización Lineal

Caso funciones lineales en términos de variables continuas.
Dados los parámetros: matriz $A \in \mathbb{R}^{m \times n}$, y los vectores columnas $b \in \mathbb{R}^m$ y $c \in \mathbb{R}^n$ y la variable de decisión: vector columna $x \in \mathbb{R}^n$,

$$\begin{array}{cl} \min & c^\tau x \\ \text{s.a} & Ax = b \\ & x \geq 0. \end{array}$$

$$\text{Donde } x^* \text{ es} \begin{cases} \text{solución,} & \text{si } Ax^* = b \\ \text{solución factible,} & \text{si es solución y } x^* \geq 0 \\ \text{solución óptima,} & \text{si es factible y } c^\tau x^* \leq c^\tau x, \ \forall x \text{ factible.} \end{cases}$$

Problemas *factible*, *no factible* y *no acotado*.

Optimización Lineal: Equivalencias y reducciones en formulación

Equivalencias

Sea la fila i-ésima de la matriz A, a_i,

1. la restricción de igualdad $a_i^\tau x = b_i$ es equivalente a las restricciones $a_i^\tau x \leq b_i$ y $a_i^\tau x \geq b_i$,
2. la restricción $a_i^\tau x \leq b_i$ es equivalente a $-a_i^\tau x \geq -b_i$,
3. el objetivo $\min c^\tau x$ es equivalente al objetivo $-\max -c^\tau x$.

Reducciones

1. *eliminación de variable libre*: la variable x_j no restringida en signo se reemplaza por la diferencia de dos variables no negativas, $x_j^+ - x_j^-$.
2. *eliminación de restricción de desigualdad*: dada $\sum_{j=1}^n a_{ij}x_j \leq b_i$, se agrega una nueva variable (de *holgura*), $s_i \geq 0$, tal que $\sum_{j=1}^n a_{ij}x_j + s_i = b_i$. Similarmente, para la otra desigualdad, se agrega una variable de *exceso* que se resta a la restricción.

Optimización Lineal: condiciones de optimalidad

En el poliedro $P = \{x : Ax = b, x \geq 0\}$ las soluciones están determinadas por los puntos extremos y rayos extremos.

Particionando $A = [B, N]$, donde B es base; además, $x = [x_B, x_N]$ y $c = [c_B, c_N]$ se tiene la solución básica $x_B = B^{-1}b, x_N = 0$, con valor $c_B^\tau B^{-1}b$.

Condición de *factibilidad*, dada por $x_B = B^{-1}b \geq 0$
Condición de *optimalidad*, dada por $c_N^\tau - c_B^\tau B^{-1}N \geq 0$.

Optimización Lineal: Dualidad

Propiedad algebraica que establece problemas equivalentes:

$$
\begin{array}{ll}
\multicolumn{2}{l}{\textit{Primal}} \\
\hline
\min & c^{\tau} x \\
\text{s.a} & Ax = b \\
 & x \geq 0.
\end{array}
\quad \rightarrow \quad
\begin{array}{ll}
\multicolumn{2}{l}{\textit{Dual}} \\
\hline
\max & \pi^{\tau} b \\
\text{s.a} & \pi^{\tau} A \leq c^{\tau}.
\end{array}
$$

Optimización Lineal: Propiedades de dualidad

- *Dualidad débil*
  Sean $\hat{x}$ primal factible y $\hat{\pi}$ dual factible $\Rightarrow c^{\tau}\hat{x} \geq \hat{\pi}^{\tau}b$
- *Dualidad fuerte*
  Existe x^* sol.óptima $\Leftrightarrow$ existe π^* sol.óptima, con $c^{\tau}x^* = (\pi^*)^{\tau}b$.

Optimización Lineal: relación valor objetivo entre primal y dual

Primal \ Dual	Finito	No-acotado	No-factible
Finito	$\surd$	$\times$	$\times$
No-acotado	$\times$	$\times$	$\surd$
No-factible	$\times$	$\surd$	$\surd$

Algoritmos

Los problemas de optimización se resuelven mediante algoritmos.
Una *instancia* de un problema se define a partir de datos que la especifican.
El problema comprende la colección de todas sus instancias.

Un *algoritmo* de resolución es un conjunto finito y ordenado de instrucciones que se diseña para un problema pero se aplica a sus instancias.
Los algoritmos emplean los recursos tiempo y espacio (memoria). El uso de los recursos depende de la dificultad y tamaño de las instancias.

Optimización Lineal: Algoritmo simplex (resumen)

El método itera de una solución (punto extremo) a otra reduciendo el valor de la función objetivo hasta que llega a la solución óptima.

La codificación computacional del método pone especial énfasis en el cálculo de la inversa de la base y en evitar quedar atrapado en secuencias de iteración con cambio de base sin reducción de valor en la función objetivo, debido a la presencia de *degeneración*.

La complejidad computacional teórica (peor caso) es exponencial en el número de variables y restricciones.
En la práctica es muy eficiente, en promedio la complejidad es proporcional al número de restricciones.

Optimización Lineal: Algoritmo simplex

Iteración de algoritmo simplex

1. Obtener la base $B := A_{B(1)}, \ldots, A_{B(m)}$, calcular su inversa B^{-1} y su solución básica $x_B := B^{-1}b$.
2. Calcular el valor dual $\pi^\tau := c_B^\tau B^{-1}$ y los costos reducidos $\widehat{c}_j := c_j - \pi^\tau A_j$, para todo j. Si todo $\widehat{c}_j \geq 0$, la solución es óptima, *Parar*; sino, seleccionar algún j para el que $\widehat{c}_j < 0$.
3. Calcular $u := B^{-1}A_j$. Si ningún componente de u es positivo, el valor óptimo es $-\infty$, el problema es no acotado, *Parar*.
4. Si alguno de los componentes de u es positivo, calcular $\theta^* := \min_{\{i=1,\ldots,m: u_i>0\}} x_{B(i)}/u_i$.
5. Sea k el índice donde $\theta^* = x_{B(k)}/u_k$. Construir la nueva base reemplazando $A_{B(k)}$ con A_j. Ir al paso 1.

Optimización No Lineal

Caso funciones no lineales de variables continuas.
Dadas la variable de decisión $x \in \mathbb{R}^n$ y las funciones $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ y $h : \mathbb{R}^n \to \mathbb{R}^l$. Se asume que $f(x)$ es convexa y que la región factible es cerrada y convexa,

$$\begin{array}{ll} \min & f(x) \\ \text{s.a} & g_j(x) \leq 0, \quad j = 1, ..., m, \\ & h_i(x) = 0, \quad i = 1, ..., l. \end{array}$$

Si existen derivadas direccionales de $f(x)$, entonces es diferenciable, con gradiente $\nabla f(x)$; en caso contrario, se establecen subgradientes, cuyo conjunto es el subdiferencial $\partial f(x)$.

Optimización No Lineal: condiciones de optimalidad

Condición necesaria: Karush-Kuhn-Tucker (KKT)

Sea x^* mínimo local regular, donde f, g y h son diferenciables continuas. Entonces existen multiplicadores de Lagrange únicos λ^* y μ^* tales que

$$\nabla f(x^*) + \sum_{j=1}^{m} \mu_j^* \nabla g_j(x^*) + \sum_{i=1}^{l} \lambda_i^* \nabla h_i(x^*) = 0$$

$$\mu_j^* g_j(x^*) = 0, \quad j = 1, ..., m$$

$$\mu_j^* \geq 0, \quad j = 1, ..., m.$$

Optimización Entera

Trata del modelado y resolución de problemas con variables discretas:

- enteras: indivisibilidad,
- binarias: decisiones.

Características:

- Formulaciones más representativas y flexibles del problema
- Explosión combinatoria de soluciones factibles
- Resolución más difícil que optimización lineal; no se conoce método eficiente general de resolución.