Práctica 11: Frentes de Pareto

Fabiola Vázquez

2 de diciembre de 2020

---

1. Introducción

En esta práctica [4] se trabaja con optimización multicriterio, es decir, a un mismo conjunto de variables se le asignan valores de tal manera que se optimicen más de dos funciones objetivo, sin que una mejora de una función empeore otra. Primero se generan polinomios pseudo aleatorios y se determina si se va a minimizar o maximizar. Se trabaja con el concepto de dominancia de Pareto, se identifican todas aquellas soluciones que dominan (mejora por lo menos un objetivo sin empeorar ningún otro) y a dicho conjunto de soluciones se le denomina frente de Pareto. El objetivo de esta práctica es visualizar el porcentaje de soluciones que pertenecen al frente al ir aumentando la cantidad de funciones objetivo. El experimento se lleva a cabo en el software R [3] en un cuaderno de Jupyter [2].

2. Experimento

Se varía la cantidad n de funciones, $n \in \{2, 3, \ldots, 12\}$. Para cada uno de los valores se realizan 200 repeticiones considerando la generación de 200 soluciones pseudo aleatorias. Los resultados se almacenan en un `data.frame`, el cuadro 1 muestra un fragmento de los datos recopilados.

En la figura 1 se tiene un gráfico comparando el porcentaje de soluciones que pertenecen al frente de Pareto para cada cantidad de funciones. Como se puede apreciar en dicha figura, al considerar cinco o más funciones el porcentaje de soluciones pertenecientes al frente es mayor que el 50 % y con siete o más se considera casi la totalidad de las soluciones como parte del frente.

Se realiza pruebas T de Student para concluir si existen diferencia significativa entre las medias de cada uno de los conjuntos de datos. Como en la figura 1 se aprecia que el porcentaje va en aumento, se realiza la prueba con grupos consecutivos. Se identifica a los conjuntos de números con la cantidad de funciones objetivos con las que se trabaja. En el cuadro 2 se muestran los valores p obtenidos en dichas pruebas. Como se puede apreciar en las primeras cinco pruebas se obtienen valores menores a 0.05, por lo que se concluye que sí existen diferencias significativas entre las medias. A partir de la prueba seis, los valores obtenidos son mayores que 0.05, por lo que no se tiene evidencia suficiente para considerar que existe una diferencia significativa entre dichos conjuntos.

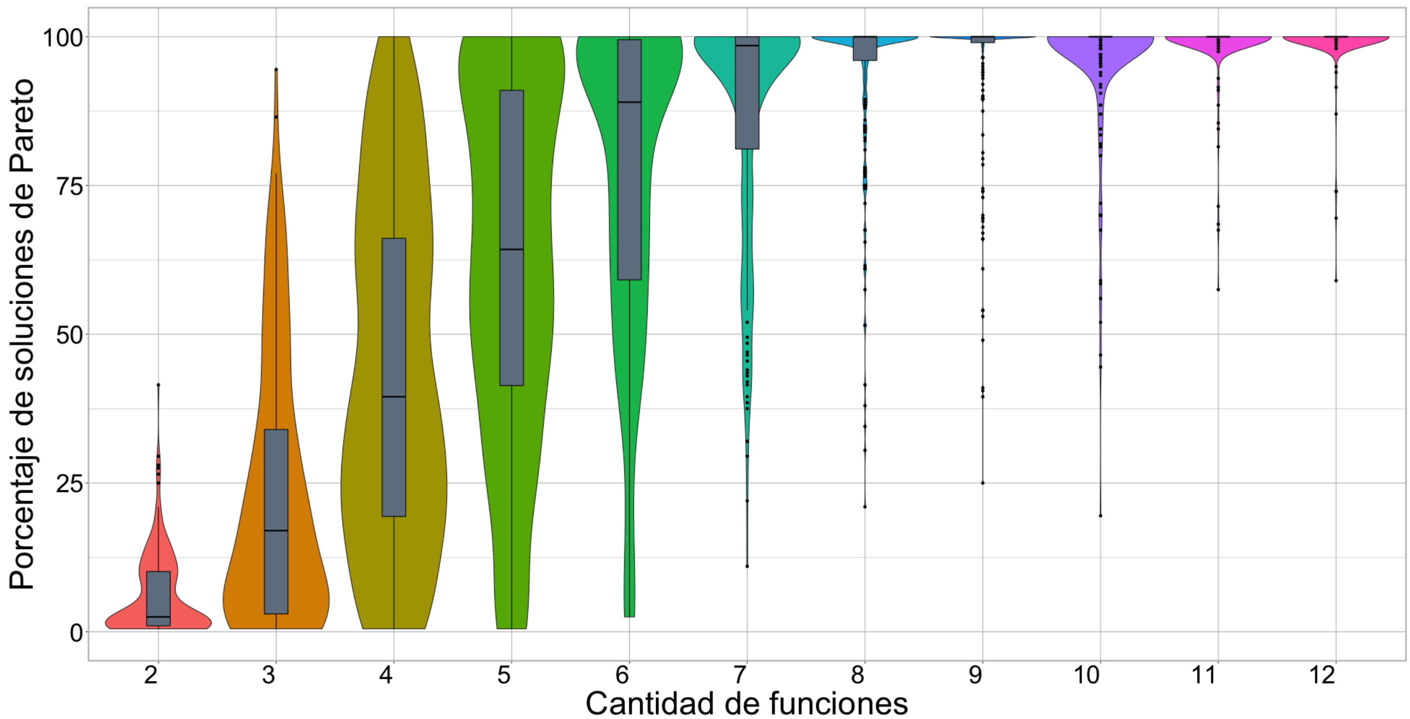


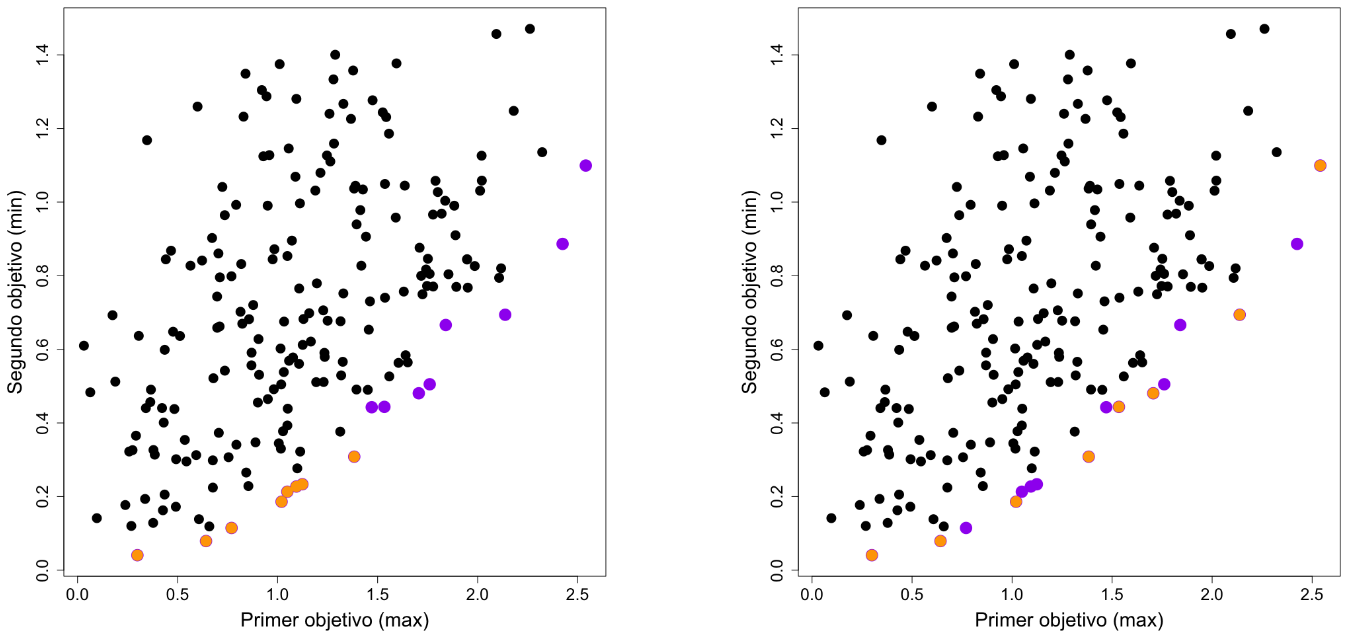
Figura 1: Gráficos de violín del porcentaje de soluciones que son parte del frente de Pareto.

Cuadro 1: Fragmento de los datos.

Funciones	Porcentaje
2	8.0
3	24.0
4	1.0
5	1.0
6	57.0
7	99.0
8	99.5
9	100.0
10	100.0
11	100.0

Cuadro 2: Valores p obtenidos de aplicar la prueba T de Student a distintos pares de conjuntos.

Conjuntos	Valor p
2 y 3	2.00×10^{-16}
3 y 4	1.18×10^{-14}
4 y 5	1.53×10^{-06}
5 y 6	1.18×10^{-05}
6 y 7	1.18×10^{-04}
7 y 8	0.46
8 y 9	0.18
9 y 10	0.03
10 y 11	0.50



(a) No diversificada.

(b) Diversificada.

Figura 2: Ejemplos de subconjuntos (naranja) del frente de Pareto (morado).

3. Reto 1

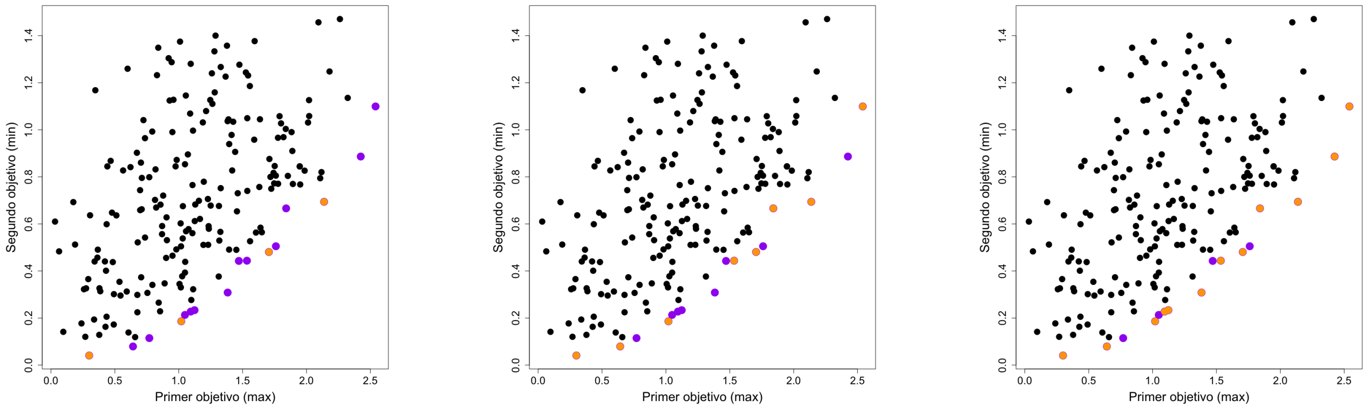
En el primer reto se selecciona un subconjunto, cuyo tamaño es un porcentaje del tamaño del frente de Pareto de tal forma que el subconjunto esté *diversificado*, es decir, que no esté agrupado en una sola zona del frente. En la figura 2a se muestra un subconjunto del frente con el 50 % del tamaño original no diversificado, como se aprecia todas las soluciones se ubican en la zona inferior izquierda, a diferencia de la figura 2b donde las soluciones están diversificadas.

Para lograr obtener el subconjunto diversificado, se trabaja con la función `kmeans` [1] que agrupa datos minimizando la distancia entre ellos. Esta función indica a que grupo pertenece cada dato, para tomar solo uno de cada grupo se utiliza la función de diversificación dada en el código 1.

```
diversidad <- function(front, porcentaje){
contador <- c()
posiciones <-c()
kmeans <- kmeans(front, round(dim(front)[1]*porcentaje/100), iter.max = 1000, nstart =
    10)
for (i in 1:length(kmeans$cluster)){
    if(sum(kmeans$cluster[i]==contador)==0){
        posiciones <- c(posiciones, i)
        contador <- c(contador, kmeans$cluster[i])
    }

}
    return(posiciones)
    }
```

Código 1: Función de diversificación.



(a) Subconjunto con el 25 % del frente original.

(b) Subconjunto con el 50 % del frente original.

(c) Subconjunto con el 75 % del frente original.

Figura 3: Variando el tamaño del subconjunto (naranja) del frente de Pareto (morado).

Referencias

[1] Rondald Delgado. Agrupamiento por K-Medios (K-Means Clustering). https://rpubs.com/rdelgado/399475.

[2] Thomas Kluyver, Benjamin Ragan-Kelley, Pérez, et al. Jupyter notebooks—a publishing format for reproducible computational workflows. In *Positioning and Power in Academic Publishing: Players, Agents and Agendas: Proceedings of the 20th International Conference on Electronic Publishing*, page 87. IOS Press, 2016.

[3] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.

[4] Elisa Schaeffer. Práctica 10: frentes de Pareto. https://elisa.dyndns-web.com/teaching/comp/par/p10.html.