APUNTES SOBRE EMPAREJAMIENTOS BILATERALES UNO-A-UNO

JUAN PABLO TORRES-MARTÍNEZ, UNIVERSIDAD DE CHILE

En un problema de *emparejamiento bilateral uno-a-uno* existen dos grupos finitos de individuos M y W. Cada individuo en un grupo tiene preferencias estrictas, completas y transitivas por los miembros del otro grupo. Así, cada individuo $m \in M$ es caracterizado por un ranking $R(m)$ de $W \cup \{m\}$. Análogamente, las preferencias de cada individuo $w \in W$ son representadas por un ranking $R(w)$ de $M \cup \{w\}$.

Un *emparejamiento* es dado por una función $\mu : M \cup W \to M \cup W$ tal que, para cada par $(m, w) \in M \times W$, se cumplen las siguientes propiedades:

(i) $(\mu(m), \mu(w)) \in (W \cup \{m\}) \times (M \cup \{w\})$;

(ii) $m = \mu(w)$ si y solamente si $w = \mu(m)$.

Diremos que un emparejamiento μ es *bloqueado por un par de individuos* si existe $(m, w) \in M \times W$ tal que m y w prefieren estar juntos a mantenerse emparejados, respectivamente, con $\mu(m)$ y $\mu(w)$. Un emparejamiento μ es *bloqueado por un individuo* si existe $z \in M \cup W$ que prefiere estar solo a mantenerse emparejado con $\mu(z)$.

Un emparejamiento μ es *individualmente racional* si no puede ser bloqueado por ningún individuo. Además, diremos que μ es *estable* si es individualmente racional y no puede ser bloqueado por ningún par de individuos.

Nuestro objetivo es caracterizar el conjunto de emparejamientos estables.

Teorema 1 (Gale y Shapley (1962))
El conjunto de emparejamientos estables es diferente de vacío. Además, siempre se puede obtener un emparejamiento estable siguiendo el algoritmo de <u>aceptación diferida</u>.

Nos referiremos a un emparejamiento estable como *M-óptimo* si ningún individuo en M prefiere el resultado de otro emparejamiento estable. Análogamente, un emparejamiento estable es *W-óptimo* si ningún individuo en W prefiere el resultado de otro emparejamiento estable.

Teorema 2 (Gale and Shapley (1962))
Suponga que se utiliza el algoritmo de aceptación diferida. Si los individuos en M hacen las propuestas, entonces el emparejamiento M-óptimo es implementado. Análogamente, el emparejamiento W-óptimo es implementado cuando los individuos en W hacen las propuestas.

Demostración. Diremos que una mujer $w \in W$ es alcanzable por un hombre $m \in M$ si existe un emparejamiento estable μ tal que $\mu(m) = w$.[1] Sea μ_M el emparejamiento estable alcanzado por el algoritmo de aceptación diferida cuando los hombres hacen las propuestas.

Suponga que al inicio de la etapa $k \geq 0$ del algoritmo que implementa μ_M ningún hombre ha sido rechazado por una mujer alcanzable.

Asuma que una mujer w rechaza a un hombre m durante la etapa k del algoritmo de aceptación diferida. Si ella prefiere estar sola a emparejarse con él, entonces cualquier emparejamiento que los junte será individualmente irracional, lo cual implica que w no es alcanzable por m. Alternativamente, suponga que w rechaza a m para aceptar o mantener (temporalmente) la propuesta de un hombre m'. Si hubiera un emparejamiento μ tal que $\mu(m) = w$, entonces el par (m', w) bloquearía a μ. Efectivamente, w prefiere a m' por sobre m, pues lo rechazo para aceptar o mantener la oferta de m' en el algoritmo de aceptación diferida, mientras que m' prefiere a w por sobre $\mu(m')$, pues por hipótesis todas las mujeres que él considera mejores que w no son alcanzables.

Por lo tanto, el algoritmo que implementa μ_M asegura que cada hombre es emparejado con la mujer alcanzable mejor rankeada. Esto es, μ_M es M-óptimo. Argumentos análogos, cambiando los papeles de M y W, demuestran que el emparejamiento μ_W alcanzado por el algoritmo de aceptación diferida cuando las mujeres hacen las propuestas es W-óptimo. $\square$

El siguiente exemplo muestra que el resultado del proceso de emparejamiento alcanzado por el mecanismo de aceptación diferida depende del lado del mercado que hace las propuestas. Por lo tanto, los emparejamientos M-óptimos y W-óptimos no necesariamente coinciden.

Ejemplo 1
Si $M = \{m_1, m_2, m_3, m_4\}$, $W = \{w_1, w_2, w_3\}$ y las preferencias son dadas por

$R(m_1)$	$R(m_2)$	$R(m_3)$	$R(m_4)$	$R(w_1)$	$R(w_2)$	$R(w_3)$
w_1	w_3	w_2	w_1	m_2	m_4	m_1
w_3	w_1	m_3	w_2	m_1	m_2	m_4
w_2	w_2	w_1	w_3	m_4	m_3	m_2
m_1	m_2	w_3	m_4	w_1	m_1	m_3
				m_3	w_2	w_3

Entonces, los emparejamientos óptimos viene dados por

$$\mu_M(m_1) = w_1, \quad \mu_M(m_2) = w_3, \quad \mu_M(m_3) = m_3, \quad \mu_M(m_4) = w_2;$$

$$\mu_W(m_1) = w_3, \quad \mu_W(m_2) = w_1, \quad \mu_W(m_3) = m_3, \quad \mu_W(m_4) = w_2.$$

$\square$

[1]Por conveniencia, nos referiremos a los individuos en M como *hombres* y a los individuos em W como *mujeres*.

Diremos que un conjunto de individuos $A \subseteq M \cup W$ considera el emparejamiento μ tan bueno cuanto μ' si para cada $a \in A$ tal que $\mu(a) \neq \mu'(a)$ tenemos que $\mu(a)R(a)\mu'(a)$. Notación: $\mu \succeq_A \mu'$.

Teorema 3 (Knuth (1976))
Si μ y μ' son emparejamientos estables, $\mu \succeq_M \mu'$ si y solo si $\mu' \succeq_W \mu$.

Demostración. Sean μ y μ' dos emparejamientos estables diferentes tales que $\mu \succeq_M \mu'$. Si μ' no es tan bueno cuanto μ para W, entonces existe $w \in W$ tal que $\mu(w)R(w)\mu'(w)$. Por lo tanto, μ' puede ser bloqueado por el par de individuos $(\mu(w), w)$, lo cual contradice su estabilidad. $\square$.

Corolario
El emparejamiento M-óptimo (resp., W-óptimo) junta a cada individuo en W (resp., M) con la alternativa menos aceptable entre aquellas que pueden ser implementadas de forma estable.

Un emparejamiento μ está en el *núcleo* del mercado bilateral si no existe una coalición de individuos en $M \cup W$ que pueden bloquearlo, i.e., emparejarse entre si y generar una asignación Pareto superior para la coalición.

Teorema 4 (Roth and Sotomayor (1990))
El núcleo del mercado bilateral uno-a-uno coincide con el conjunto de emparejamientos estables.

Demostración. Si μ está en el núcleo entonces tiene que ser estable, pues caso contrario sería bloqueable por una coalición compuesta por uno o por dos individuos. Recíprocamente, sea μ un emparejamiento estable y asuma que no está en el núcleo. Entonces, existe una coalición $A \subseteq M \cup W$ (compuesta por hombres y mujeres) y un emparejamiento $\mu' : A \to A$ tal que: (i) para algún $z \in A$, $\mu'(z) \neq \mu(z)$; (ii) para cada $z \in A$, $\mu(z) \neq \mu'(z) \Longrightarrow \mu'(z)R(z)\mu(z)$. Como μ es estable, no puede ser bloqueado por ningún individuo en A. Luego, existe al menos una mujer $w \in A$ tal que $\mu'(w) \in M$ y $\mu'(w) \neq \mu(w)$. Pero esto implica que μ es bloqueado por el par de individuos $(\mu'(w), w)$, lo cual contradice su estabilidad. $\square$

Algunos resultados adicionales sobre la estructura del conjunto de emparejamientos estables:

Teorema 5 (Gale and Sotomayor (1985))
En un emparejamiento estable el conjunto de individuos que quedan solos es siempre el mismo.

Demostración. Fije emparejamientos estables μ y μ'. Sea M_μ el conjunto de hombres que prefiere μ a μ'. Definiciones análogas se aplican para los símbolos $M_{\mu'}$, W_μ y $W_{\mu'}$. Entonces, si $m \in M_\mu$ tenemos que $\mu(m)R(m)\mu'(m)$. Luego, $w := \mu(m)$ debe cumplir $\mu'(w)R(w)\mu(w)$, pues caso contrario el par (m, w) podría bloquear a μ'. Concluimos que $\mu(M_\mu) \subseteq W_{\mu'}$. De forma análoga, como $w \in W_{\mu'}$ asegura que $\mu'(w)R(w)\mu(w)$, $m := \mu'(w)$ debe cumplir $\mu(m)R(m)\mu'(m)$ pues caso contrario el par (m, w) podría bloquear a μ. Concluimos que $\mu'(W_{\mu'}) \subseteq M_\mu$. Por lo tanto, como μ y μ' son funciones inyectivas, concluimos que $\mu(M_\mu) = W_{\mu'}$ y $\mu'(W_{\mu'}) = M_\mu$.

Ahora, suponga que existe $m \in M$ tal que $\mu'(m) = m$ y $\mu(m) \neq m$. Entonces, la estabilidad de μ asegura que $\mu(m)R(m)m$. Por lo tanto, $m \in M_\mu$. Como $\mu'(W_{\mu'}) = M_\mu$, concluimos que existe $w \in W$ tal que $\mu'(w) = m$, una contradicción. $\square$

Teorema 6 (Gale and Sotomayor (1985))
Cuando un nuevo individuo m (respectivamente, un nuevo individuo w) entra al mercado, ningún agente en W (respectivamente, M) está peor si se implementa μ_M o μ_W.

El siguiente resultado, conocido como el Segundo Teorema de Optimalidad, muestra que no existe ningún emparejamiento (estable o no) en el cual cada individuo en un grupo pueda estar mejor que en su matching óptimo.

Teorema 7 (Roth (1982), Gale and Sotomayor (1985))
Para los individuos en M (resp., en W), el emparejamiento μ_M (resp., μ_W) es débilmente Pareto óptimo en el conjunto de emparejamientos individualmente racionales.

Demostración. Haremos la demostración para μ_M. Sea μ es un emparejamiento individualmente racional tal que $\mu(m)R(m)\mu_M(m)$ para todo $m \in M$. Entonces, μ debe emparejar a cada hombre con una mujer que lo rechazó durante el proceso de aceptación diferida. Por lo tanto, la racionalidad individual de μ asegura que las mujeres en $\mu(M) \subseteq W$ están emparejadas en μ_M. Luego, todos los hombres están emparejados en μ_M, lo cual implica que la última mujer a aceptar una propuesta no rechazó a nadie. Una contradicción. $\square$

El siguiente ejemplo muestra que el emparejamiento estable M-óptimo puede ser Pareto dominado por un emparejamiento individualmente racional, el cual obviamente será inestable a consecuencia del Teorema 2.

Ejemplo 2
Suponga que $M = \{m_1, m_2, m_3\}$ y $W = \{w_1, w_2, w_3\}$.
Si las preferencias individuales son dadas por

$R(m_1)$	$R(m_2)$	$R(m_3)$	$R(w_1)$	$R(w_2)$	$R(w_3)$
w_2	w_1	w_1	m_1	m_3	m_1
w_1	w_2	w_2	m_2	m_1	m_2
w_3	w_3	w_3	m_3	m_2	m_3
m_1	m_2	m_3	w_1	w_2	w_3

entonces $\mu_M = \mu_W = [(m_1, w_1), (m_2, w_3), (m_3, w_2)]$. Note que $\mu = [(m_1, w_2), (m_2, w_3), (m_3, w_1)]$ domina en el sentido de Pareto a μ_M desde el punto de vista de los hombres. Verifique que μ es individualmente racional e inestable, pues puede ser bloqueado por el par (m_2, w_1). $\square$

El siguiente ejemplo muestra que el emparejamiento M-óptimo puede ser estrictamente Pareto dominado por un emparejamiento individualmente irracional.

Ejemplo 3
Suponga que $M = \{m_1, m_2\}$ y $W = \{w_1, w_2\}$.
Si las preferencias individuales son dadas por

$R(m_1)$	$R(m_2)$	$R(w_1)$	$R(w_2)$
w_1	w_1	w_1	m_1
w_2	w_2	m_1	m_2
m_1	m_2	m_2	w_2

entonces $\mu_M = \mu_W = [(m_1, w_2), (m_2, m_2), (w_1, w_1)]$. Además, desde el punto de vista de los hombres, $\mu = [(m_1, w_1), (m_2, w_2)]$ domina estrictamente en el sentido de Pareto a μ_M. Note que μ es individualmente irracional pues w_1 preferiría estar sola a emparejarse con m_1. □

Concluiremos esta sección con algunos resultados sobre el comportamiento estratégico de los individuos al verse enfrentados a un planificador, el cual busca implementar un mecanismo de emparejamiento centralizado luego de recibir información de los individuos sobre sus preferencias.

Asuma que en un mercado bilateral uno-a-uno caracterizado por $(M, W, (R(z))_{z \in M \cup W})$, los individuos informan preferencias $R' = (R'(z))_{z \in M \cup W}$ a un planificador central que implementa un emparejamiento $\Phi(R')$. Si para todo R' el emparejamiento $\Phi(R')$ es estable, diremos que Φ es un mecanismo de emparejamiento centralizado *estable*.[2]

Teorema 8 (Roth (1982))
No existe ningún mecanismo de emparejamiento centralizado estable en el cual reportar las verdaderas preferencias sea una estratégia dominante para cada individuo.

Demostración. Es suficiente probar que existe un mercado bilateral donde no hay mecanismos de emparejamiento centralizado estables donde reportar las verdaderas preferencias sea una estratégia dominante. Suponga que $M = \{m_1, m_2\}$, $W = \{w_1, w_2\}$ y que las preferencias vienen dadas por $w_i R(m_i) w_j R(m_i) m_i$ y $m_j R(w_i) m_i R(w_i) w_i$, donde $i \neq j$. Entonces, los únicos emparejamientos estables son $\mu = [(m_1, w_1), (m_2, w_2)]$ y $\mu' = [(m_1, w_2), (m_2, w_1)]$. Note que los hombres prefieren μ a μ' y las mujeres μ' a μ.

Por lo tanto, si Φ es un mecanismo de emparejamiento centralizado estable, entonces $\Phi(R) \in \{\mu, \mu'\}$. Note que, caso $\Phi(R) = \mu$ y $M \cup \{w_1\}$ estén reportando sus verdaderas preferencias, w_2 tiene incentivos a reportar preferencias $m_1 R'(w_2) w_2 R'(w_2) m_2$. Efectivamente, el único emparejamiento estable del mercado bilateral $(M, W, R(m_1), R(m_2), R(w_1), R'(w_2))$ es μ', el cual mejora la situación

[2] Asumiremos que el emparejamiento $\mu(R')$ está bien definido para cualquier reporte de preferencias (estrictas) R'.

de w_2 con respecto a μ. Alternativamente, cuando $\Phi(R) = \mu'$ y $W \cup \{m_1\}$ están reportando sus verdaderas preferencias, m_2 tiene incentivos a reportar preferencias $w_2 R'(m_2) m_2 R'(m_2) w_1$. Efectivamente, el único emparejamiento estable del mercado bilateral $(M, W, R(m_1), R'(m_2), R(w_1), R(w_2))$ es μ, el cual mejora la situación de m_2 con respecto a μ'.

Concluimos que cuando el mecanismo centralizado estable es dado por Φ reportar las verdaderas preferencias no es una estratégia dominante. $\square$

En el ejemplo que demuestra el Teorema 8, las mujeres tienen incentivos a manipular sus preferencias cuando el mecanismo centralizado implementa el emparejamiento M-óptimo, mientras que son los hombres los que manipulan información cuando el mecanismo centralizado implementa el emparejamiento W óptimo. Esto puede ser explicado por el siguiente resultado.

Dado un mercado bilateral uno-a-uno $(M, W, (R(z))_{z \in M \cup W})$ y $S \in \{M, W\}$, diremos que un mecanismo centralizado estable Φ es *compatible con incentivos para los individuos en* S si no existe ningún conjunto $S' \subseteq S$ que tenga incentivos a reportar preferencias falsas dado que los otros individuos están reportando sus verdaderas preferencias.

Teorema 9 (Dubins and Freedman (1981))
Dado un mercado bilateral uno-a-uno $(M, W, (R(z))_{z \in M \cup W})$ *y* $S \in \{M, W\}$*, todo mecanismo centralizado estable* Φ *tal que* $\Phi(R) = \mu_S$ *es compatible con incentivos para los individuos en* S*.*

La demostración de este resultado usará la siguiente propiedad:

Proposición.
Dado un emparejamiento μ*, sea* M' *el conjunto de hombres que prefiere* μ *a* μ_M*. Entonces, existe un par* $(m, w) \in (M \setminus M') \times W$ *que bloquea al emparejamiento* μ*.*

Demostración del Teorema 9. Nos enfocaremos en el caso $S = M$, pues el otro es análogo. Suponga que Φ no es compatible con incentivos para los hombres. Entonces, existe un conjunto $M' \subseteq M$ que tiene incentivos a reportar preferencias falsas $(R'(m))_{m \in M'}$ cuando los otros individuos están reportando las verdaderas preferencias, con el objetivo de implementar un emparejamiento estable μ del mercado $(M, W, (R'(z))_{z \in M'}, (R(z))_{z \in (M \setminus M') \cup W})$.

Por la Proposición anterior, va a existir un par $(m, w) \in (M \setminus M') \times W$ que bloquea a μ en $(M, W, (R(z))_{z \in M \cup W}$. Como m y w están reportando sus verdaderas preferencias, concluimos que (m, w) también bloquea a μ en $(M, W, (R'(z))_{z \in M'}, (R(z))_{z \in (M \setminus M') \cup W})$, lo cual contradice la estabilidad de μ. $\square$

Bibliografía

Gale, D. and L.S. Shapley (1962): "College admissions and the stability of marriage," *American Mathematical Monthly,* volume 69, pages 9.15.

Knuth, D. (1976): "Mariage stables et leurs relation avec d'autres problèmes combinatoires," Les Presses de L'Université de Montréal.

Roth, A. (1982): "The economics of matching: stability and incentives," *Mathematics of Operations Research,* volume 7, pages 617-628.

Gale, D. and M. Sotomayor (1985): "Some remarks on the stable matching problem," *Discrete Applied Mathematics*, volume 11, pages 223-232.

Roth, A. and M. Sotomayor (1990): "Two-Sided Matching: A study in game-theoretic modeling and analysis," *Econometric Society Monographs,* Cambridge University Press.