Probablemente es Tenis

Matías Salim Seda Auil
Universidad de Chile

El tenis, uno de los deportes más famosos del mundo, consiste en, básicamente, dos jugadores golpeando una pelota con una raqueta dentro un área dividida por un red. Un partida de tenis, por su parte, consiste en un conjunto de *sets*. Un set es un conjunto de juegos. Un juego es un conjunto de pelotas. En último lugar, una pelota se juega cuando un jugador saca y termina cuando el otro jugador pierde la pelota. A priori, pareciese que este deporte es relativamente simple comparado con otros deportes que involucran más factores, como el fútbol donde existen 11 jugadores por equipo. Sin embargo, detrás de ésta lógica de partidas, set, juegos y pelotas, se esconde un complejo mundo de probabilidades.
Así, en el siguiente trabajo, se estudiará un problema probabilístico asociado al deporte del tenis. En particular, dado un partido entre dos jugadores A y B, con una probabilidad p que A gane una pelota y una probabilidad $q = 1 - p$ que B gane una pelota (independiente de quién sirve), se quiere estudiar como la variación de la probabilidad p afecta el redimiento del jugador A, es decir, se quiere analizar que ocurre con la probabilidad de que A gane un juego y A gane un set cuando varía p. Para estudiar la variación de las probabilidades, se modelará la probabilidad de ganar un juego y la probabilidad de ganar un set a través de diversas funciones. Con las funciones planteadas, se analizará las estructuras de las probabilidades asociadas y, con las derivadas respectivas, se estudiará la variación de las probabilidades respecto a la variación de la probabilidad de ganar una pelota.
Finalmente, se concluirá, que la expresión el tenis *amplifica las diferencias* es cierta. Es decir, cuando dos jugadores presentan probabilidades muy similares para ganar una pelota, una pequeña variación de la probabilidad de ganar una pelota para un jugador puede inclinar completamente la balanza y amplificar sus probabilidades de ganar o perder un juego y un set de manera significativa.

PRESENTACIÓN DEL PROBLEMA: LA PROBABILIDAD DE GANAR

Matías es un estudiante de computación poco conocedor del mundo del tenis. Por el otro lado, su profesor de *Análisis Avanzado de Algoritmos* disfruta mucho de ese apasionante deporte que Matías no logra comprender. Sin embargo, el profesor Patricio planeó una tarea para el curso relacionada con el tenis para que los alumnos puedan comprender el indescifrable mundo del tenis. La explicación que el profesor le proporcionó a los alumnos sobre la lógica del tenis fue la siguiente:

- En primer lugar, en términos simples y resumidos, en un partido de tenis entre dos jugadores A y B, en cada pelota que se juega, A tiene una probabilidad p de ganar la pelota y B tiene una probabilidad q de ganar la pelota (independiente de quien está sirviendo), con $p + q = 1$.
- En un juego normal de tenis, gana el primer jugador que llega a ganar 4 pelotas, con una diferencia de 2 por sobre su adversario (estas 4 pelotas se cuentan *quince, treinta, cuarenta, juego*). Por otra parte, en un *tie-break*, gana el primero que llegar a ganar 7 pelotas, con una diferencia de 2. Nótese que, si se está jugando un juego de k pelotas y se llega a un empate de $k - 1$ pelotas, desde ese momento para adelante las diferencias son siempre de 1 *ventaja* o 0 *iguales* hasta que algún jugador llegue a alcanzar diferencia de 2.
- En un set normal de tenis, gana el primer jugador que llegue a ganar 6 juegos con una diferencia de 2 o, en caso de que ambos jugadores ganen 5 juegos, existen dos escenarios: un jugador puede ganar dos juegos más y ganar el set (es decir, el marcador termina 7-5) o cada jugador gana un juego y, por lo tanto, los jugadores empatan a 6 y, en ese caso, se juega un *tie-break*.
- En algunos campeonatos *grand slam* el quinto set es *largo*, lo que significa que al llegar a 6, no hay *tie-break* y se debe seguir jugando hasta que algún jugador consiga una diferencia de 2 juegos.

Para no desviar completamente el foco del análisis de algoritmos hacia el tenis, el profesor Patricio les pidió a los alumnos resolver la siguientes preguntas.

1. Generalizando a partir de los juegos mencionados, supóngase que se juega un juego en que gana el primero que llegar a ganar k pelotas con una diferencia de 2. Se quiere conocer la probrabilidad $G_k(p)$ de que A gane el juego de k pelotas con una diferencia de 2.
2. Ahora, dado la función $G_k(p)$, se quiere calcular la probabilidad $S(p)$ de que A gane un set.

3. Del mismo modo, se quiere calcular $L(p)$, la probabilidad de que A gane un set largo.
4. Finalmente, se quiere graficar las funciones $G_4(p)$, $G_7(p)$, $S(p)$ y $L(p)$ y comparar como cada una de ellas amplifica las diferencias entre los jugadores.

RESOLUCIÓN DEL PROBLEMA: SET LARGO DE PROBABILIDAES

En primer lugar, respecto a la notación, un juego se denotará como una palabra. Por ejemplo, la palabra $AABA$ significa que el jugador A gano la primera pelota, luego también ganó la segunda, después el jugador B ganó la tercera pelota y, finalmente, el jugador A ganó la cuarta pelota.

Ahora, en primera instancia, para encontrar una fórmula general para $G_k(p)$ se estudiarán los casos bases $G_2(p)$ y $G_3(p)$.

- $k = 2$:
  Se quiere conocer $G_2(p)$, la probabilidad que A gane un juego ganando 2 pelotas. En términos generales, existen dos posibles escenarios.

  1. A gana el juego consiguiendo una ventaja de dos pelotas ganando 2 pelotas, sin empate.
  2. A empata con B y, en el desempate, consigue una venjata de dos pelotas.

  El primer escenario solo puede ocurrir de una forma: AA.

  Los empates se pueden dar de la siguiente forma: AB y BA. Ahora, cuando se tiene un empate, viene un número indeterminado de pelotas (hasta que A consiga una ventaja de 2). En particular, la estrucutra para que, luego de un empate, A consiga una victoria, esta expresada por el siguiente autómata:

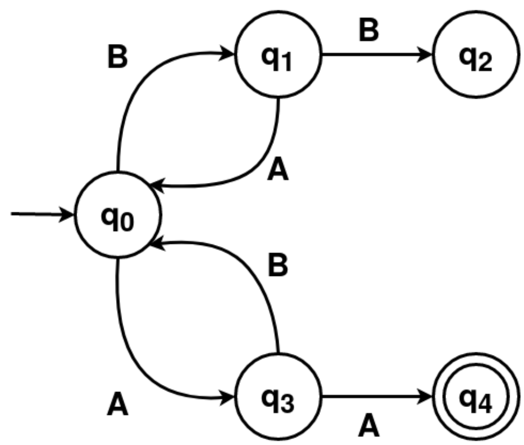

Figura 1. Autómata que modela la estrucutra de un desempate favorable para A.

  Nótese que las palabras (desempates) posibles que genera este autómata son de la forma $(AB + BA)^*AA$. Así, se puede definir $\mathcal{J}_2$ como el conjuto de juego posibles de 2 pelotas donde A gana. $\mathcal{J}_2$ viene dada por la siguiente expresión:

$$\mathcal{J}_2 = AA + (AB + BA)(AB + BA)^*AA$$

  De $\mathcal{J}_2$ se puede derivar la función $G_2(p)$, es decir, la probabilidad que A gane un juego de 2 pelotas. Recordando que la probabilidad que gane una pelota A es p y la probabilidad que gane una pelota B es $q = 1 - p$, se tiene que la expresión $G_2(p)$ viene dada por:

$$G_2(p) = pp + (p(1-p) + p(1-p))\frac{1}{1 - 2p(1-p)}pp$$

Lo que es equivalente a:

$$G_2(p) = \sum_{i=0}^{2-2} \binom{2-1+i}{i} p^{2-1}(1-p)^i p + \binom{2(2-1)}{2-1} p^{2-1}(1-p)^{2-1} \frac{1}{1-2p(1-p)} pp$$

- $k = 3$:
  Ahora, se quiere conocer $G_3(p)$, la probabilidad que A gane un juego ganando 3 pelotas. En términos generales, del mismo modo que para el caso $G_2(p)$, existen dos posibles escenarios.

  1. A gana el juego consiguiendo una ventaja de dos pelotas ganando 3 pelotas, sin empate.
  2. A empata con B y, en el desempate, consigue una venjata de dos pelotas.

  El primer escenario puede ocurrir de las siguientes formas: AAA, $AAAB$, $AABA$, $ABAA$ y $BAAA$.

  Los empates se pueden dar de la siguiente forma: $AABB$, $ABAB$, $BAAB$, $BBAA$, $BABA$ y $ABBA$. Ahora, cuando se tiene un empate, viene un número indeterminado de pelotas (hasta que A consiga una ventaja de 2) La estrucutra para que, luego de un empate, A consiga una victoria, es, del mismo modo que para la función $G_2(p)$, modelada por el autómata de la figura 1.

  Así, se puede definir $\mathcal{J}_3$ como el conjunto de juegos posibles de 3 pelotas donde A gana. $\mathcal{J}_3$ viene dada por la siguiente expresión:

  $$\mathcal{J}_3 = AAA + AAAB + AABA + ABAA + BAAA + (AABB + ABAB + BAAB + BBAA + BABA + ABBA)(AB+BA)^* AA$$

  De $\mathcal{J}_3$ se puede derivar la función $G_3(p)$, es decir, la probabilidad que A gane un juego de 3 pelotas. Recordando que la probabilidad que gane una pelota A es p y la probabilidad que gane una pelota B es $q = 1 - p$, se tiene que la expresión G_3 viene dada por:

  $$G_3(p) = \sum_{i=0}^{3-2} \binom{3-1+i}{i} p^{3-1} q^i p + \binom{2(3-1)}{3-1} p^{2-1}(1-p)^{2-1} \frac{1}{1-2p(1-p)} pp$$

Dado los ejemplos de casos bases, se puede notar que $G_k(p)$, la probabilidad que A gane un juego en k pelotas viene dada por la siguiente expresión:

$$G_k(p) = \sum_{i=0}^{k-2} \binom{k-1+i}{i} p^{k-1}(1-p)^i p + \binom{2(k-1)}{k-1} p^{k-1}(1-p)^{k-1} \frac{pp}{1-2p(1-p)}$$

En particular, el primer término de la expresión, $\sum_{i=0}^{k-2} \binom{k-1+i}{i} p^{k-1} q^i p$, representa todas las forma que A puede ganar un juego de k pelotas sin empate. Nótese que A puede ganar las k pelotas y B no ganar ni una. Solo existe un posible escenario donde ocurre lo anteriormente mencionado y la probabilidad de ese escenario es p^k. Del mismo modo, A puede ganar las k pelotas pero B puede ganar 1 pelota. Para que esto ocurra y gane A, B puede ganar cualquiera de las primeras $k - 1 + 1 = k$ pelotas pero, la última pelota, la debe ganar A. Por esta razón, existen $\binom{k-1+1}{1} = k$ formas que B gane alguna de las primeras pelotas. La probabilidad para que ocurre cualquiera de esos k escenarios es la misma y viene dada por la expresión $p^{k-1}(1-p)p$. En general, B puede ganar i pelotas, con $i \in \mathbb{N}$, $0 \leq i \leq k-2$ y existen $\binom{k-1+i}{i}$ formas en que B puede ganar cualquiera de las $k - 1 + i$ primeras pelotas. La probabilidad que ocurre cualquiera de escenarios mencionados es la misma y viene dado por la expresión $p^{k-1}(1-p)^i p$. Por esa razón, la expresión $\sum_{i=0}^{k-2} \binom{k-1+i}{i} p^{k-1} q^i p$, es la probabilidad que A gane un juego de k pelotas sin empate.

Por la otra parte, el segundo término de la expresión, $\binom{2(k-1)}{k-1} p^{k-1}(1-p)^{k-1} \frac{pp}{1-2p(1-p)}$, representa todas las posibles formas que A gane un juego de k pelotas empatando. Primero, en un juego donde gana el que gana k pelotas, para que A y B empaten, ambos deben ganar $k - 1$ pelotas. La expresión $\binom{2(k-1)}{k-1} p^{k-1}(1-p)^{k-1}$ representa todas las formas en que A y B pueden empatar. Luego, al empatar, tienen que jugar el *desempate*. Para que A gane el desempate, A debe conseguir una ventaja de dos pelotas. La expresión $\frac{pp}{1-2p(1-p)}$ representa todas las posibles formas en que A puede ganarle a B con una diferencia de dos pelotas.

Ahora, siguiendo una lógica muy similar a la planteada para conocer la expresión $G_k(p)$, se puede mostrar que $S(p)$, la probabilidad que A gane un set, viene dada por la siguiente expresión:

$$S(p)=\sum_{i=0}^{6-2}\binom{6-1+i}{i}G_4(p)^{6-1}(1-G_4(p))^iG_4(p)+\binom{2(6-1)}{6-1}G_4(p)^{6-1}(1-G_4(p))^{6-1}(G_4(p)^2+2G_4(p)(1-G_4(p))G_7(p))$$

Recordando, en un set normal de tenis, gana el primer jugador que llegue a ganar 6 juegos de 4 pelotas con una diferencia de 2 juegos o, en caso que ambos jugadores ganen 5 juegos, existen dos escenarios: un jugador puede ganar dos juegos más y ganar el set (es decir, el marcador termina 7-5) o cada jugador gana un juego y, por lo tanto, los jugadores empatan a 6 y, en ese caso, se juega un *tie-break*.

Así, el primer término de la expresión $S(p)$, $\sum_{i=0}^{6-2}\binom{6-1+i}{i}G_4(p)^{6-1}(1-G_4(p))^iG_4(p)$, representa todas las forma que A puede ganar 6 juegos de 4 pelotas sin empate.

El segundo término de la expresión, $\binom{2(6-1)}{6-1}G_4(p)^{6-1}(1-G_4(p))^{6-1}(G_4(p)^2+2G_4(p)(1-G_4(p))G_7(p))$, representa todas las posibles formas que A gane un set de 6 juegos empatando. Primero, en un set donde gana el que gana 6 pelotas, para que A y B empaten, ambos deben ganar 5 juegos. La expresión $\binom{2(6-1)}{6-1}G_4(p)^{6-1}(1-G_4(p))^{6-1}$ representa todas las formas en que A y B pueden empatar. Luego, al empatar, existen dos escenaros: A gana dos juegos más y el set termina 7-5 en favor de A, escenario representado por el término $G_4(p)^2$, o A gana un juego y B gana un juego, empatando a 6 y definiendo el set a través de un *tie-break*, escenario representado por el término $2G_4(p)(1-G_4(p))G_7(p)$

Finalmente, siguiendo una lógica muy similar a la planteada para conocer las expresiones anteriores, se puede mostrar que $S(p)$, la probabilidad que A gane un set, viene dada por la siguiente expresión:

$$L(p)=\sum_{i=0}^{6-2}\binom{6-1+i}{i}G_4(p)^{6-1}(1-G_4(p))^iG_4(p)+\binom{2(6-1)}{6-1}G_4(p)^{6-1}(1-G_4(p))^{6-1}\frac{G_4(p)G_4(p)}{1-2\cdot G_4(p)(1-G_4(p))}$$

El primer término de la expresión $L(p)$, $\sum_{i=0}^{6-2}\binom{6-1+i}{i}G_4(p)^{6-1}(1-G_4(p))^iG_4(p)$, representa todas las forma que A puede ganar 6 juegos de 4 pelotas sin empate.

El segundo término de la expresión, $\binom{2(6-1)}{6-1}G_4(p)^{6-1}(1-G_4(p))^{6-1}\frac{G_4(p)G_4(p)}{1-2\cdot G_4(p)(1-G_4(p))}$, representa todas las posibles formas que A gane un set de 6 juegos empatando. Primero, en un set donde gana el que gana 6 pelotas, para que A y B empaten, ambos deben ganar 5 juegos. La expresión $\binom{2(6-1)}{6-1}G_4(p)^{6-1}(1-G_4(p))^{6-1}$ representa todas las formas en que A y B pueden empatar. Luego, al empatar, tienen que jugar el *desempate*. Para que A gane el desempate, A debe conseguir una ventaja de dos juegos. El término $\frac{G_4(p)G_4(p)}{1-2\cdot G_4(p)(1-G_4(p))}$ representa todas las posibles formas en que A puede ganarle a B con una diferencia de dos juegos.

Finalmente, el gráfico asociado a las cuatro funciones definidas anteriormente es el siguiente:

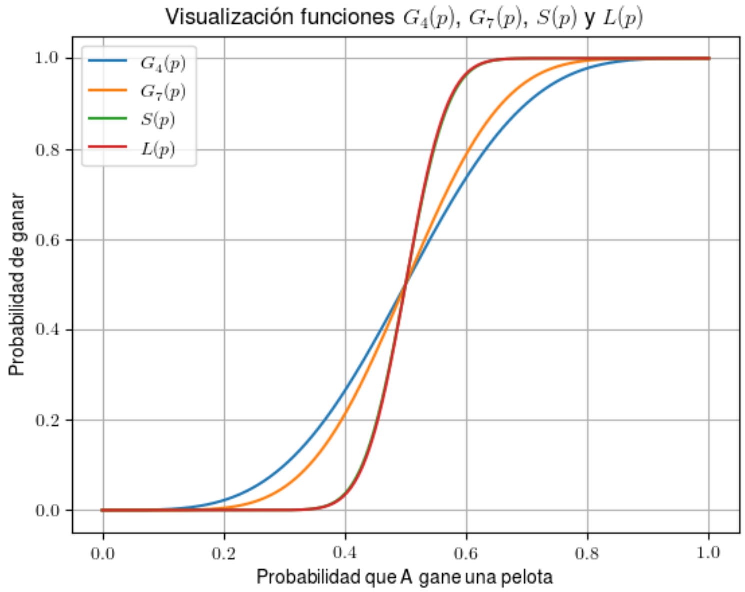

Figura 2. Visualización funciones $G_4(p)$, $G_7(p)$, $S(p)$ y $L(p)$.

ANÁLISIS DE LA SOLUCIÓN: LA IMPORTANCIA DE LAS PEQUEÑAS DIFERENCIAS

A primera vista, lo más evidente del gráfico de la figura 2 es que, cuando se tiene $p = \frac{1}{2}$, tanto la probabilidad $G_4(p)$ como las probabilidaes $G_7(p)$, $S(p)$ y $L(p)$ son iguales a $\frac{1}{2}$. Lo anterior, a priori, debiese ser evidente ya que, si la probabilidad que A gane una pelota es $\frac{1}{2}$, entonces la probabilidad que B gane una pelota también es la misma y, por tanto, la probabilidad tanto de ganar un juego normal como un tie-break, un set y un set largo debiese ser $\frac{1}{2}$ tanto para A como para B, dado que el tenis es un juego *justo*.

Sin embargo, observando el gráfico, se puede notar que el tenis *amplifica las diferencias*. La expresión *amplifica las diferencias* se refiere a que, cuando la probabilidad que A gane una pelota es relativament similar a la probabilidad que B gane una pelota , pequeñas variaciones en la probabilidad que A gane un pelota afectan notoriamente en la probabilidad que A gane un juego o un set. En particular, se puede apreciar que si A tiene una probabilidad de $\frac{2}{5}$ para ganar una pelota, su probabilidad de ganar un juego o un tie-break es cercana a $\frac{1}{5}$ y su probabilidad de ganar un set es casi nula. Por el otro lado, si A tiene una probabilidad de $\frac{3}{5}$ para ganar una pelota, su probabilidad de ganar un juego o un tie-break es cercana a $\frac{4}{5}$ y su probabilidad de ganar un set es casi 1.

Para formalizar la expresión *amplifica las diferencias*, se puede definir las funciones $G_4'(p)$, $G_7'(p)$, $S'(p)$ y $L'(p)$ tales que $G_4'(p) = \frac{d}{dp}\left(G_4(p)\right)$, $G_7'(p) = \frac{d}{dp}\left(G_7(p)\right)$, $S'(p) = \frac{d}{dp}\left(S(p)\right)$ y $L'(p) = \frac{d}{dp}\left(L(p)\right)$. Con estas funciones derivadas, se puede estudiar exactamente como se comportan las distintas funciones con pequeñas variaciones de p. Para el análisis, se graficarán las cuatros funciones derivdas.

Nótese que la hipótesis que el tenis *amplifica las diferencias* se verifica a través de la información obtenida de los gráficos de la figura 3. En valores cercanos a $p = \frac{1}{2}$, los valores de las funciones $G_4'(p)$, $G_7'(p)$, $S'(p)$ y $L'(p)$ son los mayores, en cambio, cuando p es cercana a 0 o 1, es decir, cuando p se aleja del centro $p = \frac{1}{2}$, la variación de la probabilidad de ganar respecto a cambios de p es mínima.

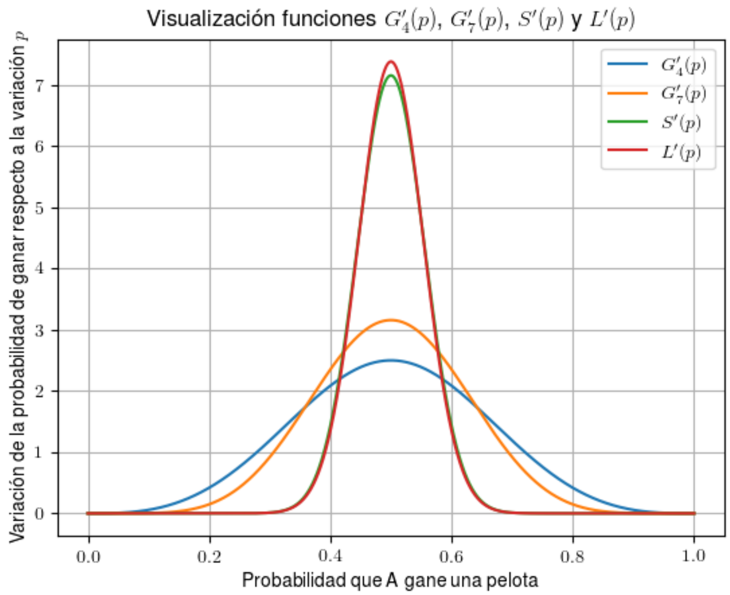

Figura 3. Visualización funciones $G'_4(p)$, $G'_7(p)$, $S'(p)$ y $L'(p)$.

DISCUSIONES Y CONCLUSIONES: SOBRE EL ORDEN DE LAS FORMAS

En primer lugar, es interesante destacar que, la estructura de una partida de tenis es una de las razones principales detrás de esta *amplificación de las diferencias*. Por decirlo de cierta forma, la probabilidad fundamental planteada en el problema es la probabilidad de ganar un pelota. Nótese que, para ganar un juego, se tiene que ganar cierta cantidad de pelotas y, por lo tanto, la probabilidad $G_k(p)$ de ganar un juego ganando k pelotas depende de la probabilidad p de ganar una pelota. Lo anteriormente dicho es evidente al observar la expresión encontrada para $G_k(p)$. Además, $G_k(p)$ no depende linealmente de p sino que, existen expresiones combinatoriales asociadas a $G_k(p)$ que modelan la dependencia de $G_k(p)$. Este tipo de dependencia *combinatorial* es lo que genera la *amplificación de las diferencias*.

Del mismo modo, nótese que, para ganar un set, se tiene que ganar cierta cantidad de juegos y, por lo tanto, la probabilidad $S(p)$ de ganar un set depende de la probabilidad $G_k(p)$ de ganar un juego de k pelotas. En particular, $S(p)$ depende de $G_4(p)$ y $G_7(p)$. Lo anteriormente mencionado es evidente al observar la expresión encontrada para $S(p)$. Asimismo, $S(p)$ no depende linealmente de $G_4(p)$ sino que, existen expresiones combinatoriales asociadas a $G_4(p)$ que modelan la dependencia de $S(p)$ con $G_4(p)$. Este tipo de dependencia *combinatorial* es lo que genera la *amplificación de las diferencias*. Nótese que, por lo observado en la figura 3, la *amplificación de las diferencias* es más significativa en $S(p)$ que en G_4 y se debe a que, por decirlo de cierta forma, en $S(p)$ existe una dependencia combinatorial de *segundo grado*. Primero, existe una dependencia combinatorial entre $S(p)$ y $G_4(p)$ y, de la misma forma, $G_4(p)$ depende combinatorialmente de p. Este dependencia combinatorial de *segundo grado* amplifica mas drásticamente las diferencias que una dependencia combinatorial de *primer grado*.

Respecto a la función $L(p)$, al presentar una expresión similar a $S(p)$, las conclusiones mencionadas respecto a $S(p)$ son, del mismo modo, extrapolables a $L(p)$. En particular, tanto las funciones $S(p)$ y $L(p)$ como $S'(p)$ y $L'(p)$ presentan gráficas extremadamente similares, lo cual no es de extrañar al observar las expresiones de $S(p)$ y $L(p)$ y notar sus semejanzas.

Finalmente y, lo más relevante a destacar del presente trabajo, es la confirmación de la hipótesis el tenis *amplifica las diferencias*. El análisis de las funciones $G_4(p)$, $G_7(p)$, $S(p)$ y $L(p)$ y sus respectivas derivadas permitió comprobar la hipótesis planteada. Así, en un partido de tenis entre dos jugadores del mismo nivel, un pequeño cambio en la probabilidad de ganar una pelota para un jugador puede inclinar completamente la balanza y amplificar sus probabilidades de ganar o perder un juego y un set de manera significativa.