



UNIVERSIDAD DE BUENOS AIRES

FACULTAD DE CIENCIAS EXACTAS Y NATURALES

DEPARTAMENTO DE COMPUTACIÓN

Un modelo de estimación de habilidad para la Asociación Argentina de Go con perspectiva internacional

Tesis presentada para optar al título de
Licenciado en Ciencias de la Computación

Martín Amigo y Tobías Carreira Munich

Directores: Esteban Mocskos y Gustavo Landfried

Buenos Aires, 2022

Resumen

El Go es un juego de mesa milenario creado en Asia que, a pesar de tener reglas relativamente sencillas, presenta una enorme complejidad combinatoria en cuanto a las posibles jugadas. Esto deriva en grandes diferencias de habilidad entre distintos jugadores. Es una práctica común que, cuando la diferencia de habilidad entre dos jugadores es muy grande, se ofrezca una ventaja para producir un juego más entretenido. En la tradición del Go existe un sistema de ranking con el que se intenta clasificar a las personas en niveles con la idea de que la diferencia de niveles represente la ventaja que una persona debería otorgarle a la otra.

Sin embargo, determinar la ventaja necesaria para alcanzar un buen equilibrio en el juego es un problema de difícil solución ya que depende de la estimación de la habilidad de cada uno de los jugadores. El problema de estimar la probabilidad de ganar comenzó a resolverse a mediados del siglo pasado con el objetivo de evaluar la habilidad de los jugadores profesionales de ajedrez. Desde entonces la mayoría de los modelos utilizan la probabilidad del resultado observado, r , para actualizar las estimaciones de habilidad s , $P(r|s_i, s_j)$. En el caso del Go, además es importante estimar el efecto que produce la ventaja en la probabilidad de ganar.

Actualmente, la Asociación Argentina de Go (AAGo) reúne a las personas aficionadas del juego de Go en nuestro país y utiliza el sistema de ranking de la Asociación Estadounidense de Go. En la práctica, este algoritmo ha mostrado tener serias dificultades según manifiestan los propios miembros de la AAGo. La dificultad particular que tiene la AAGo es el tamaño de su comunidad que además está compuesta por una gran diversidad de habilidades. En esta tesis nos propusimos desarrollar un sistema para la AAGo que mejore sustancialmente las estimaciones de habilidad con las que cuentan en la actualidad haciendo un uso eficiente de los datos.

Para ello, decidimos adaptar el modelo estado del arte en la industria del video juego, TrueSkill Through Time (TTT), de modo tal que incorpore los efectos de las ventajas, modelo que llamamos TTT-hreg. En vez de propagar la información en un único sentido, del pasado al futuro, TTT propaga la información a través de todo el sistema ofreciendo así mejores estimaciones de habilidad con menor cantidad de datos. Luego de su publicación original, se implementó un algoritmo similar denominado Whole History Rating (WHR), de referencia en la actualidad en la estimación de habilidad de los jugadores profesionales de Go. Ambos algoritmos que utilizan toda la historia (TTT y WHR) mostraron tener un desempeño altamente superior al algoritmo utilizado actualmente por la AAGo. En particular, el modelo WHR mostró tener predicciones a priori mejores que TTT. Por otro lado, este último tiene la ventaja de permitir agregar nuevos factores respetando las reglas de la probabilidad dentro del marco bayesiano, simplificando su uso y adaptación a nuevos contextos.

En la actualidad, no es posible comparar las estimaciones de habilidad obtenidas por las diferentes Asociaciones de Go del mundo, no sólo porque utilizan sistemas diferentes, sino porque en general son bases de datos disconexas. Sin embargo, unos pocos vínculos entre bases de datos le permitirían a nuestra solución propagar la información eficientemente, garantizando la comparabilidad entre las diversas asociaciones internacionales. En este trabajo dejamos sentadas las bases para extender el uso de nuestra solución como medida de referencia internacional utilizando la base de datos completa de KGS con 100 millones de partidas de todas las asociaciones de Go del mundo.

Palabras clave: Habilidad, Go, Handicap, Aprendizaje, Comparabilidad

Índice general

1. Introducción	4
1.1. Objetivos	6
2. Modelos Probabilísticos	7
2.1. Modelos causales y razonamiento probabilístico	7
2.1.1. Principio de indiferencia	7
2.1.2. Regla de la suma	8
2.1.3. Regla del producto	9
2.2. Evaluación de modelos	11
3. Modelos de habilidad	15
3.1. Modelo general de eventos	15
3.1.1. Inferencia exacta	15
3.2. Modelos de múltiples eventos	16
3.2.1. Modelos de historial completo	17
3.2.2. Modelo AGA/AAGo	18
3.2.3. Whole History Rating	19
3.2.4. TrueSkill Through Time	21
3.3. Cálculo de la evidencia	22
4. Metodología	23
4.1. Fuente de datos	23
4.1.1. Réplica de estimaciones de AAGo	24
4.2. Exploración del funcionamiento de los algoritmos de estimación	25
4.2.1. Pequeña comunidad	25
4.2.2. Intercomunidades	26
4.2.3. Gran comunidad	30
4.2.4. Evolución de la habilidad	31
5. Resultados	36
5.1. Comparación cuantitativa de los modelos	36
5.1.1. Condiciones de la evaluación	36
5.1.2. Resultados principales	37
5.2. AGA/AAGo	38
5.3. TrueSkill Through Time	39
5.4. Whole History Rating y TrueSkill Through Time	40
6. Conclusiones	43

Capítulo 1

Introducción

El Go es un juego estratégico de mesa para dos jugadores creado en China hace aproximadamente 3 o 4 milenios. En pocas palabras, el juego consiste en colocar piezas (blancas y negras, un color para cada jugador) sobre una grilla cuadrículada (generalmente de 19×19 posiciones), con el objetivo de conseguir ocupar la mayor superficie posible del tablero con piezas del propio color. Para lograrlo, cada jugador intenta rodear otras piezas con las de su color: al hacer esto, todas las piezas del contrincante son eliminadas del tablero. Al final de la partida, gana el jugador con mayor territorio y fichas capturadas.

El Go, a pesar de tener reglas relativamente sencillas, presenta una enorme complejidad debido a la inmensa cantidad de configuraciones posibles que permite el tablero. Explorar eficientemente el espacio de estrategias posibles es una tarea sumamente difícil. Esto se evidencia en la dificultad que ha tenido la inteligencia artificial para derrotar a los jugadores profesionales de Go. Si en 1996 un algoritmo de inteligencia artificial le ganó a unos de los mejores jugadores de ajedrez [CHJH02], para Go hubo que esperar otros 20 años para alcanzar el mismo objetivo [SHM⁺16]. La clave del éxito de los jugadores profesionales de Go es la herencia cultural con la que cuenta la humanidad producida de generación en generación durante milenios.

¿Pero a que nos referimos cuando hablamos de *habilidad*? Cada uno de nuestros antepasados imitó, modificó y transmitió parte de su conocimiento a la siguiente generación, llegando a nosotros sedimentado en forma de herramientas, creencias y rutinas. La especial integración de los procesos biológicos, cognitivos y sociales que permite a los humanos desarrollar culturas complejas se debió a una coevolución genético-cultural desencadenada por el surgimiento previo de la crianza cooperativa [HB20, KMH⁺20]. El surgimiento de la comprensión mutua produjo cambios radicales para nuestra especie. El conocimiento que antes debía ser redescubierto mediante costosa experiencia individual, luego de la transición comenzó a ser transmitido a la siguiente generación a través de aprendizaje social. Emergió así un nuevo sistema de información, autónomo del sistema de información biológico, lo que nos permitió ocupar todos los nichos ecológicos de la tierra como ningún otro vertebrado terrestre había logrado antes [BRH11].

La cultura es un fenómeno poblacional que emerge como consecuencia del intercambio de información entre individuos. Basada en los principios de la teoría de la evolución, se ha desarrollado un marco de **evolución cultural** que intenta explicar cómo cambia la cultura en el tiempo. La descendencia, mutación y selección de la información cultural se produce cuando una persona adopta el comportamiento de otras, sea porque es el comportamiento de la mayoría (*frequency-based strategy*), sea porque la otra persona es particularmente exitosa o prestigiosa (*payoff-based strategy*), o sea por otros tipos de sesgos de selección [BREP04, MMH16, RFH⁺11]. Todos los procesos de aprendizaje social modifican en algún grado la variante cultural adoptada. Y algunas de esas variantes culturales se propagan más rápido o son más estables que otras, predominando en el tiempo.

Debido a la enorme complejidad combinatoria y cultural del Go, existe una enorme cantidad de niveles de habilidad. Es una práctica común que si la diferencia de habilidad entre los jugadores es muy grande, el jugador más hábil ofrezca una ventaja al menos hábil para equilibrar las fuerzas relativas de ambos y producir un juego más interesante para ambos. La ventaja se otorga haciendo que el jugador menos

hábil comience primero, colocando más fichas en el tablero. Esta ventaja se la conoce como **handicap**. La cantidad de fichas dependerá de la diferencia de habilidad entre ambos jugadores, y las posiciones de las mismas en el tablero son fijas: no son elegidas por el jugador. Cuando se cree que la cantidad de fichas permite al jugador menos hábil superar al jugador más hábil, se suele ofrecer entonces uno pocos puntos de ventaja al jugador más hábil para alcanzar el “equilibrio perfecto”. Estos puntos se los conoce como **Komi**, y se suele usar un valor no entero (generalmente 6,5).

Saber cuántas piedras debe recibir cada jugador, y cuántos puntos de *komi* son necesarios para alcanzar el equilibrio perfecto, es un problema de difícil solución. En la cultura del Go se desarrolló un sistema de ranking tradicional, que se divide en dos grupos: rankings *kyu* o estudiantes y rankings *dan* o maestros. La idea es que la diferencia de niveles represente la cantidad de piedras de ventaja que un jugador debería darle al otro. El nivel más básico es el 30avo *kyu*. A medida que un jugador mejora, su nivel de *kyu* disminuye hasta llegar a primer *kyu*. Pasado este nivel, el jugador llega a los niveles de maestro, conocidos como rankings *dan*. Éstos van del primer *dan* al séptimo (de manera similar a los sistemas de artes marciales).

Si bien hay 37 niveles tradicionales, se espera que la diferencia de habilidad entre dos jugadores no supere los 9 niveles, que son la cantidad de fichas máximas de ventaja que tradicionalmente se le puede ofrecer a un jugador de menor habilidad. Esto muestra la profunda diferencia de habilidad que se puede producir al interior del Go. Conocer la habilidad real de las personas, que refleje con buena confianza cual es la probabilidad de ganar entre dos personas es un problema que comenzó a resolverse a mediados del siglo pasado para evaluar la habilidad de los jugadores profesionales de ajedrez [Elo08]. Este modelo ha servido como base para los estimadores de habilidad que se utilizan ampliamente en la actualidad en la industria del video juego y la academia [Gli99,HMG06].

Considerar sólo la frecuencia de resultados positivos como indicador de la habilidad de los individuos puede conducir a aproximaciones erróneas, fundamentalmente porque su valor depende también de la dificultad de los desafíos. Por esta razón, todos los estimadores de habilidad ampliamente usados se basan en comparaciones entre pares. Desde los primeros modelos generativos, propuestos hace casi un siglo por [Thu27] y [Zer13], se supone que la probabilidad de un resultado observado r depende del rendimiento p del agente i y de su oponente j , expresada como $P(r | p_i, p_j)$. El campo siguió progresando con los trabajos de [BT52] y [Mos51a, Mos51b, Mos51c], que condujeron al gran avance que tuvo lugar cuando [Elo08] desarrolló una metodología para la Federación de Ajedrez de los Estados Unidos (USCF), adoptada hasta el día de hoy por la Federación Internacional de Ajedrez (FIDE).

A diferencia de la FIDE, que establece a Elo como el método estandarizado para medir la habilidad a lo largo de todo el mundo, la IGF (*International Go Federation*) no establece un sistema de medida estandarizado. Cada asociación nacional decide de manera autónoma qué modelo utilizar. La mayoría de ellas utilizan distintas variantes de Elo. Sin embargo, debido a que el *handicap* y *komi* son factores importantes, los estimadores de Go requieren incorporarlos al modelo.

La Asociación Argentina de Go (AAGo)¹ reúne a las personas aficionadas del juego en nuestro país. A diferencia de las asociaciones de otros deportes, la AAGo cuenta con una cantidad reducida de jugadores, con distinto nivel de compromiso, experiencia y habilidad. La AAGo utiliza el estimadores desarrollado por la Asociación Estadounidense de Go (AGA). En particular, es esperable que en un entorno donde hay pocos jugadores, sean frecuentes las partidas dispares, por lo que el *handicap* estará muy presente. Además, en el caso de que la habilidad esté relacionada a un juego, mientras más información tengamos de un jugador (en este caso, mayor número de partidas), entonces mayor certeza tendremos acerca de su habilidad. Así, en un contexto en el que tenemos acceso a un número pequeño de partidas, alcanzar un grado alto de certeza se dificulta.

Dicho esto, quedan planteados los principales problemas encarados en esta tesis: estimar la habilidad para un juego en un entorno con pocos jugadores y partidas, teniendo en cuenta el efecto de factores externos (como el *handicap* y el *komi*). En este trabajo estudiaremos el modelo usado en Argentina y Estados Unidos; otro más complejo, utilizado en reconocidos sitios web; y uno ampliamente reconocido en la academia, con modificaciones propuestas por nosotros para adaptarlo al Go.

¹<https://go.org.ar/>

1.1. Objetivos

Esta tesis de grado tiene como objetivo principal evaluar distintos modelos matemático-computacionales para estimar la habilidad de jugadores de Go.

Utilizaremos los principales métodos de la bibliografía, teniendo como base el que usa actualmente la Asociación Argentina (AAGo). Buscaremos comprenderlos y evaluarlos para hacer un buen uso y una comparación justa de los modelos, proporcionando en esta tesis una buena documentación para próximos usuarios. Tendremos especial consideración para extender los métodos de forma que contemplen las reglas propias del Go, particularmente el *handicap* y el *komi* de cada partida.

Usaremos la base de datos de partidas de la AAGo, para adaptarnos a su contexto de uso e intentar proveerles una herramienta útil en la práctica.

Capítulo 2

Modelos Probabilísticos

La teoría de la probabilidad funciona como sistema lógico, consistente, con el cual medir la incertidumbre mediante números reales [Cox46]. Las reglas de la probabilidad han sido derivadas formalmente a partir de varios sistemas axiomáticos conceptualmente distintos e independientes entre sí, lo cual es uno de los puntos fuertes a su favor [Hal17]. En este capítulo introduciremos los principales conceptos de la teoría, utilizados como base conceptual y marco teórico para el resto de la tesis.

2.1. Modelos causales y razonamiento probabilístico

La probabilidad, desde un punto de vista bayesiano, puede ser utilizada para modelar situaciones en las cuales contamos con algún grado de incertidumbre. De esta manera, funciona como marco de trabajo para razonar sobre situaciones de sistemas abiertos estudiados en las ciencias empíricas, donde indefectiblemente desconocemos el valor de todas las variables. Para trabajar, nos basamos tanto en evidencia formal (modelos causales) como empírica (datos). Veremos los principales conceptos y reglas de la probabilidad mediante un ejemplo clásico: Monty Hall.

2.1.1. Principio de indiferencia

Supongamos que tenemos tres cajas, y sabemos que dos están vacías y una tiene un regalo.



Figura 2.1: En alguna caja está el regalo. Tenemos incertidumbre y un **espacio de hipótesis**.

Al no poder ver a través de las cajas, tenemos incertidumbre sobre dónde estará el regalo. Para modelar la situación, asignaremos una distribución de creencias. Una opción posible es la mostrada en la Figura 2.2.

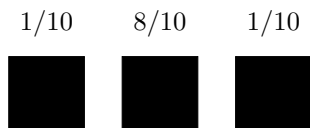


Figura 2.2: Una distribución de creencias posible.

Pero podemos ponernos de acuerdo en que esta distribución no está justificada en tanto no tenemos ninguna evidencia que nos conduzca a tener una preferencia por la caja del medio. En este caso, lo más honesto realmente es utilizar una distribución equiprobable, como muestra la Figura 2.3

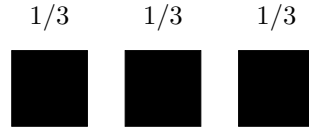


Figura 2.3: Distribución de creencias honesta cuando no tenemos información previa.

Esto se conoce como el **principio de indiferencia**. Si llegamos a un acuerdo respecto a que no tenemos información, entonces estaremos de acuerdo en esta distribución de creencias. Es un principio de validez intercultural y aparece repetidas veces en la historia de la humanidad. En la historia de la teoría de la probabilidad, fue Joseph Bernoulli quien propuso este principio.

Principio de indiferencia
Dividir las creencias en partes iguales

2.1.2. Regla de la suma

El principio de indiferencia puede generalizarse a modelos causales multidimensionales. Para ejemplificar, supongamos que nos permiten elegir una caja c y que después alguien nos señala con certeza que el regalo no está en la caja señalada, s . Es decir, tenemos tres variables y hay dos dependencias causales, la pista no señala ni la posición del regalo ni la caja elegida, que podemos representar con el modelo causal de la figura 2.4. Para simplificar el problema, consideraremos que la caja elegida es $c = 1$.

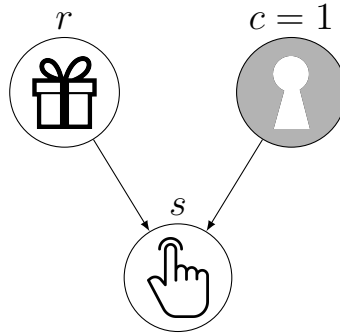


Figura 2.4: La pista depende de la posición del regalo y la de la caja elegida previamente. Los círculos representan las variables y la flecha representa la relación causa \rightarrow efecto.

Cada modelo causal tiene su propio conjunto finito de universos paralelos. Siguiendo con el principio de indiferencia, dividimos la creencia en partes iguales en cada una de las bifurcaciones de los caminos paralelos del modelo causal. En la figura 2.5 mostramos cómo se dividen las creencias. La pista genera dos universos sólo cuando el regalo está en la misma caja que elegimos, $r = c = 1$. Esto nos permite definir creencias honestas respecto de los universos paralelos.

El conjunto de creencias asignadas a los caminos paralelos del modelo causal representa nuestra *creencia conjunta* honesta de que el regalo se encuentre en la caja i y la pista señale la caja j (todas las posibles combinaciones de hipótesis). Cada una de las celdas de la Tabla 2.1 representa una terminal de los caminos del modelo causal representados en la Figura 2.5. Tenemos entonces un principio para alcanzar acuerdos sobre las creencias conjuntas cuando la única información previa es el modelo causal multidimensional.

Habiendo definido la creencia conjunta sobre hipótesis mutuamente excluyentes (universos paralelos), con la regla de la suma podemos calcular la creencia correspondiente a cada una de las variables individuales. Para calcular la creencia sobre una variable, calculamos la suma de todos los caminos paralelos en los que esa variable está presente.

En términos generales, la regla de la suma permite que cualquier distribución marginal pueda ser obtenida integrando la distribución conjunta:

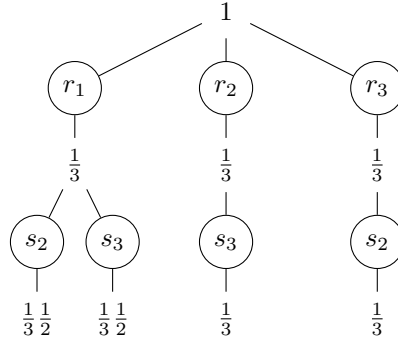


Figura 2.5: Los caminos paralelos que se generan a partir de los modelos causales alternativos. El principio de máxima incertidumbre (o de indiferencia) divide la creencia en partes iguales en cada una de las bifurcaciones de los caminos paralelos del modelo causal.

$$P(r = i, s = j)$$

	r_1	r_2	r_3
s_1	0	0	0
s_2	1/6	0	1/3
s_3	1/6	1/3	0

Tabla 2.1: Creencia conjunta para el modelo causal. El valor de las celdas representa la creencia honesta sobre cada uno de los universos paralelos.

	r_1	r_2	r_3	
s_1	0	0	0	0
s_2	1/6	0	1/3	1/2
s_3	1/6	1/3	0	1/2
	1/3	1/3	1/3	1

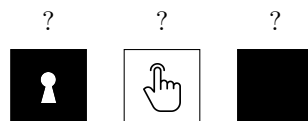
Tabla 2.2: Creencia marginal obtenida como la suma de las creencia conjuntas honesta de cada uno de los caminos paralelos de los modelos causales.

$$P(x) = \sum_y P(x, y) \quad \text{ó} \quad p(x) = \int p(x, y) dy \quad (\text{regla de la suma})$$

donde $P(\cdot)$ y $p(\cdot)$ representan distribuciones de probabilidad discretas y continuas, respectivamente.

2.1.3. Regla del producto

¿Pero cómo preservar los acuerdos intersubjetivos cuando recibimos nueva información? El principio de coherencia establece que la nueva creencia honesta, después de haber visto un nuevo dato, es la creencia previa que sigue siendo compatible con ese dato. Supongamos que alguien nos indica con certeza que el regalo no está en la caja del medio. ¿Cuál es la nueva creencia marginal honesta sobre el regalo después de haber visto esta pista?



Para actualizar las creencias posiblemente nos veamos tentados a aplicar el principio de indiferencia nuevamente, asignando 0,5 a las dos cajas restantes. Aunque esa propuesta coincida con la solución

correcta en algunas situaciones, esa decisión conduce a errores en términos generales. El principio de indiferencia sólo se aplica una única vez, al inicio. Después sólo actualizaremos esa creencia en función de la nueva información que vayamos incorporando.

Para actualizar las creencias simplemente nos quedamos con la creencia previa que es compatible con el dato. Para ello nos quedamos con los caminos paralelos del modelo causal que siguen siendo compatibles con el observable, $P(r_i, s_2)$.

$$P(r = i, s = 2)$$

	r_1	r_2	r_3	
s_2	1/6	0	1/3	1/2

Tabla 2.3: La creencia conjunta y marginal que sobrevive luego de ver los datos.

En la Tabla 2.3 nos quedamos con la creencia conjunta y marginal que sigue siendo compatible con el dato. Esto es exactamente lo mismo que quedarnos con los caminos del modelo causal que son posibles dado el modelo causal y el observable. En la Figura 2.6 mostramos los caminos que son compatibles en negro y los incompatibles en gris.

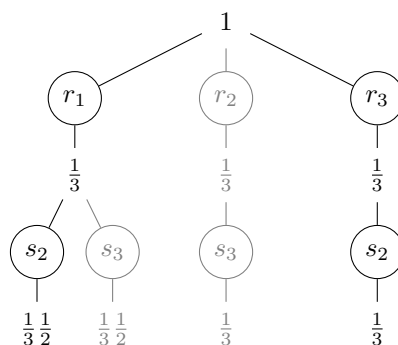


Figura 2.6: Los caminos paralelos compatibles (negro) e incompatibles (gris) con el dato $s = 2$.

Es decir, para actualizar nuestra creencia nos quedamos con la creencia a priori que es compatible con los datos. Lo que nos permitirá cumplir el objetivo que nos habíamos propuesto, actualizar la creencia sobre el regalo luego de haber visto la pista. **La creencia que sobrevive**, $P(r_i, s_2)$, es ahora nuestra nueva creencia total. Para expresarla nuevamente como tal, la normalizamos para que vuelva a sumar 1. Luego de este procedimiento, se mantiene la propiedad de ser la distribución de creencias que maximiza la incertidumbre, dada la evidencia formal (modelo causal) y empírica (datos). En otras palabras, se conserva el acuerdo intersubjetivo, si el modelo causal y los datos fueran ciertos.

$$P(r_i|s_2) = \frac{P(r_i, s_2)}{P(s_2)} \quad (2.1)$$

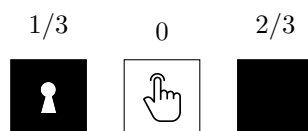


Figura 2.7: Nueva creencia honesta dada el dato y el modelo

Generalizada, la regla del producto señala que cualquier distribución conjunta puede ser expresada como el producto de distribuciones condicionales unidimensionales.

$$p(x, y) = p(x|y)p(y) \quad (\text{regla del producto})$$

De estas reglas, se deduce el Teorema de Bayes:

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)} \quad (\text{Teorema de Bayes})$$

Aplicada de manera sistemática, la teoría de la probabilidad no es más que la herramienta para filtrar la creencia previa que sigue siendo compatible con la evidencia empírica y formal. Teniendo esto en mente, se llega a la interpretación del Teorema de Bayes:

$$\overbrace{P(\theta|D, M)}^{\text{creencia a posteriori}} = \frac{P(D|\theta, M)P(\theta|M)}{\int P(D|\theta, M)P(\theta|M)d\theta} = \frac{\overbrace{P(D|\theta, M)}^{\text{verosimilitud}} \overbrace{P(\theta|M)}^{\text{creencia a priori}}}{\underbrace{P(D|M)}_{\text{evidencia}}} \quad (2.2)$$

donde θ es el espacio de hipótesis, D es el conjunto de datos observados y M es el modelo causal asumido. Aunque muchas veces no se explicita, siempre debemos recordar que estamos asumiendo un modelo.

Nótese que llamamos **evidencia** al denominador: $P(D|M)$. Como veremos más adelante, ésta refiere a la probabilidad a priori de obtener los datos observados que se puede calcular integrando el producto de la **verosimilitud** $P(D|\theta, M)$ y **creencias a priori** $P(\theta|M)$ sobre el espacio de hipótesis (θ).

2.2. Evaluación de modelos

Al trabajar con distintos modelos, vamos a querer compararlos para saber cuál se aproxima mejor a la realidad. En este capítulo describiremos la metodología bayesiana basada en probabilidad, usando la evidencia de los modelos. Esta forma introduce un factor que penaliza modelos más complejos de lo necesario, mediante un término conocido como Factor de Occam, que surge naturalmente de las reglas de la probabilidad. Esto evita sobre ajustes, sin necesidad de acudir a penalizadores ad-hoc. Este marco de trabajo fue introducido por Jeffreys [Jef39], con continuaciones de Gull [Gul88], Kass y Raftery [KR95], y MacKay [Mac91] [Mac99].

Inicialmente, vamos a considerar dos niveles de inferencia:

- En el primer nivel, suponemos que el modelo con el que trabajamos es el verdadero, y ajustamos nuestro modelo a los datos
- En el segundo nivel, tenemos la comparación de modelos, en la cual usamos la evidencia del primer nivel

En el primer nivel, suele haber un espacio de hipótesis sobre el cual queremos aprender utilizando los datos:

$$P(\theta|D, M) = \frac{P(D|\theta, M)P(\theta|M)}{\int P(D|\theta, M)P(\theta|M)d\theta} = \frac{P(D|\theta, M)P(\theta|M)}{P(D|M)} \quad (2.3)$$

Notamos que en 2.3, la evidencia es la **probabilidad de los datos, dado que el modelo M es el verdadero**: $P(D|M)$.

El segundo nivel de inferencia utiliza la evidencia del primer nivel para calcular la probabilidad del modelo, dados los datos:

$$P(M_i|D) = \frac{P(D|M_i)P(M_i)}{\sum_k P(D|M_k)P(M_k)} \quad (2.4)$$

En el caso de que no tengamos información previa que nos haga preferir un modelo por sobre el resto, $P(M)$ será equiprobable, por lo que elegir el modelo con mayor probabilidad equivale a quedarse con el modelo de mayor evidencia. De aquí podemos derivar el concepto de Factor Bayesiano [KR95]:

$$\begin{aligned}
\log_e \left(\frac{P(M_i|D)}{P(M_j|D)} \right) &= \log_e \left(\frac{P(D|M_i)P(M_i)}{P(D|M_j)P(M_j)} \frac{\sum_k P(D|M_k)P(M_k)}{\sum_k P(D|M_k)P(M_k)} \right) \\
&= \log_e \left(\frac{P(D|M_i)}{P(D|M_j)} \right) \\
&= \log_e(P(D|M_i)) - \log_e(P(D|M_j))
\end{aligned} \tag{2.5}$$

que resume en un número (en escala logarítmica) la evidencia provista por los datos a favor de un modelo o teoría por sobre otra.

Por la regla del producto, podemos descomponer la evidencia en la probabilidad de cada dato, dado el modelo y los datos anteriores (ecuación (2.6)). En nuestro contexto, esto se traduce en calcular la probabilidad a priori de cada partida, dado que el método conoce el resultado de las partidas anteriores. Es destacable el uso de la productoria, ya que una sola partida con probabilidad 0 provoca que la evidencia sea 0.

$$P(D|M) = \prod P(d_i|M, d_1, \dots, d_{i-1}) \tag{2.6}$$

Como la evidencia es un número pequeño y poco interpretable, en múltiples ocasiones usaremos el promedio geométrico de las evidencias individuales de las partidas. En el contexto de un proceso multiplicativo, como es el caso de la evidencia, el promedio geométrico es el correcto a usar (a diferencia del promedio aritmético clásico, que es usado generalmente).

Otra forma de entender a la evidencia es verla como verosimilitud marginal (ecuación (2.7)). Es decir, la evidencia es la verosimilitud de los datos, pero integrando todas las hipótesis con las que trabajamos, ponderadas por la probabilidad a priori de estas hipótesis. Este es un aspecto fundamental de la evidencia como métrica bayesiana: en vez de quedarse con una sola hipótesis (como suele hacerse al maximizar verosimilitud), hace uso de todas en simultáneo.

$$P(D|M) = \int P(D, \theta|M) d\theta = \int P(D|\theta, M) P(\theta|M) d\theta \tag{2.7}$$

Esto debe verse acompañado de un principio de honestidad intelectual. Uno podría mejorar la evidencia dejando de lado las hipótesis que peor explican los datos, pero sería un mal uso de la teoría de la probabilidad. Además, esto puede fomentar el sobre-ajuste, problema común en los métodos que solamente maximizan verosimilitud.

En cambio, al integrar todas las hipótesis, agregar variables que no son necesarias para explicar los datos se ve penalizado sin necesidad de agregar costos ad-hoc. A este penalizador se lo denomina Factor de Occam, y surge naturalmente de la aplicación de las reglas de la probabilidad, tanto en casos de regresión como clasificación [Mac91]. Para ejemplificar, veamos un caso clásico de regresión: supongamos que necesitamos ajustar una función sinusoidal mediante modelos polinomiales de diferente grado.

A medida que aumentamos el grado del polinomio, nuestro modelo adquiere mayor flexibilidad y por lo tanto es capaz de acercarse más y más a los datos. Podemos ver los resultados con polinomios de distintos grados en la Figura 2.8. Si bien en la figura solo se muestran los valores máximos a posteriori, estos modelos fueron ajustados mediante inferencia bayesiana, haciendo uso de todas las hipótesis a priori.

El criterio de evaluación que se suele usar en estadística frecuentista es minimizar las distancias, que es equivalente a maximizar verosimilitud. Aumentando la complejidad de los modelos, siempre podemos reducir más y más las distancias a los datos. Sin embargo, cuando evaluamos las predicciones a priori (evidencia) conjunta de los distintos modelos polinomiales, vemos que el mejor modelo tiene una complejidad intermedia. En este caso, como se puede ver en la Figura 2.9, el modelo polinomial de grado 3 resulta ser claramente superior al resto. Este modelo tiene la particularidad de tener la flexibilidad mínima necesaria según la función objetivo que genera los datos que estamos analizando.

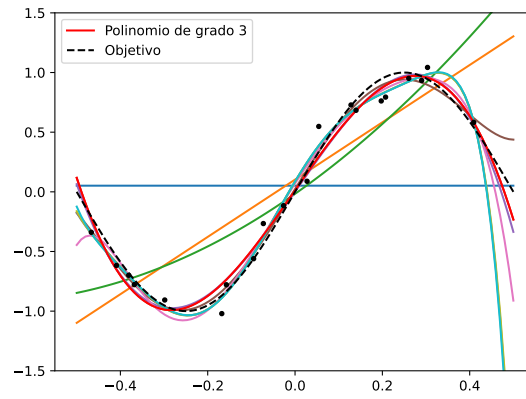


Figura 2.8: La función objetivo (línea punteada) es una sinusoidal. Los modelos alternativos analizados son polinomios de grado 0 a 9, ajustados mediante inferencia bayesiana (en el gráfico no se muestra la incertidumbre, solo el máximo a posteriori). Los puntos son tomados aleatoriamente de la función objetivo, sumándoles un ruido normal con desvío estándar de 0,1.

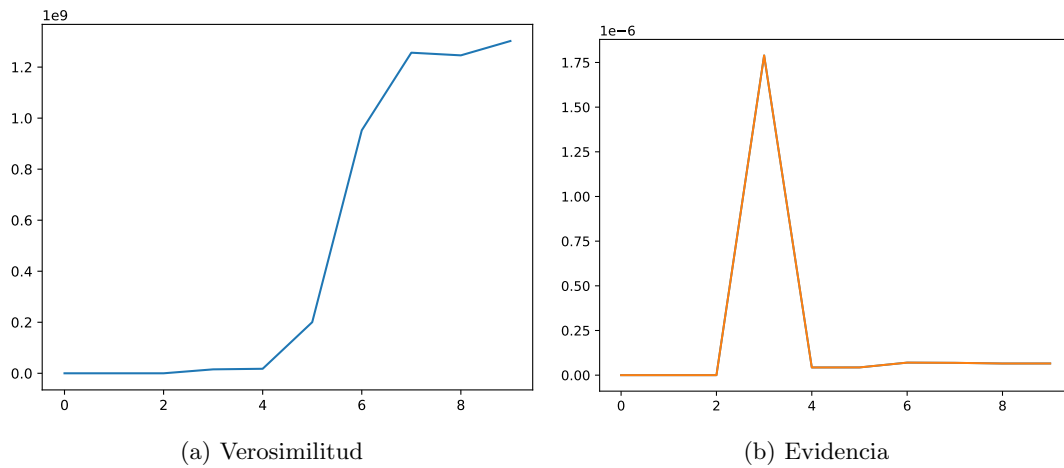


Figura 2.9: Predicción a priori de los modelos polinomiales alternativos

Esto tiene que ver con una balance natural que se produce por el hecho de que las predicciones son distribuciones de probabilidad y por lo tanto siempre tienen que integrar 1. A medida que se suman más dimensiones, los modelos se ven obligados a distribuir esa misma integral en un espacio cada vez más grande. Por ejemplo, supongamos que conocemos el valor en el eje X. En la Figura 2.10 mostramos dos distribuciones de probabilidad sobre el eje Y (predicciones). Los modelos más rígidos van a proponer distribuciones de probabilidad con menos incertidumbre que los modelos más flexibles, que admiten muchas más posibilidades. Si el dato aparece en la región donde apuntan los modelos más simples, entonces les va a ir mejor que a los más complejos, y viceversa. El modelo de grado 3 tiene la característica de ser el modelo más simple capaz de ajustar correctamente la posición de esa distribución a medida que cambiamos el valor en el eje x. Los modelos con grado mayor a 4 también pueden ajustar la posición a medida que cambiamos el eje X, pero la incertidumbre suele ser más grande que la de grado 3, por lo que pierden frente a ellos. Los modelos con grado menor a 3 tienen menos incertidumbre, pero no pueden adaptar la posición a medida que cambiamos el eje X. Así, la aplicación estricta de las reglas de la probabilidad garantiza que la evaluación de modelo favorezca a aquellos con la flexibilidad mínima necesaria.

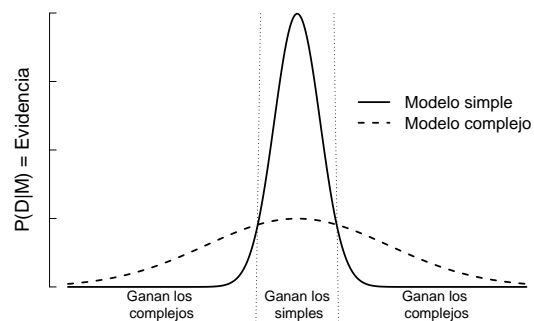


Figura 2.10: Ejemplo de predicción con un modelo más simple y uno más complejo. Los modelos más rígidos van a proponer distribuciones de probabilidad con menos incertidumbre que los modelos más flexibles que admiten muchas más posibilidades. Si el dato aparece en la región donde apuntan los modelos más simples, entonces les va a ir mejor que a los más complejos, y viceversa.

Capítulo 3

Modelos de habilidad

3.1. Modelo general de eventos

La gran mayoría de modelos de estimación de habilidad conceptualizan cuatro variables: habilidad, desempeño, diferencia y resultado. Mientras la habilidad, el desempeño y la diferencia son variables latentes, el resultado es observado. Este modelo puede verse expresado en el modelo causal de la Figura 3.1.

Como no tenemos razones para considerar que un jugador va a rendir por encima o por debajo de su habilidad real, el desempeño es modelado como una variable gaussiana centrada en la habilidad. El parámetro β regula la incertidumbre esperada, aunque en la práctica sirve como escala del modelo. El resultado de un partido queda determinado por la diferencia entre los desempeños: el jugador con mayor desempeño gana.

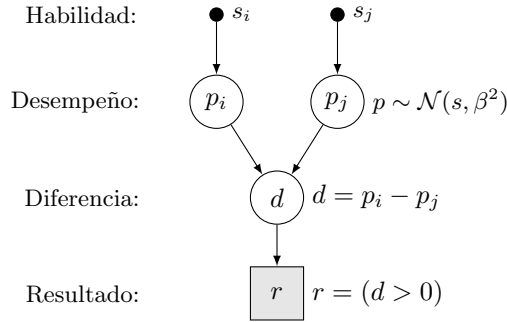


Figura 3.1: Modelo causal de eventos donde el resultado observable se considera como consecuencia directa de los desempeños, variables aleatorias centradas en la verdadera habilidad.

A diferencia de Elo, que modela las habilidades como un número real puntual, los modelos estudiados en esta tesis agregan la incertidumbre alrededor de la estimación, de forma que las habilidades quedan dadas por variables gaussianas.

3.1.1. Inferencia exacta

Utilizando las reglas de la probabilidad y las propiedades de las distribuciones gaussianas se puede calcular [LM21] la evidencia exacta (probabilidad a priori del resultado) y la creencia a posteriori de la habilidad al ganar una partida quedan, dadas por las ecuaciones (3.1) y (3.2) respectivamente.

$$\underbrace{P(r)}_{\text{Evidencia}} = 1 - \Phi\left(0 \mid \underbrace{\mu_i - \mu_j}_{\text{Diferencia esperada}}, \underbrace{2\beta^2 + \sigma_i^2 + \sigma_j^2}_{\text{Incertidumbre total}}\right) \quad (3.1)$$

$$\underbrace{p(s_i | r)}_{\text{Posterior}} \propto \underbrace{1 - \Phi(0 | s_i - \mu_j, 2\beta^2 + \sigma_j^2)}_{\text{Verosimilitud } P(r|s_i)} \underbrace{\mathcal{N}(s_i | \mu_i, \sigma_i^2)}_{\text{Prior } p(s_i)} \quad (3.2)$$

El posterior normalizado se obtiene dividiendo el lado derecho con la evidencia, $P(r)$. Es interesante notar las similitudes y diferencias entre la verosimilitud y la evidencia. La evidencia es la acumulada de una distribución gaussiana que está centrada en la diferencia de las estimaciones medias, con una varianza que incluye todas las incertidumbres. La verosimilitud cuantifica la misma densidad acumulada que la evidencia, pero centrada ahora en la diferencia entre la hipótesis que estamos evaluando s_i y la estimación media del oponente μ_j , con una varianza que incluye todas las incertidumbres salvo la de la propia hipótesis s_i .

El posterior no es más que la densidad del prior no filtrada por la verosimilitud. La sorpresa, defi-

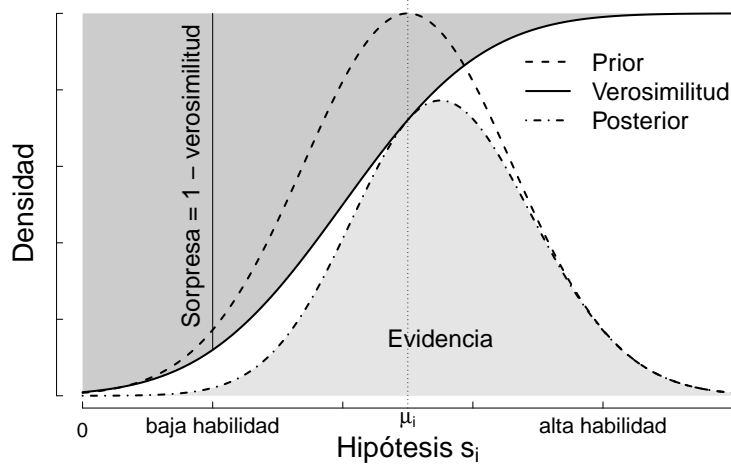


Figura 3.2: Actualización de creencias para el caso ganador. El posterior proporcional se obtiene como el producto de la distribución a priori (distribución gaussiana) y la verosimilitud (distribución gaussiana acumulada). La evidencia es la integral del posterior proporcional. Las distribuciones no están necesariamente en la misma escala: la distribución a priori integra 1, mientras que la verosimilitud va de 0 a 1.

nida como el complemento de la verosimilitud, funciona como un filtro para el prior. En la región de hipótesis de muy alta habilidad, donde el resultado ganador no nos hubiera generado casi ninguna sorpresa ($\lim_{s_i \rightarrow \infty} P(r|s_i) = 1$), el posterior recibe casi toda la densidad del prior. En cambio, en la región de hipótesis de muy baja habilidad, donde el resultado habría generado mucha sorpresa ($\lim_{s_i \rightarrow -\infty} P(r|s_i) = 0$), el posterior no recibe casi nada de la densidad del prior.

Es importante remarcar que la posterior, aunque se parezca, no es una distribución gaussiana, lo que nos impedirá usar la ecuación (3.2) de forma exacta cuando propaguemos esa información a otros eventos. Todos los modelos que analizamos aquí aproximan el posterior exacto mediante una distribución gaussiana. Las técnicas de aproximación son distintas. Además existen diferentes formas de modelar la serie de eventos.

3.2. Modelos de múltiples eventos

Otro punto a considerar es el cambio de la habilidad a través del tiempo. Los modelos estudiados consideran un proceso de Wiener. Esto se traduce una dinámica sin memoria (o markoviana, ecuación 3.3) con variables gaussianas (ecuación 3.4). γ^2 es un hiperparámetro que regula cuánta incertidumbre se agrega por cada unidad de tiempo.

$$P(s_{t+k+m} | s_t, s_{t+k}) = P(s_{t+k+m} | s_{t+k}) \quad (3.3)$$

$$s_{t+k}|s_t \sim \mathcal{N}(s_t, k \cdot \gamma^2) \quad (3.4)$$

Este modelo supone tiempo discreto. En este trabajo, discretizamos el tiempo en días. Notar que puede haber más de una partida de un mismo jugador por día t . En este caso, se utiliza la misma variable aleatoria s_t . Esto implica que las partidas que se juegan un mismo día no tienen un orden entre sí, se considera que entran en una misma tanda o *batch*.

3.2.1. Modelos de historial completo

Algo importante de destacar es que algunos modelos, como el de AGA/AAGo, difunden la información en un solo sentido: del pasado hacia el futuro. Cada vez que obtenemos los resultados de un torneo, se actualizan las estimaciones de la habilidad de los jugadores participantes, utilizando esta información para un evento futuro. Los *posteriors* de un torneo son los *priors* del siguiente (actualizados por el modelo temporal, que agrega incertidumbre). Sin embargo, esto es una simplificación de la teoría de la probabilidad. Un uso correcto de las reglas de la suma y el producto implican también difundir la información en el sentido contrario: del futuro al pasado. Esto produce varios problemas:

- las estimaciones de los primeros torneos de un jugador resultan malas, sin importar la cantidad de torneos que juegue luego.
- se pierde la posibilidad de comparar las habilidades de dos momentos históricos distintos.
- las estimaciones, al contar con menor información, aumentan su incertidumbre

Estos problemas son encarados por los modelos de **historial completo**. Por ejemplo, supongamos que tenemos dos partidas y dos jugadores que se ganan mutuamente inmediatamente uno después del otro. Al ganar una partida cada jugador, no hay evidencia de que alguno de los dos jugadores sea mejor que el otro. En la Figura 3.3 mostramos en rojo las estimaciones de los modelos que propagan la información del pasado al futuro, y en verde las que propagan la información en todas las direcciones.

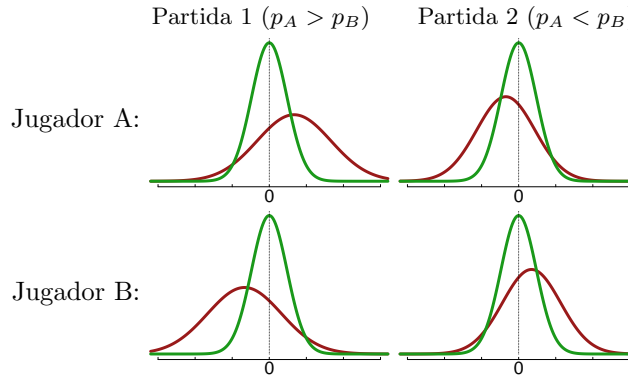


Figura 3.3: Habilidades estimadas con dos partidas. En la primera partida, gana el jugador A. En la segunda, gana el jugador B. En rojo se muestran las estimaciones de un modelo que solamente propaga información del pasado al futuro, y en verde se muestran las estimaciones finales de un modelo de historial completo. La primera columna muestra las estimaciones del primer partido, mientras que la segunda muestra las estimaciones del segundo partido.

Los modelos que propagan la información del pasado al futuro muestran varias limitaciones. El primero es que las estimaciones de la primera partida no utilizan la información del futuro, por lo que se considera que el jugador A debe ser mejor que el jugador B sin importar que luego pierda contra éste. El segundo es que no utiliza la información de forma correcta: uno esperaría que luego de la segunda partida, las estimaciones de los dos jugadores sean las mismas. Sin embargo, eso no se cumple. Por último, al no usar de forma óptima la información disponible, las estimaciones contienen mayor incertidumbre.

Estudiamos dos estimadores que utilizan de formas distintas la información de toda la historia de partidas: Whole History Rating (3.2.3) y TrueSkill Through Time (3.2.4).

3.2.2. Modelo AGA/AAGo

El modelo utilizado por la AAGo está basado en el sistema utilizado por la American Go Association (AGA) [Wal10]. Estudiaremos el segundo para luego presentar las adaptaciones realizadas por la asociación argentina.

Las distribuciones de creencias sobre las habilidades de las personas también son descritas mediante distribuciones normales, centradas en μ y con desviación estándar σ . En vez de actualizar las estimaciones luego de cada partida, la American Go Association espera hasta el final de cada torneo para actualizar los valores con la información nueva. El método de aproximación de los posteriors exactos usado por la AGA utiliza el valor que maximiza el posterior como media de las nuevas estimaciones y agrega incertidumbre mediante otro estimador conocido como matriz de información de Fisher.

Como hemos visto, la verosimilitud de cada partido está dada por las habilidades de los jugadores. Pero en el modelo de AGA depende además del efecto de la ventaja (*handicap* y *komi*) de la partida j , d_j .

$$P(\text{gana jugador blanco} | s_{\text{blanco}}, s_{\text{negro}}) = \Phi\left(\frac{r_{\text{blanco}} - r_{\text{negro}} - d_j}{\sigma_{px}}\right) \quad (3.5)$$

En el sistema AGA, el valor de d_j depende de forma lineal positiva de la cantidad de *handicap* (ya que favorece al jugador negro), y de forma lineal negativa de la cantidad de *komi* (favoreciendo al jugador blanco). El modelo AGA establece una tabla de constantes para las diferentes combinaciones de *handicap* y *komi* para determinar el valor de d_j . Algo a destacar es que con *handicap* = 0 y *komi* = 0 el valor de d_j es positivo: se supone que el jugador negro tiene ventaja por jugar primero. Por otra parte, el σ_{px} representaría el desvío de la distribución normal de diferencia de desempeños.

Diferencia de habilidad (escala AGA/AAGo)	Probabilidad de ganar
0,1	0,5374
0,2	0,5744
0,5	0,6806
0,75	0,7593
1	0,8261
2	0,9698

Tabla 3.1: Ejemplos numéricos de la verosimilitud, cuando no hay *handicap* ni *komi*, y sin tener en cuenta el factor d_j para que la función sea simétrica.

En la sección anterior hemos mostrado cómo se calcula la incertidumbre de la diferencia de desempeños, y hemos visto que depende de la incertidumbre de las estimaciones previas. Sin embargo, debido que la AGA decide maximizar el posterior conjunto (reemplaza todas las estimaciones por las hipótesis), la incertidumbre σ_{px} no contiene las incertidumbres de las estimaciones previas. Por lo tanto, la incertidumbre de la diferencia de desempeños depende entonces del valor de base (dado por el parámetro β en el modelo causal) y varía solo en función de la cantidad de *handicap* y *komi*.

$$\sigma_{px} = \begin{cases} \text{incertidumbre_de_base} + \text{incertidumbre_por_komi}(k) & \text{si } h \in \{0, 1\} \\ \text{incertidumbre_de_base} + \text{incertidumbre_por_handicap_y_komi}(h, k) & \text{sino} \end{cases} \quad (3.6)$$

Cuando la partida no tiene *handicap*, $h \in \{0, 1\}$, la incertidumbre de la diferencia de desempeños solo aumenta en función del *komi*. Cuando la partida tiene *handicap*, la incertidumbre tanto aumenta en función del *handicap* como del *komi* (ver artículo original [Wal10]).

En el artículo citado, el posterior conjunto de un torneo resulta proporcional a la función $L(s_1, \dots, s_n)$, que es definida como el producto de todas las distribuciones a priori, $P(s_i)$, y las verosimilitudes de todas las partidas (3.5). Los valores medios de las nuevas distribuciones de creencias serán los valores que maximizan el posterior. A esta aproximación se la conoce como **Máximo A Posteriori** (MAP).

$$\arg \max_{s_1, \dots, s_n} L(s_1, \dots, s_n) = \arg \max_{s_1, \dots, s_n} \prod_i^{\text{jugadores}} P(s_i) \cdot \prod_j^{\text{partidas}} P(r_j | s_{\text{ganador}_j}, s_{\text{perdedor}_j}) \quad (3.7)$$

Los priors son las últimas estimaciones con incertidumbre agregada debido al paso del tiempo (el proceso de Wiener explicado en la sección 3.1, con $\gamma = 0,0005$). Para resolver la maximización primero se aplica el logaritmo, que preserva el máximo pero convierte la productoria en su equivalente sumatoria de logaritmos. La conveniencia de estas transformaciones reside en el bajo valor que daría una productoria de (potencialmente) miles de términos menores a 1. Tan bajos valores son muy vulnerables a errores de redondeo, lo que es preferible evitar. Para resolverlo se utiliza el método del gradiente conjugado de Fletcher-Reeves, teniendo el algoritmo simplex como alternativa de respaldo.

Por otro lado, para obtener las nuevas desviaciones estándar, se calcula el Hessiano del posterior. Si el posterior cambia mucho en el entorno del máximo, es decir si es muy picuda, entonces hay bastante certeza de que la hipótesis está cerca del máximo. Si el posterior cambia poco en el entorno del máximo, entonces cualquiera de las hipótesis cercanas son suficientemente buenas como la seleccionada. Este concepto es implementado por la matriz de información de Fisher. La inversión de matrices (necesaria para Fisher), los algoritmos de minimización y otros cálculos se resuelven con métodos de GSL (GNU Scientific Library).

Categorías declaradas, uso del software y código

Debido a que los sistemas AGA y AAGo tienen algunas dificultades para estimar correctamente las estimaciones, cada vez que una persona se anota a un torneo se le pide que declare su habilidad autopercebida (en términos de *kyu*/dan). En caso de ser el primer torneo del jugador, esta categoría es utilizada para definir la distribución de habilidad a priori. Pero en caso de que ya exista una estimación previa en el sistema, la declaración de categoría permite a los jugadores corregir las estimaciones que hace el sistema. La autopromoción solo se hace efectiva si el jugador demuestra con sus resultados durante el torneo que este aumento de habilidad es real (según un conjunto de reglas definidas por AGA).

El sistema de la AAGo hace dos modificaciones sobre este modelo. La primera es un ajuste de escala en el komi, de una unidad, por la diferencia entre las reglas chinas y japonesas. La segunda agrega restricciones para aceptar una auto-promoción. La asociación argentina agrega condiciones más estrictas para aceptar la categoría declarada. En particular, requiere que el jugador gane al menos una partida contra un oponente que sea a lo sumo una categoría peor que la categoría a la cual se quiere autopromocionar.

El código fuente del software utilizado por la AAGo se encuentra disponible en GitHub¹ y es una modificación del que utiliza AGA. Tener disponible este código nos permitió estudiarlo y profundizar nuestra comprensión de su funcionamiento. Además, hizo posible probar modificaciones del mismo (de las cuales hablamos en la sección 5.2).

El sistema consta de un programa cuya entrada son los datos de un torneo:

- la lista de jugadores que participaron, con sus estimaciones a priori, la cantidad de días desde el último torneo en el que participaron, y la categoría declarada
- la lista de partidos, con los jugadores correspondientes, la cantidad de piedras de *handicap*, el *komi* y el resultado del partido

La salida del programa es la lista de jugadores, con la distribución a posteriori de sus habilidades.

3.2.3. Whole History Rating

El modelo Whole History Rating (WHR) [Cou08] se presenta como una mejora a Elo. Es utilizado para comparar la habilidad de jugadores profesionales de Go a nivel internacional², además de ser el sistema de ranqueo de varios servidores de distintos juegos en línea³.

A diferencia del modelo general que hemos mostrado al principio, en el que la verosimilitud se describe como la acumulada de una distribución normal, en WHR siguiendo la tradición del modelo Elo original, describe la verosimilitud de los eventos utilizando una función logística. Ambas funciones tienen la misma

¹<https://github.com/elsantodel90/RAAGo>

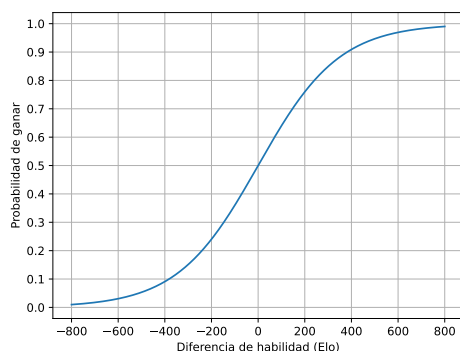
²Accesible en <https://www.goratings.org/en/>

³Puede encontrarse una lista en <https://www.remi-coulom.fr/WHR/>

forma de tipo S, que comienza en 0 y sube hasta alcanzar el 1. Sin embargo la forma de parametrizar las funciones difiere.

$$P(r|s_{ganador}, s_{perdedor}) = \frac{1}{1 + 10^{-\frac{(elo_{ganador} - elo_{perdedor})}{400}}} \quad (3.8)$$

Notar que esta es una verosimilitud, y por lo tanto no tiene en cuenta la incertidumbre de las estimaciones de las habilidades. El modelo WHR, siguiendo la tradición de Elo, utiliza su escala, muy conocida especialmente en el ambiente de ajedrez. El valor 400 tiene que ver con la incertidumbre de base que en el modelo general había sido modelado con el parámetro β . En esta escala, la probabilidad del resultado de una partida queda dada por una función logística (ecuación (3.8)). En la Tabla 3.4(b) mostramos valores concretos en función de la diferencia de habilidad entre los jugadores $elo_{ganador} - elo_{perdedor}$.



(a) Verosimilitud logística

Diferencia de habilidad (Elo)	Probabilidad de ganar
10	0,514
20	0,528
50	0,571
100	0,64
200	0,76
400	0,91

(b) Ejemplos numéricos de la verosimilitud logística

Figura 3.4: La verosimilitud o probabilidad de ganar una partida en función de la diferencia de habilidad es modelada como una función logística en vez de con una acumulada de una normal

Para calcular la evidencia, se debe integrar la probabilidad conjunta con respecto a las habilidades.

Los modelos de clasificación basados en distribuciones normales, como WHR, no cuentan con una solución analítica exacta y tratable para la distribución a posteriori. Existen distintas aproximaciones para encarar este problema. WHR decide utilizar la aproximación de Laplace, que consiste en buscar numéricamente el Máximo A Posteriori (MAP) y estimar la incertidumbre alrededor utilizando el Hessiano. Es un método **local al máximo**, pero apropiado cuando la distribución a posteriori real es similar a una normal (por ejemplo, por teorema central del límite). Como método numérico para encontrar el Máximo A Posteriori, utiliza el método de Newton para buscar una raíz en el gradiente.

El modelo fue implementado originalmente en Ruby⁴ para su uso en el servidor GoShrine, con una posterior conversión a Python⁵. En esta tesis utilizaremos la versión hecha en Python.

El paquete provee un objeto al cual se le agregan partidas. Por cada partida, se indica el día en que se juega (como número entero), el identificador de cada jugador, el ganador y el efecto del *handicap* (como número flotante, en escala de Elo). Este último es agregado a la habilidad del jugador negro al calcular la verosimilitud de una partida (ecuación (3.8)). Además, tiene métodos para hacer iteraciones del método de Newton durante un tiempo, o hasta que converja.

Dos consideraciones son destacables de esta interfaz: no permite asignar *priors* a los jugadores (de hecho, no aparecen de forma explícita en todo el código), y no modela la transformación de número de piedras de *handicap* en una partida a la habilidad en Elo que estas agrega. Lo primero resulta inconveniente cuando se tiene información de los jugadores más allá de las partidas, además de imposibilitar ciertos experimentos. Utilizando el sistema, se puede ver que la habilidad a priori calculada para un jugador

⁴Repositorio público: https://github.com/goshrine/whole_history_rating

⁵Repositorio público: https://github.com/pfmonville/whole_history_rating

sin partidas es una normal con media 0,0 y desvío estándar 245,4. Lo segundo requiere un modelado extra por parte del usuario. Decidimos modelar el efecto del *handicap* de forma lineal, teniendo dos hiperparámetros: uno para la pendiente y otro para la ordenada al origen.

3.2.4. TrueSkill Through Time

TrueSkill Through Time (TTT) es una mejora al modelo TrueSkill, desarrollado por Microsoft. Al igual que WHR, usa el historial completo de partidas para estimar las habilidades de los jugadores en cada momento, pero usa una técnica distinta.

El modelo TrueSkill Through Time (TTT) [DHMG07] implementa un enfoque de *smoothing* que realiza la inferencia usando una única red bayesiana que incluye todos los eventos históricos, lo que permite que la información se propague por todo el sistema, garantizando estimaciones iniciales fiables y comparabilidad entre estimaciones distantes. Esta red bayesiana se basa en la conexión de redes menores, una por cada partida, utilizando el modelo causal explicado en la Figura 3.1, utilizando las reglas del proceso de Wiener definido en las ecuaciones 3.4 y 3.3.

El éxito de la solución TrueSkill [HMG06] se basa en la aplicación de un método eficiente para calcular la gaussiana que mejor aproxima al posterior exacto, en términos de minimización de la divergencia Kullback-Leibler entre la distribución verdadera y la aproximada [Min01].

$$\hat{p}(s_i|r, s_j) = \arg \min_{\mu, \sigma} \text{KL}(p(s_i|r, s_j) || \mathcal{N}(s_i|\mu, \sigma^2)) \quad (3.9)$$

El uso de una red bayesiana permite, bajo ciertas condiciones, realizar inferencia de manera exacta mediante el método de pasaje de mensajes y el algoritmo de la suma-producto [KFL01]. Este algoritmo no hace más que aplicar las reglas de la probabilidad de forma local sobre cada nodo de la red bayesiana, mediante programación dinámica, haciendo que sea muy eficiente (lineal en la cantidad de ejes del grafo). Sin embargo, la presencia de ciclos, sumada a la aproximación de los posteriors, hace que sea necesario hacer varias iteraciones de pasaje de mensajes hasta que el sistema converja.

Otra ventaja de usar un modelo gráfico de forma explícita en el proceso de inferencia es que nos permite agregar nuevos factores a tener en cuenta de una manera simple y cumpliendo con las reglas de la probabilidad. TrueSkill fue ideado para poder usarse en deportes con equipos, asumiendo que el rendimiento de un equipo será la suma de los rendimientos de sus jugadores: $t = \sum_i w_i \cdot p_i$, donde p_i son los rendimientos de los jugadores de los equipos y w_i es un multiplicador que puede ser configurado para cada partida. Esto puede ser aprovechado para agregar otros factores, como en nuestro caso el *handicap* y *komi*.

Contamos con una implementación abierta creada en nuestro laboratorio para Julia, Python y R⁶. En este trabajo usamos la versión en Julia ya que es la más eficiente. Estos paquetes permiten agregar jugadores con *priors* y valores de γ y β personalizados. Al agregar una partida, debe indicarse los identificadores de los jugadores, el ganador y el día de la partida. Permite representar equipos, además de aplicarles multiplicadores a cada jugador.

Modelos de *handicap* y *komi*

Sea h la cantidad de piedras de *handicap* de una partida. Decidimos dos posibles alternativas para integrar el efecto del *handicap* y del *komi* en TTT, aprovechando la posibilidad de poner compañeros de equipos. La primera opción es mediante modelos lineales. Definimos el rendimiento final del jugador negro como $t_{negro} = p_{negro} + a \cdot h + b$, donde a es la variable aleatoria que representa la pendiente y b es la variable aleatoria que representa la ordenada al origen, en un modelo lineal del efecto del *handicap*. En términos de TTT, a y b son jugadores que participan en equipo junto al jugador negro, configurados con $\gamma = 0$ (el efecto del *handicap* no varía en el tiempo) y $\beta = 0$ (el efecto del *handicap* siempre es el mismo en cada partida), y h es el multiplicador aplicado a a . El efecto del *komi* puede ser agregado de igual manera a la habilidad del jugador blanco, utilizando solamente una pendiente ya que el efecto de la ordenada al origen es tenido en cuenta en el *handicap*.

⁶Disponible en <https://github.com/glandfried/trueskillThroughTime>

Otra opción considerada es la de usar una variable aleatoria nueva para cada *handicap* y cada *komi* posible. Es decir, tener diez jugadores $handicap_i$ para i entre 0 y 10, y uno por cada valor de *komi* existente en el conjunto de datos. Este modelo resulta mucho más complejo que el anterior, lo cual debería ser penalizado en la evidencia, pero permite captar efectos no lineales. Será seleccionado únicamente si el efecto no lineal es suficientemente importante para explicar el fenómeno.

En base a estas variantes, definimos cuatro modelos alternativos:

- TTT-h: modela *handicap* con una variable por cada cantidad de piedras posible
- TTT-h-k: igual a TTT-h, agregando una variable por cada valor posible de *komi*
- TTT-hreg: modela *handicap* lineal
- TTT-hreg-kreg: igual a TTT-hreg, agregando *komi* lineal

3.3. Cálculo de la evidencia

Al ser la métrica para comparar los modelos, resulta imprescindible evaluar la evidencia de los distintos modelos de la forma correcta y justa. Los tres métodos cuentan con una fórmula para calcular la verosimilitud de una partida. A partir de esta, podemos integrar para llegar a la evidencia.

$$P(d_i) = \int_{-\infty}^{\infty} P(d_i|s_1, s_2)P(s_1)P(s_2)ds_1ds_2 \quad (3.10)$$

Sin embargo, en el caso de WHR, al usar la función logística, no cuentan con fórmula cerrada para la evidencia. Es por eso que aproximamos la integral con una sumatoria. Asumiendo que $s_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ y $s_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$, *steps* y *kSigmas* parámetros de precisión, $cota = \lfloor \frac{steps}{2} \rfloor$ y $gap_x = \frac{2 \cdot kSigmas \cdot \sigma_x}{steps}$.

$$P(d_i) \approx \sum_{a=-cota}^{cota} \sum_{b=-cota}^{cota} P(d_i|s_1 = \mu_1 + a \cdot gap_1, s_2 = \mu_2 + b \cdot gap_2) \phi\left(\frac{a \cdot gap_1}{\sigma_1}\right) \phi\left(\frac{b \cdot gap_2}{\sigma_2}\right) \quad (3.11)$$

Otro factor a tener en cuenta son las partidas agregadas al sistema de inferencia antes de evaluar la evidencia, además de la convergencia para los casos de WHR y TTT. Para esto, decidimos usar la misma metodología que usa la AAGo para cargar sus datos: dividimos las partidas en tandas, o *batches*. Las tandas son agrupadas según el torneo al que pertenecen las partidas, y se utiliza la fecha de finalización del torneo como fecha de las partidas de esa tanda. Estas son procesados en orden según su fecha de finalización:

1. Calculamos la evidencia para cada partida individual del *batch*.
2. Agregamos las partidas al sistema de inferencia.
3. En el caso de TTT y WHR, hacemos converger el sistema.

De esta manera, la evidencia final es la productoria de las evidencias individuales de cada partida. Al ser números pequeños, para evitar los problemas numéricos conocidos, realizamos e informamos la suma de los logaritmos de las evidencias.

Capítulo 4

Metodología

En este capítulo nos enfocamos en estudiar las propiedades de dos elementos centrales: los datos, y los modelos. Primero hacemos un resumen de la base de datos de la AAGo, incluyendo los preprocesamientos requeridos y las propiedades básicas de los datos utilizados para estimar. Luego, ilustramos distintas propiedades prácticas del funcionamiento de los tres modelos utilizados, que aportan a la comprensión de su uso y sus resultados.

4.1. Fuente de datos

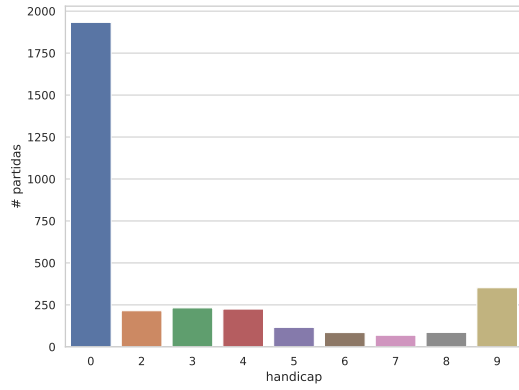
Trabajamos con el conjunto de datos provisto por la AAGo, del cual nos quedamos con las partidas válidas. Cuenta con 3313 partidas entre el 27 de junio del 2016 y el 26 de abril del 2020, entre 188 jugadores. Las partidas se dividen en 70 torneos o eventos, entre los cuales hay una gran diversidad: el más grande cuenta con 187 partidas, mientras que hay algunos de menos de cinco. El jugador blanco, que suele ser el de mayor habilidad estimada, gana en el 55,5 % de las partidas. La distribución de *handicap* utilizado en partidas puede observarse en la Figura 4.1(a). El 58,3 % de las partidas no usan *handicap*. Por el lado del komi, solo hay dos valores utilizados: 0,5 (en 1644 partidas) y 6,5 (en 1669 partidas).

Una de las primeras tareas constó en filtrar las partidas que son consideradas válidas por la asociación¹. Esto incluyó eliminar las partidas no rankeadas, las que no fueron ganadas por ningún jugador y las que terminaron por *walkover* (por ejemplo, por ausencia o descalificación del oponente). Además, un dato vital para el funcionamiento de los modelos es la fecha de las partidas. Un torneo puede durar varios días e incluso semanas. Por eso, cada partida tiene disponible tres fechas: la de inicio del torneo, la de la partida particular, y la de finalización de torneo. En la sección 5.1.1 explicaremos cómo se utilizan estos datos, y en la sección 5.3 veremos su impacto en las predicciones.

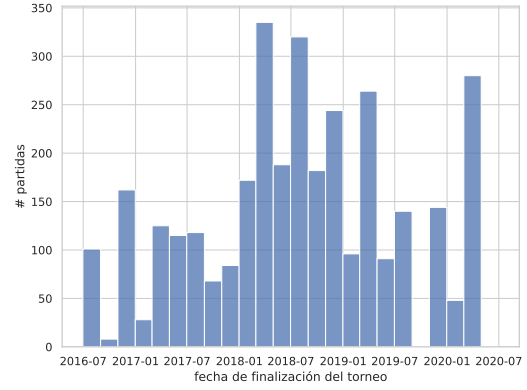
En la Figura 4.1(c) podemos ver la media de las últimas estimaciones de cada jugador de la AAGo. Estos valores pueden encontrarse en la base de datos provista por la asociación, y fueron utilizados para verificar el correcto uso del sistema (ver 4.1.1). Los jugadores con habilidad positiva son dan, el resto *kyu*. Se considera que la categoría de un jugador es el piso de la media de habilidad. Por eso no existen jugadores con habilidad entre -1 y 1 : la distancia entre estos valores es nula para el sistema. Esto debe ser tenido en cuenta cada vez que se calcula una distancia al utilizar los valores de AGA/AAGo.

La base de datos está almacenada en formato SQL. Para nuestros experimentos, pasamos a formato CSV las tablas que contienen la información respectiva a las partidas, a las estimaciones históricas calculadas por la AAGo, a la fecha de cada evento, y a las categorías auto-declaradas por los jugadores en cada torneo.

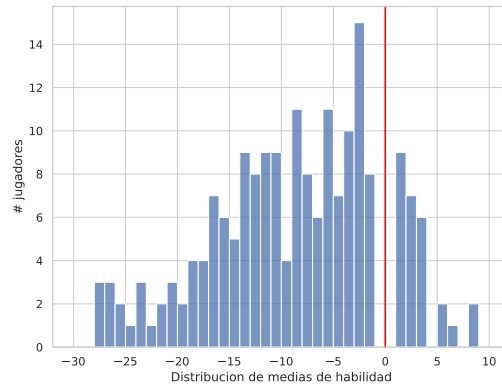
¹Basados en los scripts utilizados por ellos: <https://github.com/elsantodel90/RAAGo>



(a) Distribución del *handicap*



(b) Distribución de partidas en el tiempo.



(c) Distribución de estimaciones.

Figura 4.1: Descripción de la base de datos de la AAGo, con datos hasta abril del 2020. En (a) se exhibe la cantidad de partidas con cada valor de *handicap*. El 58,3% de las partidas no usan *handicap*. En (b) cada barra representa la cantidad de partidas en un período de dos meses calendario. En (c) se presenta el ranking de jugadores de la AAGo, calculado por la Asociación.

4.1.1. Réplica de estimaciones de AAGo

Para que los resultados que obtengamos sean válidos, es necesario verificar que estemos usando de forma correcta el código de la AAGo. Con este experimento corroboramos que los resultados obtenidos posteriormente emergen del funcionamiento del modelo. En términos generales nos proponemos correr el modelo sobre todas las partidas de la base de datos, y esperamos obtener las mismas estimaciones que las registradas por la asociación. Como contamos en la subsección 3.2.2, esto debe hacerse iterando los torneos y procesando de a uno por ejecución. Los torneos se ordenan con el mismo criterio que utiliza la AAGo: primero por fecha de finalización y, en caso de empates, se desempata primero por la fecha de inicio del torneo, y luego por el id del torneo.

Se realizó este procedimiento con todos los torneos válidos, y comparamos nuestras estimaciones con las de la asociación. Estas mostraron una diferencia (por errores de redondeo) de a lo sumo $1,4 \times 10^{-3}$ y $4,6 \times 10^{-4}$ en las medias y los desvíos estándares de las estimaciones, respectivamente. Como esta diferencia es insignificante, podemos concluir que estamos utilizando el sistema de la misma manera que la AAGo.

4.2. Exploración del funcionamiento de los algoritmos de estimación

Para hacer una comparación justa y equilibrada, realizamos distintos experimentos para explorar y validar el funcionamiento de los distintos modelos. De esta manera, esperamos ayudar a comprender las propiedades prácticas de cada uno y de esta forma dar nociones de los posibles resultados al trabajar con datos reales. Cada caso sintético está pensado para evaluar alguna propiedad de interés para la problemática de estimación de habilidad. Todos los casos dejan de lado el estudio del *handicap* y el *komi*: simulan partidas en las que ambas variables valen 0 (excepto el de evolución de la habilidad en AAGo, como se explicará en la Sección 4.2.4).

4.2.1. Pequeña comunidad

El primer experimento busca ver que el orden de las partidas en un mismo día no afecte el resultado de las estimaciones. Si bien es algo simple a primera vista, métodos iterativos como Elo y Glicko no cumplen con esta condición. El conjunto de datos queda compuesto por tres partidas con tres jugadores distintos: en la primera partida *jugador*₁ le gana a *jugador*₂, en la segunda *jugador*₂ le gana a *jugador*₃ y en la última *jugador*₃ le gana a *jugador*₁. Esperamos que los tres jugadores terminen con la misma habilidad, ya que no hay razones para que un jugador se destaque por sobre el resto.

AAGo

Nos propusimos dos formas diferentes de realizar este experimento: Una incluyendo a todos los partidos en un solo torneo, la otra separándolos en tres torneos distintos (en el mismo día). En ambos casos, se ingresa como categoría la de primer dan, por lo que esta es utilizada si no se tiene datos de un jugador. Como el modelo que usa la AAGo no incorpora la información de la historia completa, los resultados no son iguales.

En la Tabla 4.1 se muestran las estimaciones resultantes del experimento con todas las partidas en un solo torneo. Este resultado es el esperado, se estima la misma habilidad para todos los jugadores.

Las estimaciones resultantes del experimento separado en tres torneos se encuentran en la Tabla 4.2. Se puede observar claramente que no estima la misma habilidad para los tres jugadores. Esta divergencia se debe a que el modelo no recuerda los resultados de las partidas anteriores sino por medio de las estimaciones resultantes. Además, tampoco utiliza los datos de partidas posteriores como dato de entrada para calcular las estimaciones de partidas anteriores. Estos resultados nos muestran la principal limitación del modelo: no integrar la información de toda la historia de partidas. Además, al hacer un uso menos eficiente de los datos, las estimaciones resultantes cuentan con mayor incertidumbre.

Jugador	Media de la habilidad	Desvío estándar de la habilidad
<i>jugador</i> ₁	1,5	0,715881
<i>jugador</i> ₂	1,5	0,715881
<i>jugador</i> ₃	1,5	0,715881

Tabla 4.1: Resultados del experimento Pequeña Comunidad con AAGo, en un sólo torneo

Jugador	Media de la habilidad	Desvío estándar de la habilidad
<i>jugador</i> ₁	2,0906	0,967062
<i>jugador</i> ₂	1,46187	0,807324
<i>jugador</i> ₃	-1,34088	0,932484

Tabla 4.2: Resultados del experimento Pequeña Comunidad con AAGo, separando en tres torneos

WHR y TTT

Los resultados con WHR pueden verse en la Tabla 4.3. Efectivamente, la habilidad de los tres jugadores es 0, por lo que se cumple lo que buscamos. Lo mismo sucede con TTT (Tabla 4.4), lo cual comprueba

una de las principales ventajas de los modelos de historial completo: el orden de las partidas dentro de un mismo día no debe modificar las estimaciones.

Jugador	Media de la habilidad	Desvío estándar de la habilidad
$jugador_1$	0,0	173,63
$jugador_2$	0,0	173,63
$jugador_3$	0,0	173,63

Tabla 4.3: Resultados experimento Pequeña Comunidad con WHR

Jugador	Media de la habilidad	Desvío estándar de la habilidad
$jugador_1$	0,0	2,395
$jugador_2$	0,0	2,395
$jugador_3$	0,0	2,395

Tabla 4.4: Resultados experimento Pequeña Comunidad con TTT

4.2.2. Intercomunidades

El siguiente experimento consta de dos comunidades de tres jugadores cada una. En cada comunidad, los jugadores tienen la misma habilidad, pero hay una comunidad que es mejor que la otra. El conjunto de datos cuenta con dos subconjuntos de partidas: algunas intracomunidad, donde miembros de una misma comunidad juegan entre sí, y otras intercomunidades, donde juega un representante de cada comunidad. Las partidas intracomunitarias tienen el mismo formato que en el experimento de Pequeña Comunidad (ver 4.2.1), repetidas varias veces.

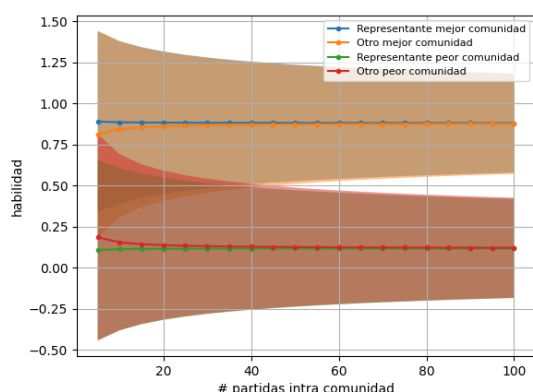
Buscamos, a partir de los datos de partidas, reconstruir las habilidades reales. Veremos los cambios en las estimaciones con respecto a la cantidad de partidas intracomunitarias y extracomunitarias. Todas las partidas se registran en el mismo día, ya que no nos interesa estudiar el efecto del tiempo en este experimento. En el caso de AAGo, las partidas se corren en un solo *batch*, para que el orden de las partidas no afecte los resultados.

AAGo

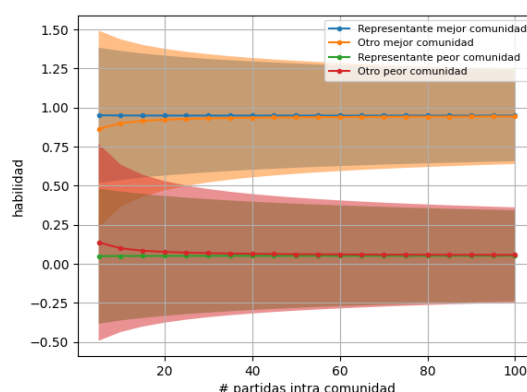
En la Figura 4.2 se pueden observar los resultados de este experimento usando el modelo AAGo, y distintas cantidades de partidas intra e intercomunitarias. Si bien cada comunidad tiene tres jugadores, se ven dos curvas por cada una porque los jugadores que no son el representante de cada comunidad tienen curvas indistinguibles. En primer lugar, observamos que el orden de las habilidades obtenido sigue lo esperado. Se estima más habilidad para todos los jugadores de la mejor comunidad que para todos los de la peor, con cualquier cantidad de partidas. Además, dentro de la mejor comunidad el jugador con más habilidad es el representante, quien jugó las partidas contra el representante de la otra comunidad. Análogamente, el jugador de la peor comunidad con la estimación más baja es su representante. Si bien lo deseable sería que todos los jugadores de cada comunidad alcancen la misma estimación, es esperable que los representantes se destaquen (para arriba o para abajo), dado que son quienes efectivamente se enfrentaron al otro grupo.

En esta figura se puede observar una tendencia en la relación entre la cantidad (y tipo) de partidas y las distancias entre habilidades que también era esperable. Cuantas más partidas intracomunitarias se hayan jugado, más cercanas están las estimaciones dentro de cada comunidad. Cuantas más partidas intercomunitarias se jueguen, más distantes están las comunidades entre sí.

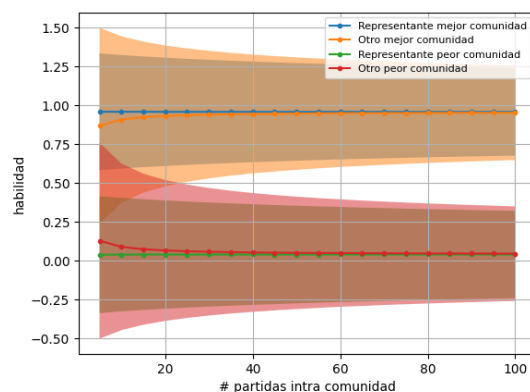
Sobre la diferencia de habilidades, esperábamos que dentro de cada comunidad los jugadores tengan estimaciones iguales o muy similares. Basta mirar el gráfico para corroborar que esto efectivamente se dio, sobre todo a partir de las 20 partidas intracomunitarias. Para hablar de la diferencia de habilidad entre las comunidades, expresémoslo en términos de probabilidad de ganar. Como en todos los casos la cantidad de partidas que le gana el representante de una comunidad al de la otra es el 80 % del total de partidas, esperábamos que las estimaciones de estos dos jugadores impliquen una probabilidad de ganar



(a) La mejor comunidad gana 8 de 10 partidas intercomunitarias



(b) La mejor comunidad gana 40 de 50 partidas intercomunitarias



(c) La mejor comunidad gana 80 de 100 partidas intercomunitarias

Figura 4.2: Habilidad de jugadores en distintas comunidades, según la cantidad de partidas intracomunitarias e intercomunitarias, utilizando AAGo. La escala utilizada corresponde a la que usa el modelo (*kyu/dan*). Como el modelo no asigna habilidades entre 1 y -1, se le restó 1 unidad a las habilidades positivas y sumó 1 a las negativas, para cerrar la brecha entre las categorías y que el gráfico sea más claro.

del primero de 0,8. El valor de esta probabilidad con 100 partidas intracomunitarias se puede observar en la tabla 4.5:

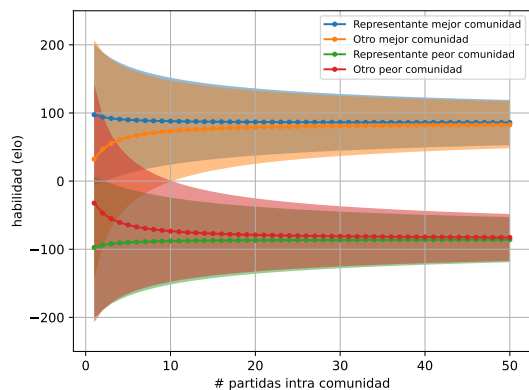
Partidas intercomunitarias	Distancia considerando incertidumbre	Distancia sin considerar incertidumbre
10	0.7486	0.7502
50	0.7871	0.7886
100	0.7929	0.7942

Tabla 4.5: Distancias entre estimaciones de representantes con 100 partidas intracomunitarias, utilizando AAGo

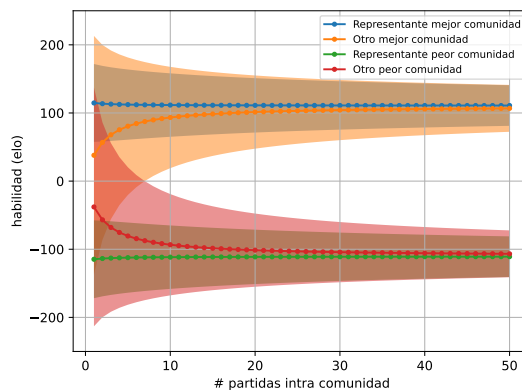
Estos resultados siguen lo esperado. Además, se puede observar que la probabilidad se acerca a 0,8 a medida que aumenta la cantidad de partidas intercomunitarias. Esta tendencia es lógica, si consideramos que estas partidas son las que expresan la diferencia de habilidad entre comunidades, por lo que tener mayor número de ellas permite al modelo acercarse más a la diferencia real.

WHR

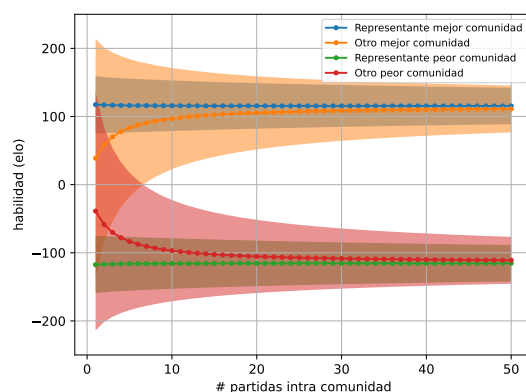
Pasamos ahora a WHR. En la Figura 4.3 podemos observar que el método diferencia la habilidad del representante de cada comunidad del resto de jugadores.



(a) La mejor comunidad gana 8 de 10 partidas intercomunitarias



(b) La mejor comunidad gana 40 de 50 partidas intercomunitarias



(c) La mejor comunidad gana 80 de 100 partidas intercomunitarias

Figura 4.3: Habilidad de jugadores en distintas comunidades, según la cantidad de partidas intracomunitarias e intercomunitarias, utilizando WHR

Esto se atenúa con una mayor cantidad de partidas intracomunitarias, pero se necesita una cantidad apreciable de partidas para que esto suceda. En cuanto a la diferencia entre comunidades, si bien las medias de las estimaciones se alejan, siguen teniendo poca distancia con respecto a los desvíos estándar. Tanto las partidas intra como intercomunitarias bajan la incertidumbre en las estimaciones. Sin embargo, es destacable que las partidas intercomunitarias solo disminuyen la incertidumbre de los representantes de cada grupo, que son los que juegan las partidas, sin afectar a la incertidumbre en las estimaciones del resto de jugadores.

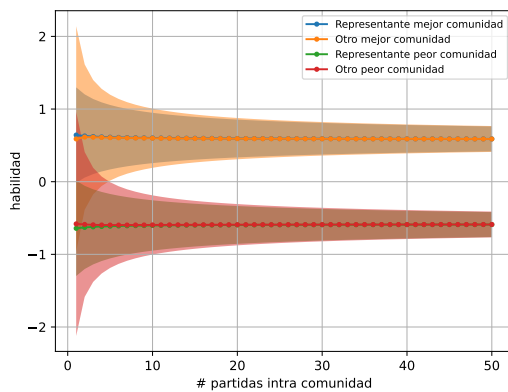
Nuevamente, esperamos que la diferencia entre jugadores de dos comunidades sea de alrededor de 0,8 en probabilidad de ganar. En la Tabla 4.6 comparamos las estimaciones de los representantes de cada grupo, con 100 partidas intracomunitarias. Podemos ver que aumentando la cantidad de partidas intercomunitarias, nos acercamos cada vez más al objetivo. Es notorio que con 10 partidas, esta distancia es particularmente grande.

Partidas intercomunitarias	Distancia considerando incertidumbre	Distancia sin considerar incertidumbre
10	0,72580	0,7290
50	0,7791	0,7819
100	0,7883	0,7906

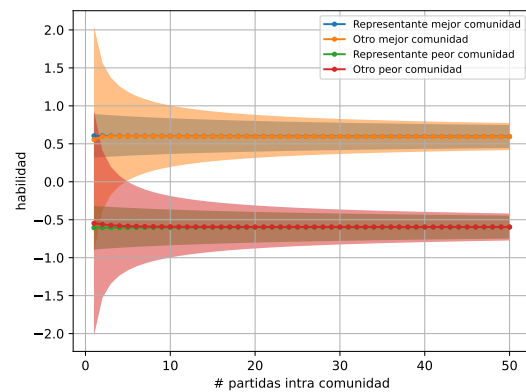
Tabla 4.6: Distancias entre estimaciones de representantes con 100 partidas intracomunitarias, utilizando WHR

TTT

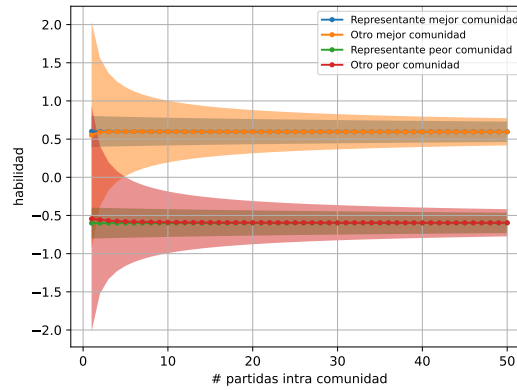
En la Figura 4.4 podemos observar los resultados del mismo experimento con TTT. Si bien comparten mucho con lo visto en los otros modelos, se puede destacar la poca distancia entre las estimaciones de habilidad de un jugador representante con los otros jugadores de su comunidad. Además, vemos que la incertidumbre en las estimaciones baja mucho más rápido al aumentar la cantidad de partidas.



(a) La mejor comunidad gana 8 de 10 partidas intercomunitarias



(b) La mejor comunidad gana 40 de 50 partidas intercomunitarias



(c) La mejor comunidad gana 80 de 100 partidas intercomunitarias

Figura 4.4: Habilidad de jugadores en distintas comunidades, según la cantidad de partidas intracomunitarias e intercomunitarias, utilizando TTT

En cuanto a distancia entre jugadores de distintas comunidades, en la Tabla 4.7 podemos ver los resultados. TrueSkill Through Time es el modelo que mejor capta esta distancia. Además, consigue resultados excelentes con tan solo 10 partidas intercomunitarias.

Partidas intercomunitarias	Distancia considerando incertidumbre	Distancia sin considerar incertidumbre
10	0,7943	0,7977
50	0,7978	0,8005
100	0,7985	0,8006

Tabla 4.7: Distancias entre estimaciones de representantes con 100 partidas intracomunitarias, utilizando TTT

4.2.3. Gran comunidad

Cuando uno no tiene información previa sobre un jugador, debe iniciar su creencia con una **estimación a priori**. En el caso de la AAGo, como cada jugador declara su categoría autopercibida al momento de anotarse a un torneo, su sistema utiliza esta categoría para seleccionar una estimación acorde, aunque el sistema también permite que uno decida otro *prior* manualmente. En TTT, por defecto se utiliza un *prior* con media 0 y un desvío estándar suficientemente grande, y también da la opción para seleccionar otra estimación manualmente. A diferencia de los otros paquetes, por como está programado WHR, no permite designar *priors* distintos al del sistema. Esto complica algunos casos de uso. En este experimento sintético buscamos ver si se puede conseguir un efecto similar entrenando el modelo con datos.

Utilizamos una comunidad de varios jugadores, cuyas habilidades se encuentran equiespaciadas en un rango. La comunidad cuenta con 21 jugadores, separados entre sí por 40 puntos de habilidad en escala de Elo. A partir de esta comunidad, simulamos una partida entre cada par de jugadores repetidas veces. Es decir, por cada partida, generamos un resultado aleatorio a partir de la probabilidad de ganar de cada jugador, utilizando la Ecuación (3.8) de verosimilitud de WHR. Cada jugador tiene partidas contra los 10 siguientes de mayor habilidad, y esto es repetido una cantidad de veces. Nos interesa ver si las habilidades estimadas se acercan a las reales, y si la distancia en habilidad entre jugadores es respetada. Al ser experimentos aleatorios, repetimos cada uno 20 veces para poder tomar estadísticas.

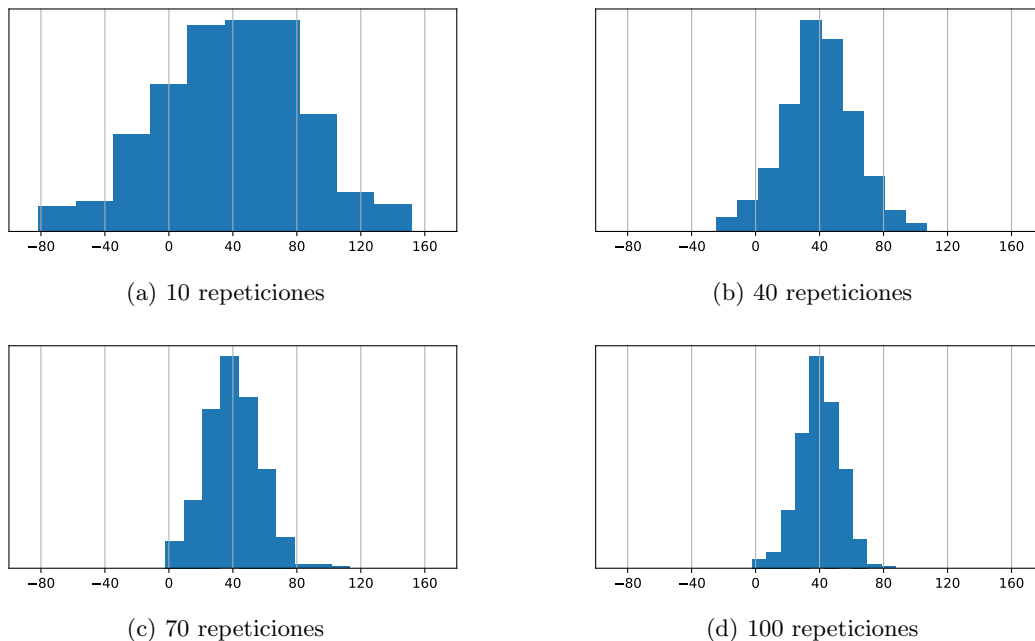


Figura 4.5: Histogramas de distancias relativas estimada entre un jugador y el siguiente de mayor habilidad, en función de la cantidad de repeticiones de las partidas. La comunidad cuenta con 21 jugadores, separados entre sí por 40 puntos de habilidad en escala de Elo.

En la Figura 4.5 podemos ver histogramas de distancias relativas entre un jugador y el siguiente de mayor habilidad, en función de la cantidad de repeticiones de las partidas. Podemos ver que a medida

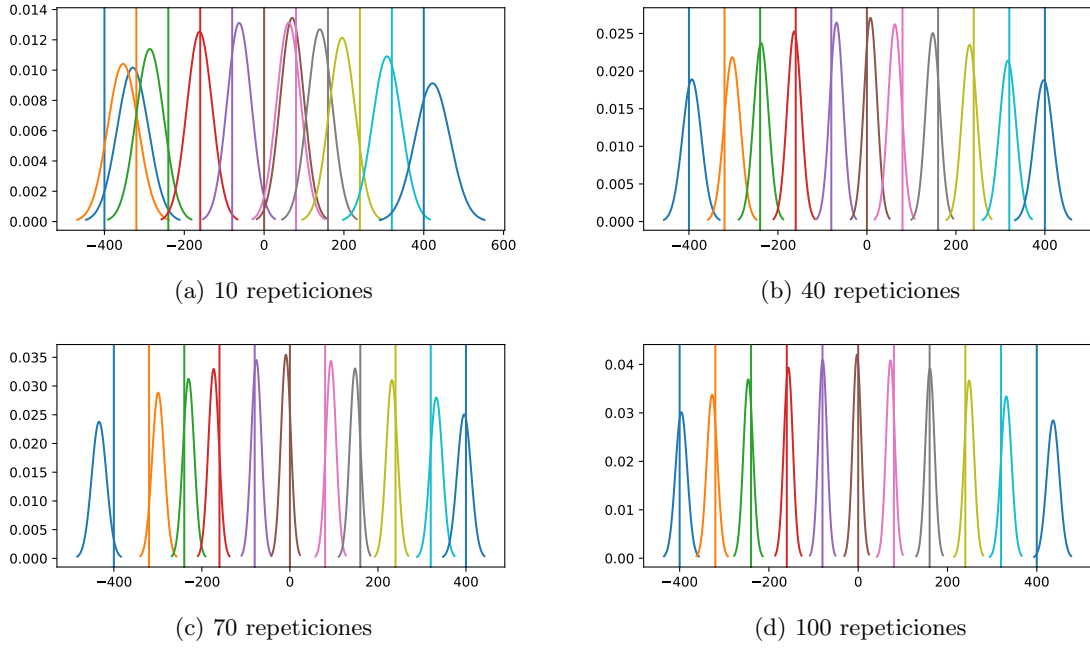


Figura 4.6: Ejemplos de resultados de simulaciones con distinta cantidad de repeticiones de partidos. Cada línea vertical representa la habilidad real de un jugador, y las campanas gaussianas son las estimaciones obtenidas.

que aumenta la cantidad de repeticiones, se forman campanas cada vez menos anchas. A partir de las 40 repeticiones conseguimos que la mayoría de distancias se encuentren a menos de 40 puntos de distancia.

En la Figura 4.6 podemos ver ejemplos de resultados de simulaciones con distinta cantidad de repeticiones de partidos. Cada línea vertical representa la habilidad real de un jugador, y las campanas gaussianas son las estimaciones obtenidas. Podemos ver que rápidamente se consiguen estimaciones cercanas a la habilidad real. Sin embargo, cuesta más estimar la habilidad de los jugadores de los extremos.

4.2.4. Evolución de la habilidad

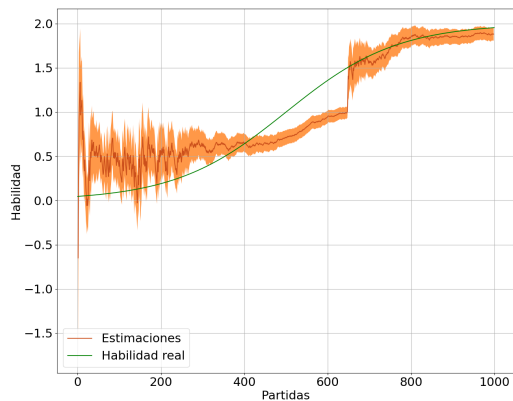
El último experimento que realizamos busca ver si los modelos logran captar el cambio de la habilidad de un jugador en el tiempo. Asumimos que la habilidad real en el tiempo cambia siguiendo una curva logística. Simulamos 1000 partidas, una por día, entre nuestro jugador especial y otro de habilidad cercana. En el caso de AAGo, probamos con otras opciones de distancia entre partidas, ya que el sistema no permite cambiar el valor del factor dinámico. Los otros jugadores son seleccionados de una gran comunidad a la que se le conoce con poca incertidumbre la habilidad verdadera. Tanto con AAGo como con TTT, esta comunidad está conformada por jugadores cuyas estimaciones de habilidad a priori se encuentra determinada por una grilla equidistante. En el caso de WHR las estimaciones a priori se generan como fue explicado en la sección anterior.

AAGo

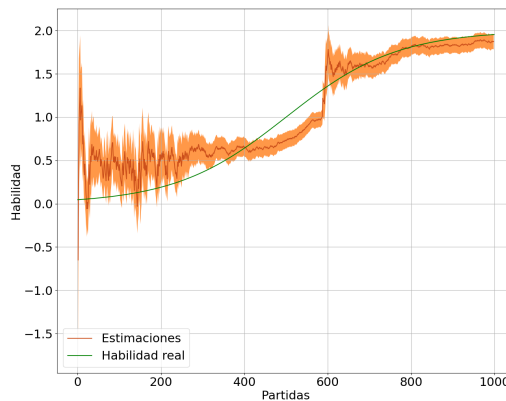
Para este experimento utilizamos un *komi* de 6 para generar partidas balanceadas. Esta necesidad viene de que un *komi* de 0 genera una partida desbalanceada, quien juega primero tiene una ventaja, que es tenida en cuenta por el modelo. Es decir, la probabilidad de ganar de un jugador contra otro de idéntica estimación, en un juego con $komi = 0$, no es 0,5, sino que es mayor para quien juega con las negras (por la ventaja de jugar primero). Esto se puede ver tanto en las fórmulas del modelo (como dijimos en la sección 3.2.2, cuando $komi = 0$, $d_j > 0$), como empíricamente (la probabilidad de ganar recién descripta equivale a un 0,63 a favor de las negras), pero esta ventaja alcanza su mínimo con 6 de *komi* (en ese caso, la probabilidad baja a 0,51). Así, creando partidas con un *komi* de 6 compensamos la ventaja que

el modelo agrega por orden de juego.

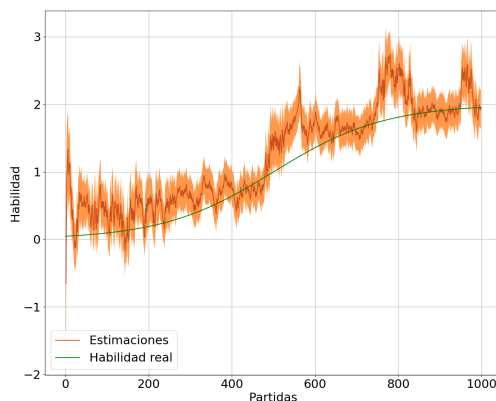
Aquí experimentamos con distintas sensibilidades al paso del tiempo. Dado que este modelo no permite ingresar como parámetro el factor dinámico, usamos distintas cantidades de días separando partidas. Recordar que este sistema agrega incertidumbre según el producto entre el factor dinámico al cuadrado y la cantidad de días desde la última estimación. Los resultados se exhiben en la Figura 4.7.



(a) 1 día de separación entre partidas



(b) 10 días de separación entre partidas



(c) 100 días de separación entre partidas

Figura 4.7: Evolución de la habilidad y de las estimaciones de un jugador utilizando el modelo de AAGo, a lo largo de 1000 partidas, y con distinta cantidad de días de separación. El área sombreada muestra el margen entre la media de la estimación y un desvío estándar hacia arriba y hacia abajo.

Se puede observar que a medida que aumenta la cantidad de días de separación la curva de estimaciones sigue más la forma de la habilidad real, pero también es más inestable y ruidosa. En los tres experimentos se nota al principio un gran movimiento, con alta incertidumbre. Esta incertidumbre, y el exagerado movimiento que la acompaña, se pueden explicar por la poca cantidad de partidas con las que se cuenta al momento de hacer estas estimaciones. Como no es un modelo de historial completo, sólo se tienen en cuenta las anteriores.

Observando los resultados en detalle, podemos ver que las estimaciones correspondientes al experimento con 100 días de separación entre partidas exhiben un alto grado de incertidumbre, expresado tanto en el área sombreada alrededor de la curva como en su carácter cambiante. En cuanto a los experimentos con 1 y 10 días, se repite una estructura interesante. Luego del período de variación inicial, las estimaciones se estabilizan en un valor cercano al de la habilidad real. Sin embargo, posteriormente se anclan en ese valor, distanciándose de la curva real y manteniendo una baja incertidumbre. Eventualmente, la diferencia entre

las estimaciones y la habilidad real confronta al modelo con un resultado muy diferente del esperado, provocando un salto en la estimación para ajustarse al valor real y aumentando la incertidumbre. Este salto se da alrededor de las 650 partidas, en el experimento con 1 día; y de las 580 partidas, en el de 10 días. Como en el segundo hay mayor incertidumbre, es esperable que sea más sensible a los cambios y por eso que suceda antes. Hacia el final las estimaciones recuperan estabilidad, a la vez que el ritmo de cambio de la habilidad real se decrementa. Estos resultados nos sugieren una limitación en la capacidad del modelo de la AAGo de adaptarse a los cambios en la habilidad.

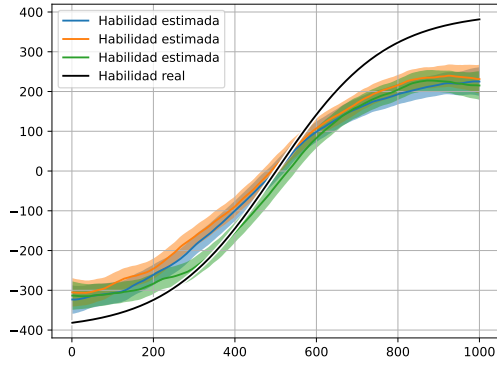
WHR y TTT

A continuación mostramos los resultados utilizando modelos de historial completo. En la Figura 4.8 mostramos los resultados del experimento sobre WHR, con distintos valores de factor dinámico, o γ . Se pre-entrenó una gran comunidad de oponentes utilizando 60 repeticiones, como fue explicado en la sección anterior.

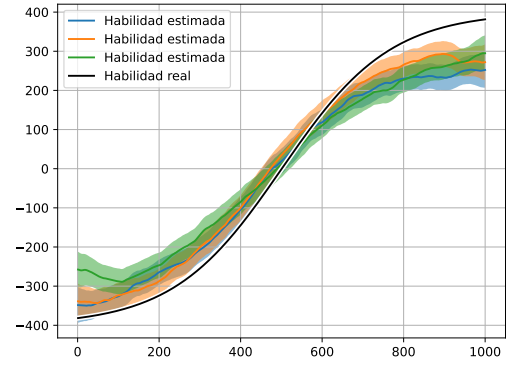
Podemos ver que con valores bajos, las curvas no llegan a captar todo el cambio de habilidad, sobre todo al principio y al final del experimento. Incluso con valores más altos, las estimaciones siempre van por abajo de la habilidad real en las últimas partidas de los experimentos. También con valores altos de factor dinámico, vemos que el sistema da curvas menos suaves. De todos modos, los resultados son mucho mejores que con AGA/AAGo: las estimaciones son mucho más cercanas a la curva real y no hay grandes saltos.

En la Figura 4.9 mostramos los resultados del experimento sobre TTT. Al igual que en WHR, con valores bajos de γ la habilidad no llega a cambiar al ritmo que se espera. Una diferencia se puede ver con valores altos de este hiperparámetro: las curvas de aprendizaje siguen siendo suaves, aunque distintas ejecuciones dieron curvas muy distintas entre sí. Esto se debe a una mayor sensibilidad: para la habilidad de un día, se cree más en el nuevo dato y menos en las estimaciones del pasado y futuro.

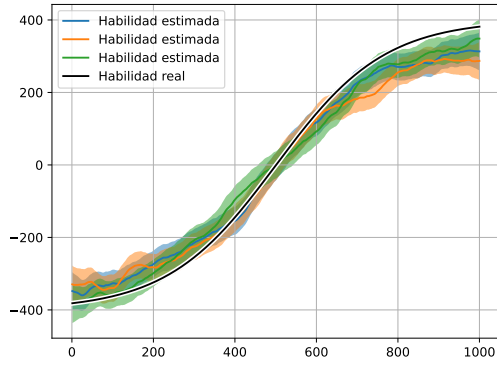
Como era esperable, los modelos de historial completo dan curvas más suaves y ajustan mejor a la habilidad real. También podemos ver la importancia de elegir correctamente el factor dinámico. Mientras un valor muy bajo puede ser insuficiente para captar los cambios, un valor demasiado alto resulta en curvas abruptas o un exceso de incertidumbre en los resultados.



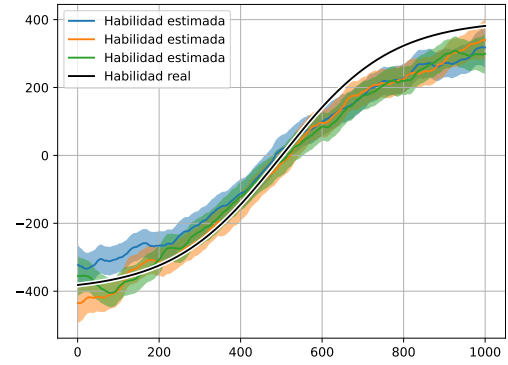
(a) $\gamma^2 = 14$



(b) $\gamma^2 = 36$

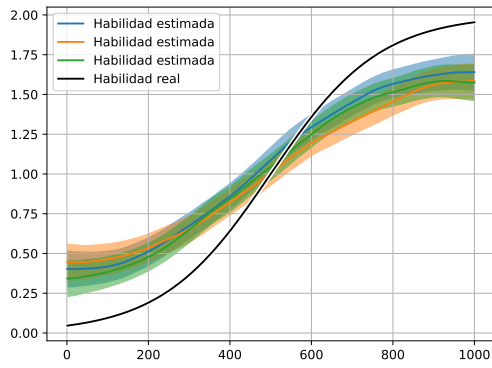


(c) $\gamma^2 = 64$

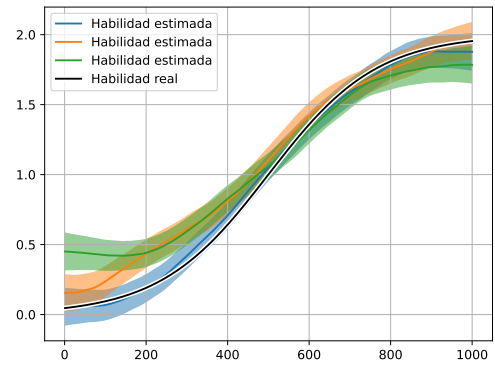


(d) $\gamma^2 = 100$

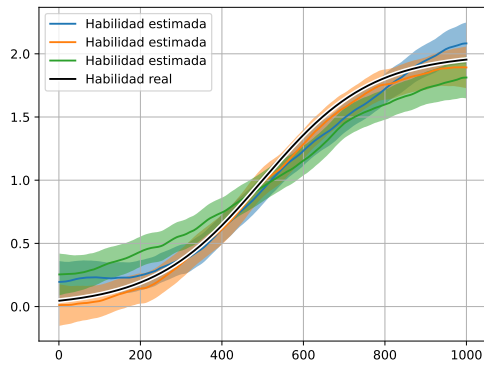
Figura 4.8: Resultados del experimento de Evolución de la habilidad con WHR, con un día de separación entre partidas. Por cada valor de factor dinámico, se muestran 3 curvas de aprendizaje estimadas. Cada curva de aprendizaje muestra la media de las estimaciones, sombreando un desvío estándar. Se pre-entrenó una gran comunidad de oponentes utilizando 60 repeticiones. El valor de $\gamma^2 = 14$ es tomado de la publicación original de WHR, los otros son cuadrados perfectos.



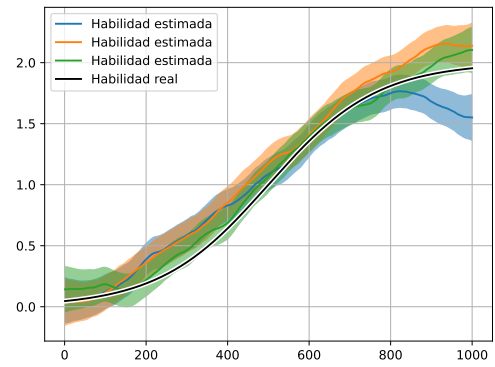
(a) $\gamma = 0,0075$



(b) $\gamma = 0,01$



(c) $\gamma = 0,015$



(d) $\gamma = 0,02$

Figura 4.9: Resultados del experimento de Evolución de la habilidad con TTT, con un día de separación entre partidas. Por cada valor de factor dinámico, se muestran 3 curvas de aprendizaje estimadas. Cada curva de aprendizaje muestra la media de las estimaciones, sombreando un desvío estándar.

Capítulo 5

Resultados

En este capítulo comparamos cuantitativamente el desempeño de los distintos modelos y analizamos sus propiedades cualitativas a través de la visualización de curvas de aprendizaje estimadas de diferentes miembros de la Asociación Argentina de Go (AAGo). En términos generales, encontramos una mejora indudable de los modelos con historia completa respecto del estimador actual de la AAGo, tanto en términos cuantitativos como cualitativos. Por el contrario, la comparación entre los modelos con historia completa no puede considerarse definitiva. Al cierre de esta tesis registramos una ventaja importante a favor del modelo WHR por sobre TTT en términos de probabilidad de los modelos dado el conjunto de datos de la AAGo, $P(\text{Modelo} \mid \text{Datos} = \text{AAGo})$. Sin embargo la diferencia no es suficiente para descartar ningún modelo, ya que existe la posibilidad de que se inviertan los resultados si se cambia algún elemento del análisis (la base de datos, los parámetros con los que fueron analizados los modelos, u otras decisiones de implementación involucradas). De hecho, hemos visto cómo ciertas modificaciones de criterio que hemos introducido durante el período de análisis tuvieron un impacto de varios ordenes de magnitud en las predicciones a priori, $P(\text{Datos} \mid \text{Modelo})$. Sin haber obtenido una diferencia indudable con los elementos que integran el condicional que utilizamos para evaluar, menos aun podemos afirmar que bajo otras condiciones la diferencia relativa que hemos observado se mantenga o se profundice. Por este motivo, nuestra recomendación para la Asociación Argentina de Go es elegir el modelo que mejor les convenga en términos prácticos y nuestra recomendación para posibles extensiones de esta tesis es volver a evaluar el desempeño de ambos modelos, WHR y TTT.

Este capítulo está dividido en cuatro secciones. En la primera, explicamos las condiciones sobre las cuales evaluamos los modelos y damos los principales resultados cuantitativos. Luego, pasamos a explicar más detalladamente el rendimiento de cada uno, incluyendo resultados cualitativos. En la segunda sección, mostramos lo obtenido con el modelo de AGA/AAGo. Vemos los problemas encontrados y las hipótesis puestas a prueba. En la tercera sección pasamos a mostrar las alternativas basadas en TrueSkill Through Time y su rendimiento. Por último, en la cuarta sección hacemos una comparación entre TTT y WHR.

5.1. Comparación cuantitativa de los modelos

Como hemos visto en 2.2, el desempeño de los modelos debe medirse en base a la predicción a priori de los datos, $P(\text{Datos} \mid \text{Modelos})$ (la evidencia) y el prior de los modelos, $P(\text{Modelos})$. En esta tesis hemos decidido otorgar a cada modelo la misma probabilidad a priori, por lo que la probabilidad de los modelos queda determinada exclusivamente por la evidencia. Queremos entonces el modelo que tenga menor sorpresa conjunta. La sorpresa de los modelos depende de las condiciones de evaluación. En las siguientes subsecciones detallamos las condiciones de evaluación y los resultados finales.

5.1.1. Condiciones de la evaluación

El estimador de la asociación (AGA) fue corrido replicando en detalle las condiciones con las que fue utilizado por la AAGo. En primer lugar, se agruparon todas las partidas de un mismo torneo. Los torneos

se ordenan por fecha de finalización. En caso de empates, se desempata primero por la fecha de inicio del torneo, y luego por el id del torneo. Cada torneo se considera una tanda o *batch*, y son procesados en orden. El sistema de AAGo recibe, además de la información de las partidas, la categoría declarada por los jugadores y su estimación a priori (en caso de existir). La estimación a priori de un jugador es la estimación a posteriori del último partido en el que participó. Además, se le informa cuantos días pasaron desde este último torneo, para que se pueda agregar incertidumbre de acuerdo con el modelo temporal.

Además de las fechas de inicio y fin del torneo, contamos con la fecha de cada partida. Tanto a WHR como a TTT los evaluamos utilizando la misma generación y orden de *batches*. Además, agregamos resultados obtenidos al evaluarlos utilizando la fecha de las partidas. En este caso, cada tanda es un día, y queda conformado por las partidas que se jugaron dicho día.

En caso de no existir estimaciones previas para un jugador, el modelo de la AAGo utiliza la categoría para determinar el *prior*. La media de la estimación a priori queda dada por el número de la categoría, mientras que el desvío estándar viene *hardcodeado* de una tabla, con valores optimizados por AGA. Como nombramos en la sección 3.2.3, WHR no modela ni permite agregar una estimación a priori para jugadores nuevos. Luego de inspeccionar el sistema, encontramos que los cálculos dan un *prior* con media 0,0 y desvío estándar 245,4 (en escala de Elo). Por último, TTT le pide al usuario del sistema asignar un *prior* a cada jugador. Decidimos utilizar valores como los de WHR, para hacer justa la comparación. Luego del cambio de escala, el prior queda dado por una media de 0,0 y 1,23 de desvío estándar.

Los modelos que propagan la información histórica por todo el sistema se ven obligados a realizar una serie de iteraciones hasta converger. El criterio de corte que elegimos para WHR fue de iterar hasta que se consiga una precisión de 10^{-3} , mientras que el criterio para TTT fue de cuatro iteraciones entre *batches*. Si bien puede parecer poco, probando con diez iteraciones se consiguieron resultados prácticamente idénticos.

5.1.2. Resultados principales

En la Figura 5.1 reportamos las predicciones a priori de los diferentes modelos. En el caso de TTT, los valores informados son los de TTT-hreg optimizado, como explicaremos en más detalle en la sección 5.4. Como la predicciones a priori conjunta tiene valores cercanos a 0, se reportan dos valores que permiten su interpretación: la media geométrica, y su valor absoluto en escala logarítmica. Se puede ver que el modelo

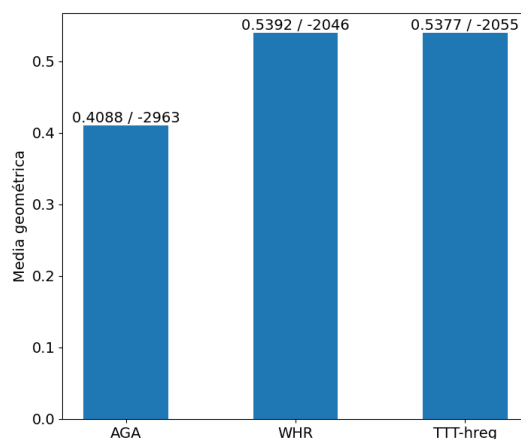


Figura 5.1: La media geométrica y evidencia en escala logarítmica de los modelos.

de AGA quedó muy por debajo de los modelos de historia completa, mientras que estos se parecen entre sí con una ventaja a favor de WHR por sobre TTT.

En primer lugar es importante notar que la promedio geométrico del modelo AGA se encuentra por debajo de 0,5. Esto puede resultar sorprendente debido a que un modelo nulo, que predice siempre resultados con probabilidad 0,5, tendría una media geométrica mayor al modelo AGA. Esto no debe interpretarse de manera lineal. En efecto, debido a que la evidencia conjunta es una producto de predicciones a priori (ver sección 2.2), existe la posibilidad de que todas las predicciones sea perfectas, salvo una. En definitiva con que un único elemento de esa productoria sea igual a cero alcanza para hacer que llevar el valor de toda la evidencia conjunta también a cero. En la sección 5.2 analizamos en detalle los motivos por los cuales hemos observado una evidencia tan baja.

En segundo lugar también es importante notar que si bien las medias geométricas de WHR y TTT se parecen, la evidencia a favor de WHR es unos ordenes de magnitud mayor que la TTT. En términos de probabilidad de los modelos dado los datos, WHR se queda con aproximadamente 99,985 % de la creencia y TTT con 0,015 %. Esta diferencia puede parecer mucha, pero no alcanza a descartar la hipótesis alternativa bajo las condiciones analizadas. Por ese motivo, en regla general no es recomendable rechazar hipótesis cuando la diferencia entre las evidencias de los modelos es menor a 20 o 10 ordenes de magnitud.

5.2. AGA/AAGo

En esta sección analizamos las razones por las cuales el modelo AGA/AAGo tuvo una evidencia tan baja: 0,4088 en promedio geométrico y -2963 de evidencia en escala logarítmica. Este fue un resultado sorprendente, pues no esperábamos ver una promedio geométrico por debajo de 0,5, que sería el desempeño de un modelo nulo que predice todas las partidas con 0,5 sin importar quien participe. Dedicamos un buen tiempo a verificar que no hubiera un error de cálculos por nuestra parte. Pudimos descartar esta hipótesis gracias a que nuestras estimaciones resultaron ser las mismas que las obtenidas por la AAGo. Tampoco se debió al bajo rendimiento de unas pocas partidas: analizando los resultados numéricos, encontramos malas predicciones en un porcentaje apreciable de los eventos. Entonces, las razones debían encontrarse en otro lado.

Luego observamos el comportamiento de las curvas de aprendizaje. En la Figura 5.2(a) mostramos las curvas de aprendizaje estimadas por el modelo AGA/AAGo para tres personas activas. Como los modelos

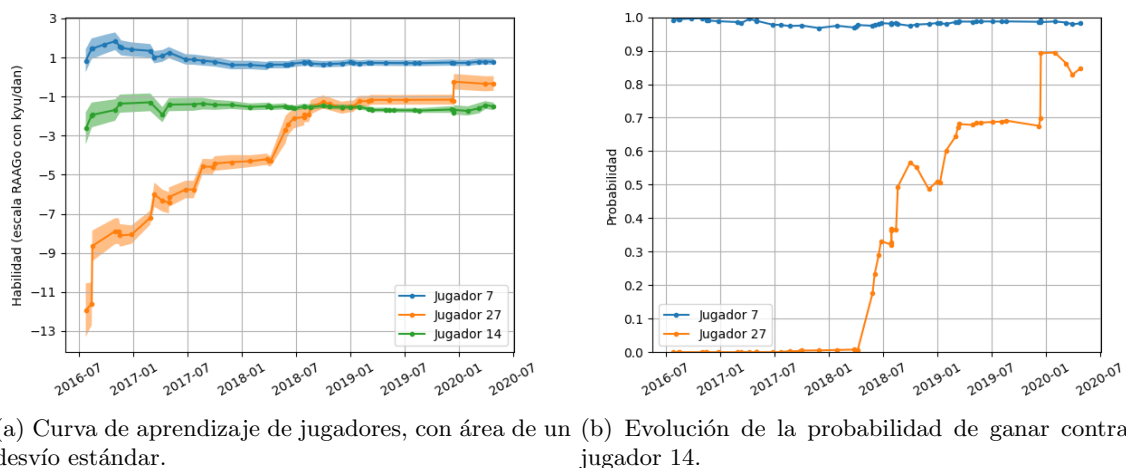


Figura 5.2: Evolución de las estimaciones de AAGo para tres personas activas.

no comparten la escala, una forma interesante de comprender mejor las curvas es utilizando la distancia en términos de probabilidad de ganar. Es decir, dado un punto de referencia, ¿cuál es la probabilidad de que los jugadores le ganen una partida sin *handicap* ni komi? En cada momento se compara a cada jugador con la habilidad estimada del jugador de referencia. En la Figura 5.2(b) mostramos las estimaciones de los jugadores 7 y 27, en términos de probabilidad de ganar, poniendo como referencia al jugador 14.

Observando su comportamiento se puede notar que las estimaciones varían solo al principio, produciéndose

se una estancamiento por pérdida de incertidumbre. El único jugador cuyas estimaciones no siguen esta regla es el 27. Sin embargo, analizando la base de datos, puede verse que esto se debe al uso de la autopromoción en múltiples ocasiones. Nuestra interpretación del gráfico coincide con las opiniones de los miembros de la asociación, que manifiestan que el modelo es “lento”.

Tanto la Asociación Estadounidense de Go, como la Argentina, intentan solucionar este problema a través de un sistema de autopromociones. Hemos descrito esto en la sección 3.2.2. En breve, se le pide a todas las personas que declaren su habilidad autopercibida y si se cumplen una serie de criterios durante el torneo (que sea verosímil la propuesta dados los resultados) entonces se reemplaza la estimación del modelo por la categoría declarada. Estas actualizaciones son consideradas en el computo de la evidencia. Es decir, no resuelven el problema.

Si el problema es la pérdida de incertidumbre, se nos ocurrió que aumentando el factor dinámico γ , que se agrega por el paso del tiempo, podría eventualmente mejorar las estimaciones del modelo. Para evaluar esta hipótesis, modificamos el código fuente del modelo AAGo con el objetivo de probar diferentes valores de incertidumbre dinámica y optimizar la evidencia con respecto a este valor. De esta forma, el factor dinámico paso de 0,0005 originalmente a 0,035 luego de optimizar, llevando el promedio geométrico de la evidencia a 0,5, y la evidencia en escala logarítmica a -2290 . Si bien esto mejoró cientos de órdenes de magnitud las predicciones con respecto al modelo original, los resultados siguen siendo insatisfactorio no solo en términos cuantitativos, sino también en términos cualitativos.

En la Figura 5.3 podemos observar el comportamiento de las curvas de aprendizaje estimadas por el modelo optimizado. A diferencia del modelo original, ahora las estimaciones tienen suficiente incertidumbre

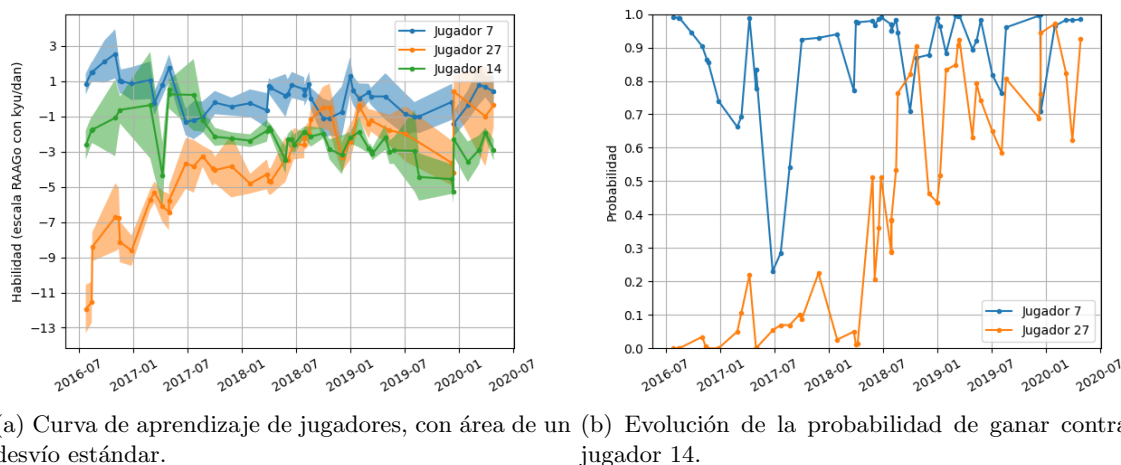


Figura 5.3: Evolución de las estimaciones de AAGo calculadas con el factor dinámico optimizado, para tres personas activas.

como para cambiar en el tiempo. Sin embargo, ahora las curvas contienen saltos sucesivos demasiado grandes como para ser verosímiles.

En vez de intentar encontrar el origen del problema en este modelo, decidimos trabajar sobre modelos que propagan la información histórica por todo el sistema.

5.3. TrueSkill Through Time

El modelo sobre el que trabajamos para desarrollar un estimador de habilidad alternativo para la Asociación Argentina de Go fue el modelo considerado estado del arte en la industria del video juego. Como explicamos en la sección 3.2.4, desarrollamos cuatro modelos alternativos a evaluar: TTT-h, TTT-h-k, TTT-hreg y TTT-hreg-kreg. Estos se diferencian en cómo modelan el efecto del *handicap* y el *komi* en el resultado de una partida.

Modelo	Evidencia por torneo	Evidencia por día de partida
TTT-h	−2093	−2036
TTT-h-k	−2113	−2050
TTT-hreg	−2097	−2026
TTT-hreg-kreg	−2113	−2038

Tabla 5.1: Resultados de los diferentes modelos de TTT integrando (no optimizando) los espacios de hipótesis relativos al *handicap* y el *komi*.

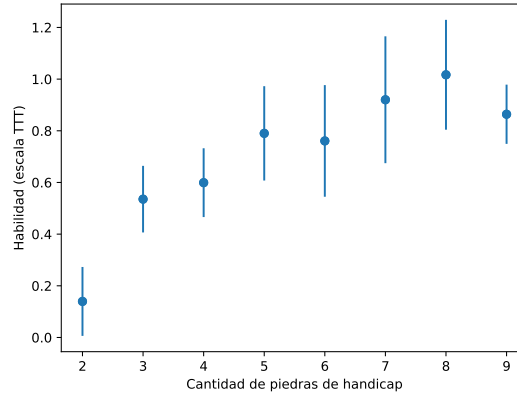


Figura 5.4: Estimaciones de *handicap* en TTT-h, agrupando partidas en tandas por torneo. Las barras representa un desvío estándar de incertidumbre. Puede verse que las estimaciones no serían fácilmente resumidas con un modelo lineal.

En la Tabla 5.1 pueden verse los resultados en términos de evidencia utilizando el criterio de *batches* por torneo (el utilizado originalmente por la AAGo) o por fecha de partida. Por un lado, vemos que procesar las partidas según el día en el que se juegan mejora la evidencia en todos los casos. Esto tiene sentido, ya que para estimar el resultado de una partida los sistemas pueden utilizar información de partidas anteriores en el mismo torneo.

Por último, vemos que con el criterio de tandas por día de partida, el mejor modelo de TTT resulta TTT-hreg, mientras que separando por torneos gana TTT-h. En el gráfico 5.4 vemos las estimaciones de habilidad para distinta cantidad de piedras de *handicap*, según TTT-h, utilizando el criterio de tandas por torneo. Puede verse que las estimaciones no serían fácilmente resumidas con un modelo lineal.

5.4. Whole History Rating y TrueSkill Through Time

Hemos explicado en la sección 3.2.1 que los modelos que propagan la información histórica por todo el sistema garantizan estimaciones iniciales fiables y comparabilidad entre estimaciones distantes en el tiempo. También mostramos que la evidencia de estos modelos es indudablemente mejor que la que se obtiene con el modelo de AGA/AAGo. Veamos ahora algunas diferencias entre WHR y TTT.

Whole History Rating (WHR) [Cou08] es un algoritmo de estimación de habilidad que se publicó luego de la aparición de TrueSkill Through Time (TTT) [DHMG07]. A diferencia de TTT que resuelve la inferencia usando aproximaciones que siempre minimizan la divergencia respecto del posterior exacto, WHR utiliza aproximaciones con técnicas clásicas: máxima verosimilitud, con estimación de incertidumbre usando la matriz de información de Fisher. Son modelos que resuelven el mismo problema usando enfoques diferentes.

Una diferencia importante entre ellos es que WHR no permite agregar factores adicionales que sean evaluados integralmente, sino que sus valores deben ser hardcoded. Debido a que los modelos de TTT evalúan el espacio integralmente, y esto penaliza naturalmente la complejidad del modelo, nos vimos

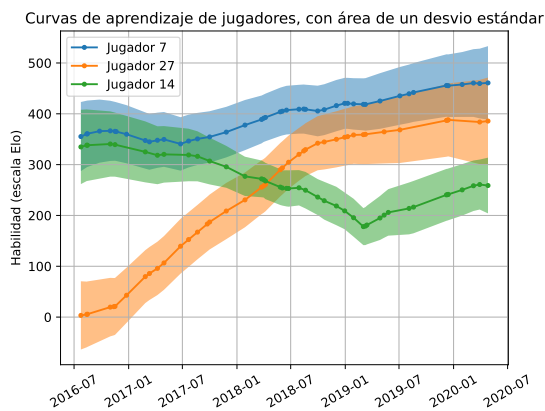
obligados a modificar la implementación de TTT para que también realizara una optimización como WHR. Es decir, cambiamos las variables aleatorias que modelan la habilidad lineal del *handicap* por dos hiperparámetros a optimizar, de forma que sea equivalente a lo hecho por WHR.

Esta comparación la hicimos mediante el criterio de *batches* por torneo. Usando WHR, tanto la pendiente como la ordenada al origen del modelo lineal encontraron óptimo valiéndose 10, con $\gamma = 6$. Por el lado de TTT, la pendiente óptima fue de 0,064 y la ordenada al origen de 0,034, mientras que el factor dinámico óptimo fue 0,025.

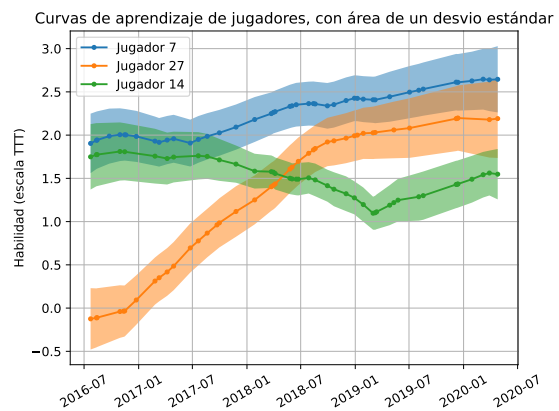
La diferencia de predicciones a priori entre los modelos optimizados otorgan una ventaja a favor de WHR de 9 órdenes de magnitud. A pesar de esta diferencia, consideramos que no es suficiente para dar el caso por cerrado. Es posible que evaluando ambos modelos en otras bases de datos, u modificando criterios que a primera vista pueden parecer insignificantes produzcan cambios en la evidencia de los modelos que cambie el orden relativo reportado en esta tesis.

De hecho, hasta último momento estuvimos convencidos de que existía una ventaja de por lo menos 20 órdenes de magnitud a favor de TTT, por lo que creíamos que la probabilidad de WHR había sido prácticamente anulada. Solo a último momento descubrimos una diferencia de criterio que dio vuelta el resultado. Por eso, durante el transcurso de la tesis centramos nuestra atención más en TTT que en WHR. En WHR no llegamos a evaluar los cuatro modelos que sí evaluamos con TTT. Si tuviéramos que comparar todos modelos entre TTT y WHR, es importante notar que el modelo TTT-h, que evaluado integralmente recibió una penalidad por complejidad, que si fuera optimizado superaría a TTT-hreg-optimizado.

Pasando a los resultados cualitativos, en ambos modelos las curvas de aprendizaje resultan más creíbles que en AAGo: la habilidad varía lentamente en el tiempo (Figuras 5.5(a) y 5.5(b)). Es destacable que esto se consigue sin utilizar la información de las auto promociones. Los modelos logran adaptarse a los cambios de forma automática, aprovechando mejor los datos de las partidas. Ambos modelos logran curvas muy similares entre si.

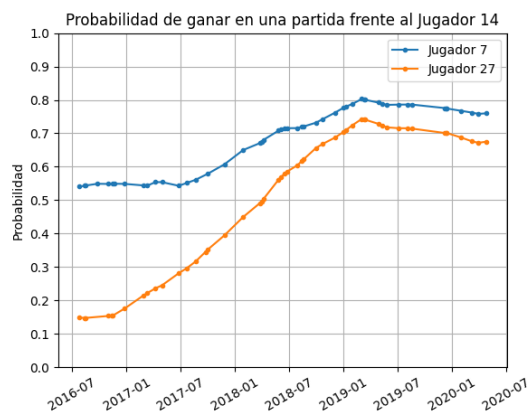


(a) Evolución de las estimaciones con WHR

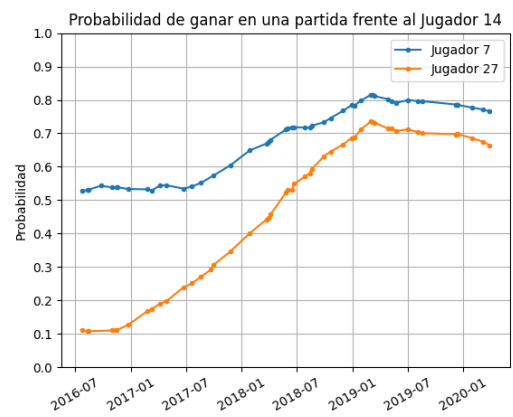


(b) Evolución de las estimaciones con TTT-hreg

En las Figuras 5.6(a) y 5.6(b) mostramos las estimaciones de los jugadores 7 y 27, en términos de probabilidad de ganar, poniendo como referencia al jugador 14. Las diferencias entre ambos modelos son pequeñas. Al principio, TTT-hreg asigna menor probabilidad al jugador 27, pero luego pasa a mostrar una curva prácticamente idéntica a la de WHR.



(a) Evolución de la probabilidad de ganarle al Jugador 14, calculado con las estimaciones de WHR



(b) Evolución de la probabilidad de ganarle al Jugador 14, calculado con las estimaciones de TTT-hreg

Capítulo 6

Conclusiones

En esta tesis nos propusimos estudiar y comparar distintos modelos de estimación de habilidad, contextualizados en los usos y problemáticas de la Asociación Argentina de Go (AAGo), con el objetivo de desarrollar y ofrecer un modelo superador. El sistema de ranking que utiliza actualmente la AAGo está basado en el estimador de habilidad de la Asociación Estadounidense de Go. Según la experiencia de los propios miembros de la AAGo, este sistema tiene serias dificultades para actualizar las habilidades, es “lento”. Este problema se agrava en el contexto de la AAGo debido a que al ser una asociación pequeña con una gran diversidad de habilidades se requiere un uso eficiente de los datos.

En esta tesis estudiamos, validamos y documentamos el sistema de ranking de la AAGo junto a otros dos modelos: el sistema WHR, que es actualmente un modelo de referencia para medir las habilidades de los jugadores profesionales de Go; y el sistema TTT, un modelo bayesiano considerado estado del arte en la industria del video juego. Pudimos mostrar que el sistema utilizado por la AAGo no hace uso eficiente de los datos en tanto propaga la información histórica en solo un sentido, del pasado al futuro. Los sistemas WHR y TTT, en cambio, al propagar la información histórica por todo el grafo de partidas garantizan estimaciones no sesgadas con baja incertidumbre en todo momento, incluso en las estimaciones iniciales, ofreciendo así mejores estimaciones de habilidad con menor cantidad de datos. Esto lo pudimos comprobar tanto en términos matemáticos, como a través de las simulaciones de validación.

Para poder utilizar los sistemas que llamamos de “historial completo” en el contexto del juego del Go fue necesario incluir en sus modelos el efecto de las ventajas (*handicap* y *komi*) en la estimación de habilidad. Originalmente WHR había sido publicado para estimar la habilidad de jugadores de Go profesionales, por lo que había sido diseñado sin incluir este tipo de efecto. Afortunadamente, encontramos que este punto había sido resuelto parcialmente por GoShrine, a través de la implementación del paquete `whole_history_rating` en Ruby (posteriormente traducido y mantenido en `Python`), adicionando el efecto del *handicap* al desempeño del jugador negro. Sin embargo, no existía una traducción directa de cantidad de piedras de *handicap* a su efecto esperado en el resultado. Decidimos resolver esto mediante un modelo lineal, cuyos coeficientes determinamos maximizando la evidencia del modelo. Por otra parte, el sistema TTT tampoco fue diseñado incluyendo los efectos de las ventajas de Go. En este caso diseñamos la solución completa de cero, que a diferencia de WHR es completamente bayesiana.

Para comparar el desempeño de los modelos utilizamos exclusivamente la base de datos original de la AAGo. La evaluación de los modelos se realizó aplicando estrictamente las reglas de la probabilidad, esto es calculando la predicción a priori de los datos D dado el modelo M , $P(D|M)$. De esta forma confirmamos que el modelo de la Asociación Estadounidense de Go, utilizado por la AAGo, tiene serias dificultades en la estimación de habilidad: el promedio geométrico de las predicciones fue menor a 0,5. Realizando un análisis minucioso del problema encontramos que las curvas de aprendizaje estimadas por la AAGo se caracterizan por perder rápidamente la incertidumbre, produciendo un estancamiento de las estimaciones. Los modelos de historial completo, por el contrario, resultan mejores tanto en términos cuantitativos como cualitativos: predicen mejor los resultados (en términos de la evidencia de los modelos) al mismo tiempo que estiman curvas más suaves y realistas.

Debido al bajo desempeño del modelo AAGo nos vimos obligados a realizar varias tareas para entender sus causas. En primer lugar, verificamos en detalle que el uso que le estuviéramos dando al modelo fuera correcto. Una vez que rechazamos un error de nuestra parte en la implementación, exploramos algunas explicaciones relativas al modelo en sí mismo. La primera intuición fue que si las estimaciones tenían tendencia a estancarse, entonces esto podría deberse a una falta de incertidumbre dinámica, que se agrega debido al paso del tiempo. La optimización de este parámetro efectivamente resultó ser un orden de magnitud mayor, lo que llevó a que la media geométrica de las predicciones llegara a 0,5. Sin embargo, este número sigue sin ser satisfactorio para un modelo de estimación de habilidad. Además, las curvas de aprendizaje resultan poco suaves, con cambios abruptos. Si bien ninguna de estas exploraciones permitieron resolver los problemas estructurales del modelo de la AAGo, sí nos sirvieron para reforzar el convencimiento sobre el objetivo original: la necesidad de ofrecer un modelo alternativo que mejore sustancialmente las estimaciones.

Ambos modelos de historial completo cumplieron el objetivo de mejorar sustancialmente las estimaciones y las curvas de aprendizaje obtenidas con el sistema de la AAGo. En términos de media geométrica y visualización de las curvas de aprendizaje, los resultados de WHR y TTT-hreg son similares. Sin embargo, si se comparan en términos de la probabilidad del modelo dados los datos, se observa una ventaja a favor de WHR. Es importante tener en cuenta que TTT-hreg se ve penalizado en la evidencia por las variables agregadas para modelar el *handicap* (mientras WHR utiliza hiperparámetros). Al pasar estas variables a hiperparámetros optimizados, TTT-hreg acorta la distancia con WHR, haciendo que quede poco claro que la diferencia sea significativa. Por otro lado, TTT cuenta con la ventaja de permitir agregar nuevos factores respetando las reglas de la probabilidad, sin necesidad de salirse del marco bayesiano, simplificando su uso y adaptación a nuevos contextos. Por estos motivos, nuestra recomendación para la Asociación Argentina de Go es elegir el modelo que mejor les convenga en términos prácticos y nuestra recomendación para posibles extensiones de esta tesis es volver a evaluar el desempeño de ambos modelos, WHR y TTT.

Consideramos que el objetivo original de la tesis se ve cumplido. Ofrecimos un modelo que mejora significativamente las estimaciones y las predicciones, que es eficiente en el uso de datos, y que puede ser utilizado en el futuro como base para un estimador internacional.

Durante nuestro análisis, surgieron algunas líneas de investigación que quedaron sin finalizar. La más importante tiene que ver con la posibilidad de ofrecer estimaciones de habilidad que sean comparables en términos internacionales. Todos los modelos producen estimaciones de habilidad que solo adquieren significado cuando se las compara con otras estimaciones de habilidad. Es decir, el valor absoluto de las estimaciones es arbitrario. En la actualidad no es posible comparar las estimaciones de habilidad obtenidas por las diferentes asociaciones de Go del mundo, no sólo porque utilizan sistemas diferentes, sino porque en general son bases de datos desconexas. Sin embargo, en el capítulo 4 mostramos con experimentos sintéticos que los modelos de historial completo son capaces de ajustar el valor absoluto utilizado por comunidades disjuntas si se ofrece al modelo algunas cuantas partidas intercomunitarias. Unos pocos vínculos entre bases de datos le permitirían a nuestra solución propagar la información eficientemente, garantizando la comparabilidad entre las diversas asociaciones internacionales.

En el transcurso de esta tesis, creamos dos bases de datos a partir de sitios de internet reconocidos internacionalmente: KGS¹ y OGS². En este tipo de plataformas suelen jugar personas de todos los países, varias de las cuales son miembros de las asociaciones de Go locales. En particular, obtuvimos los identificadores KGS y OGS de 20 miembros de la AAGo que permitirían realizar las estimaciones de habilidad integrando todas las bases de datos. Esta integración, además de mejorar las estimaciones que actualmente tiene la AAGo, por aumento de la cantidad de datos, le serviría a la asociación para tener estimaciones comparables a nivel internacional. Además, en un futuro puede servir de punto de contacto entre asociaciones, de modo de poder crear una plataforma internacional de estimaciones para jugadores de Go, que permita comparar tanto en el tiempo como entre jugadores.

Además de los datos recolectados por nosotros, los administradores de la plataforma KGS nos entregaron su base de datos completa con casi 100 millones de partidas. Este conjunto de datos, de más de 7 GB, es especialmente desafiante ya que no es posible correr los modelos en computadoras comunes y abre la

¹<https://www.gokgs.com/>

²<https://online-go.com/>

puerta a pensar optimizaciones a los paquetes de inferencia utilizados. Además de servir para el objetivo mencionado en el párrafo anterior, permite analizar en detalle ciertas hipótesis que tienen los jugadores expertos de Go. Hemos visto que la escala tradicional intenta separar las categorías por una distancia equivalente al efecto que produce una piedra de ventaja. Sin embargo, se cree que el efecto de la piedra sobre la probabilidad de ganar no es constante, sino que depende de la habilidad de la persona que recibe la ventaja. Se supone que las personas expertas son capaces de sacarle mayor provecho a la ventaja, por lo que la distancia entre categorías tradicionales debería agrandarse cada vez más a medida que las personas adquieren mayor habilidad. Sería de interés estudiar este problema, junto a nuevos modelos no lineales de efecto de *handicap*.

El modelo que hemos desarrollado, TTT-hreg, junto a la base de datos completa de KGS ofrece la condiciones ideales para resolver ambos problemas: crear un estimador internacional, y estimar la relación no lineal de las escalas tradicionales.

Bibliografía

- [BREP04] William M Baum, Peter J Richerson, Charles M Efferson, and Brian M Paciotti. Cultural evolution in laboratory microsocieties including traditions of rule giving and rule following. 25(5):305–326, September 2004.
- [BRH11] Robert Boyd, Peter J Richerson, and Joseph Henrich. The cultural niche: Why social learning is essential for human adaptation. 108(Supplement 2):10918–10925, June 2011.
- [BT52] Ralph Allan Bradley and Milton E Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345, December 1952.
- [CHJH02] Murray Campbell, A Joseph Hoane Jr, and Feng-hsiung Hsu. Deep blue. *Artificial intelligence*, 134(1-2):57–83, 2002.
- [Cou08] Rémi Coulom. Whole-history rating: A bayesian rating system for players of time-varying strength. In *of the 6th on Computers and Games*, pages 113–124, Berlin, Heidelberg, 2008. Springer-Verlag.
- [Cox46] Richard T Cox. Probability, frequency and reasonable expectation. 14(1):1–13, 1946.
- [DHMG07] Pierre Dangauthier, Ralf Herbrich, Tom Minka, and Thore Graepel. Trueskill through time: Revisiting the history of chess. In John C. Platt, Daphne Koller, Yoram Singer, and Sam T. Roweis, editors, *of the Twenty-First Annual on Neural Information Processing Systems*, pages 337–344. Curran Associates, Inc., 2007.
- [Elo08] Arpad E Elo. *The Rating of Chess Players, Past and Present*. Ishi Press, May [1978] 2008.
- [Gli99] Mark E. Glickman. Parameter estimation in large dynamic paired comparison experiments. 48(3):377–394, 1999.
- [Gul88] Stephen F. Gull. *Bayesian Inductive Inference and Maximum Entropy*, pages 53–74. Springer Netherlands, 1988.
- [Hal17] Joseph Y Halpern. *Reasoning about uncertainty*. MIT press, 2nd edition, 2017.
- [HB20] Sarah Blaffer Hrdy and Judith M Burkart. The emergence of emotionally modern humans: implications for language and learning. *Philosophical Transactions of the Royal Society B*, 375(1803):20190499, 2020.
- [HMG06] Ralf Herbrich, Ralf Minka, and Thore Graepel. Trueskill: A bayesian skill rating system. In *of the in Advances in Neural Information Processing Systems*, pages 569–576, Cambridge, MA, 2006. MIT Press.
- [Jef39] H. Jeffreys. *Theory of Probability*. Oxford University Press, 1939.
- [KFL01] Frank R Kschischang, Brendan J Frey, and H-A Loeliger. Factor graphs and the sum-product algorithm. 47(2):498–519, 2001.
- [KMH⁺20] Jeremy Koster, Richard McElreath, Kim Hill, Douglas Yu, Glenn Shepard, Nathalie Van Vliet, Michael Gurven, Benjamin Trumble, Rebecca Bliege Bird, Douglas Bird, Brian Coddington, Lauren Coad, Luis Pacheco-Cobos, Bruce Winterhalder, Karen Lupo, Dave Schmitt, Paul Sillitoe, Margaret Franzen, Michael Alvard, Vivek Venkataraman, Thomas Kraft, Kirk Endicott, Stephen Beckerman, Stuart Marks, Thomas Headland, Margaretha Pangau-Adam, Andres Siren, Karen Kramer, Russell Greaves, Victoria Reyes-García, Maximilien Guèze, Romain Duda, Álvaro Fernández-Llamazares, Sandrine Gallois, Lucentezza Napitupulu, Roy Ellen, John Ziker, Martin Nielsen, Elspeth Ready, Christopher Healey, and Cody Ross. The life history of human foraging: Cross-cultural and individual variation. *Science Advances*, 6(26), 2020.
- [KR95] Robert E Kass and Adrian E Raftery. Bayes factors. *Journal of the american statistical association*, 90(430):773–795, 1995.
- [LM21] Gustavo Landfried and Esteban Mocskos. Trueskill through time: reliable initial skill estimates and historical comparability with julia, python and r. <https://github.com/glandfried/TrueSkillThroughTime/releases/download/doc/landfried-learning.pdf>, 2021.
- [Mac91] David J. C. MacKay. Bayesian model comparison and backprop nets. In *Proceedings of the 4th International Conference on Neural Information Processing Systems, NIPS’91*, page 839–846, San Francisco, CA, USA, 1991. Morgan Kaufmann Publishers Inc.
- [Mac99] David J. C. MacKay. Comparison of approximate methods for handling hyperparameters. *Neural Comput.*, 11(5):1035–1068, July 1999.
- [Min01] Thomas P Minka. Expectation propagation for approximate bayesian inference. In *Conference on Uncertainty in Artificial Intelligence*, pages 362–369. Morgan Kaufmann Publishers Inc., 2001.

- [MMH16] Michael Muthukrishna, Thomas JH Morgan, and Joseph Henrich. The when and who of social learning and conformist transmission. *Evolution and Human Behavior*, 37(1):10–20, 2016.
- [Mos51a] Frederick Mosteller. Remarks on the method of paired comparisons: I. the least squares solution assuming equal standard deviations and equal correlations. *Psychometrika*, 16(1):3–9, March 1951.
- [Mos51b] Frederick Mosteller. Remarks on the method of paired comparisons: Ii. the effect of an aberrant standard deviation when equal standard deviations and equal correlations are assumed. *Psychometrika*, 16(2):203–206, June 1951.
- [Mos51c] Frederick Mosteller. Remarks on the method of paired comparisons: Iii. a test of significance for paired comparisons when equal standard deviations and equal correlations are assumed. *Psychometrika*, 16(2):207–218, June 1951.
- [RFH⁺11] Luke Rendell, Laurel Fogarty, William JE Hoppitt, Thomas JH Morgan, Mike M Webster, and Kevin N Laland. Cognitive culture: theoretical and empirical insights into social learning strategies. *Trends in cognitive sciences*, 15(2):68–76, 2011.
- [SHM⁺16] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.
- [Thu27] Louis L Thurstone. Psychophysical analysis. 38(3):368–389, 1927.
- [Wal10] Philip Waldron. The aga ratings system. 2010.
- [Zer13] Ernst Zermelo. The calculation of the results of a tournament as a maximum problem in the calculus of probability (translation of 1929 paper). In Heinz-Dieter Ebbinghaus and Akihiro Kanamori, editors, *Ernst Zermelo - Collected Works Volume II*, Schrifmath. [1929] 2013.