

Trabajo practico 1

The Multiarmed Bandit

Integrantes: Tomas Anderson, Alejo Vaschetti, Manuel Moresi

Profesores: Tomas Capretto, Ignacio Evangelista

Fecha de entrega: 17/04/2024

Introduccion

To do

Desarrollo

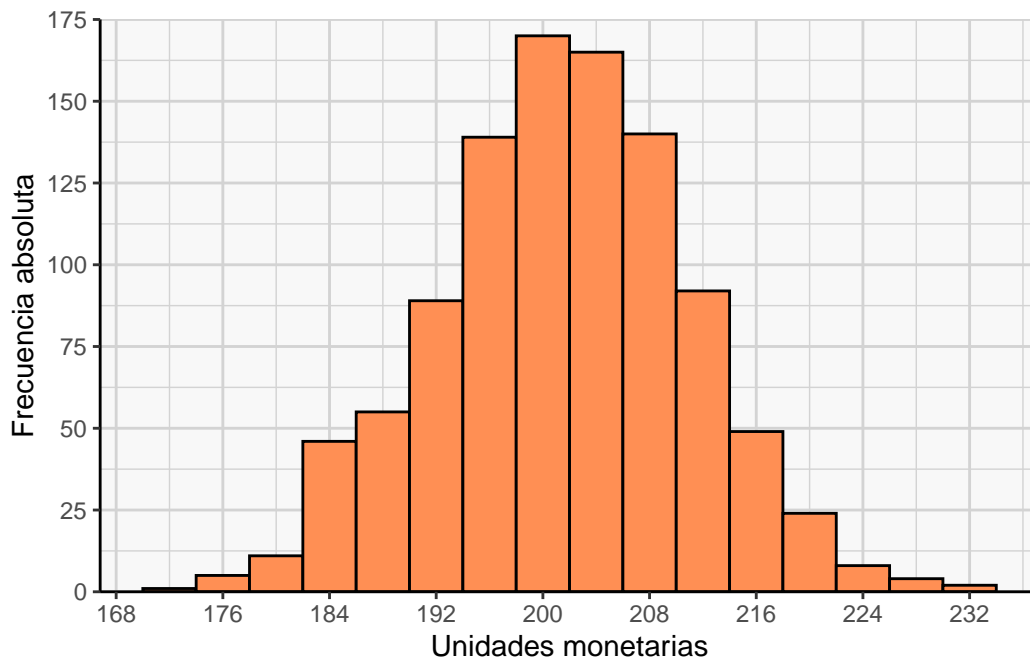
Cosas que se corren siempre

Situación ideal

La situación ideal sería conocer de antemano cual máquina es la que tiene la mayor probabilidad de ganar y jugar siempre con ella. Esta sería la mejor forma de ganar dinero, por lo que una buena estrategia tiene que asemejarse a dicha situación.

Para ver la cantidad de dinero que gana en promedio este método se simulan a 1000 personas jugando 1 año utilizándolo.

Estos fueron los resultados:



En este modus operandi una persona que juegue un año se espera que va a ganar alrededor de 201.986 unidades monetarias.

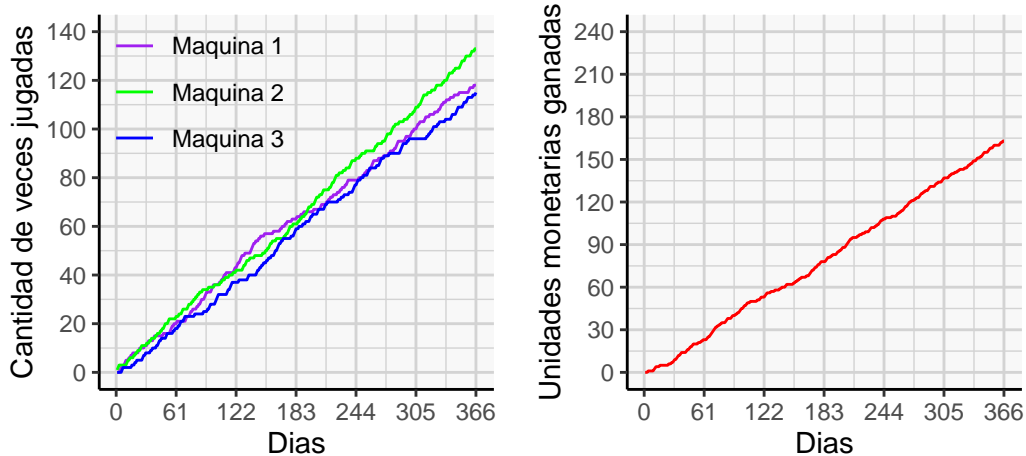
En el 95% de los años simulados las ganancias totales se encontraron entre 184 y 219 unidades monetarias. Dicho intervalo, y los que serán presentados en las próximas estrategias, se encuentran entre el 2,5% y el 97,5% de las ganancias totales del método ordenadas de menor a mayor.

Para ver que tan beneficioso es cada estrategia, se comparan la cantidad de veces que se gana en promedio esperando que se acerque lo máximo posible a esta forma de jugar. Además, se contrastan los límites de los intervalos propuestos.

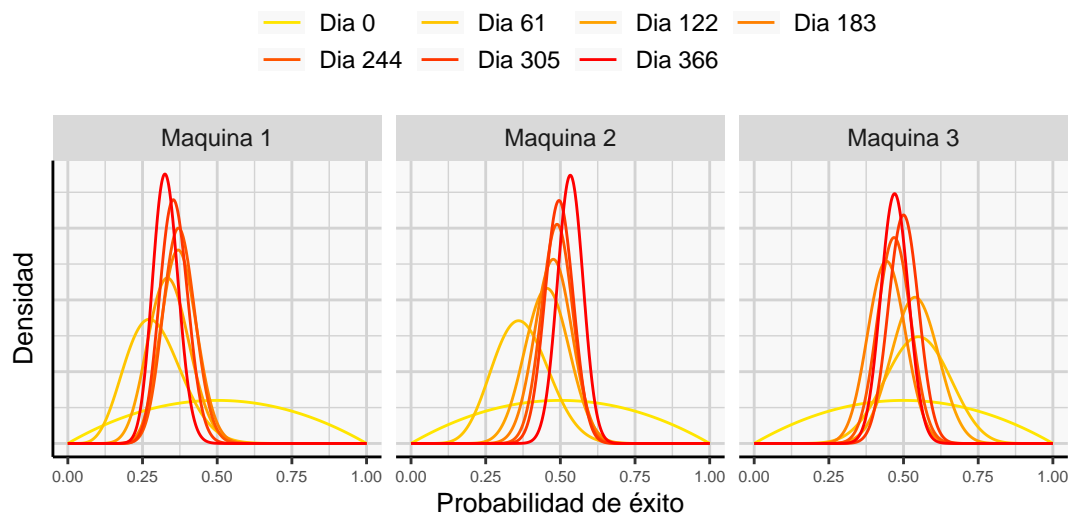
Estrategia 1: Completamente al azar

Esta estrategia consiste en elegir cada día al azar qué máquina jugar, teniendo las tres la misma probabilidad de ser selectas. Esta es la estrategia más básica, ya que no requiere ningún tipo de información previa ni evoluciona en el tiempo.

Se juega un año con esta estrategia y se obtienen los siguientes resultados:

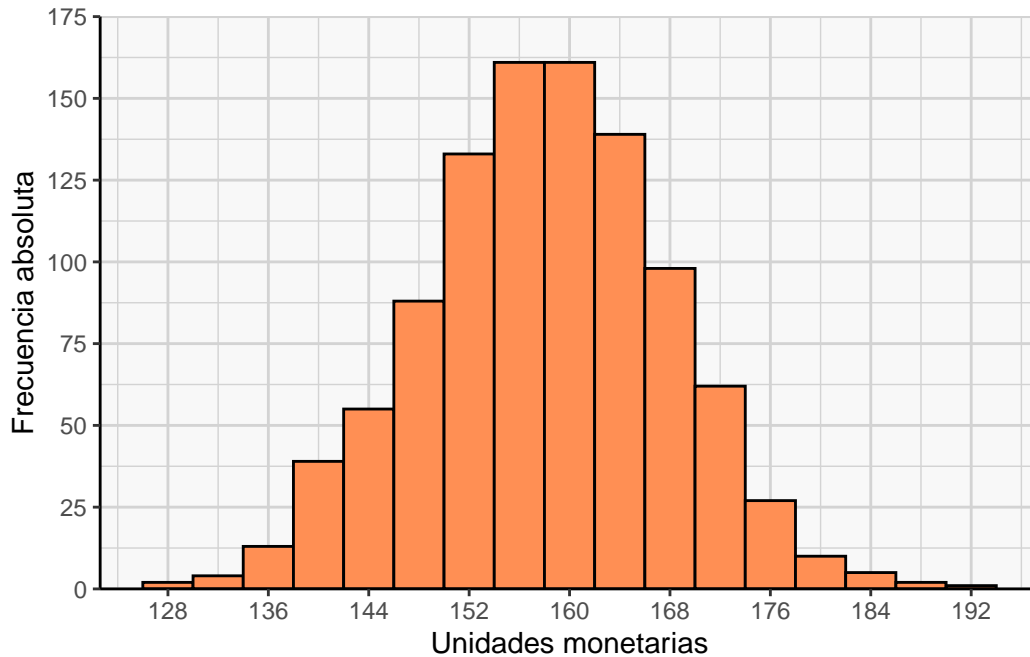


La cantidad de veces que se juegan las 3 máquinas son similares, y la diferencia entre ellas se debe puramente al azar. Este hecho impacta en la evolución de la creencia sobre la probabilidad de éxito de cada máquina, como se ve a continuación.



Al jugarse una cantidad similar de veces en cada máquina la creencia se concentra en igual medida alrededor de los valores de las probabilidades de éxito de cada una. Estos son 0.3278689, 0.5328467 y 0.4705882 para la máquina 1, 2 y 3 respectivamente.

Las simulaciones de los 1000 años arrojaron los siguientes resultados:



En promedio, siguiendo esta estrategia, se espera ganar 158.499 unidades monetarias en un año.

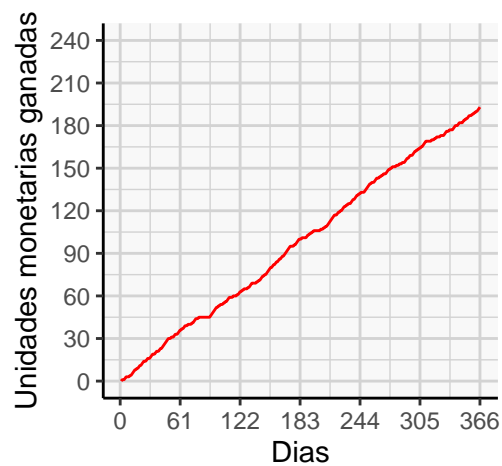
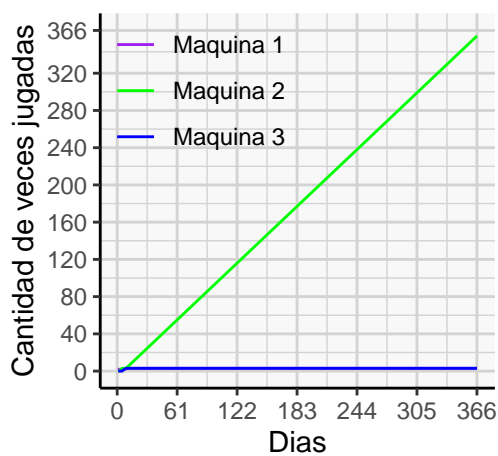
El 95% de las ganancias al final del año estuvieron entre 140 y 177.

El promedio de ganancias al final del año es sustancialmente mas chico que el de la situacion ideal y los intervalos siquiera se superponen. Se puede pensar que esta estrategia no es recomendable al diferir tanto de un resultado ideal.

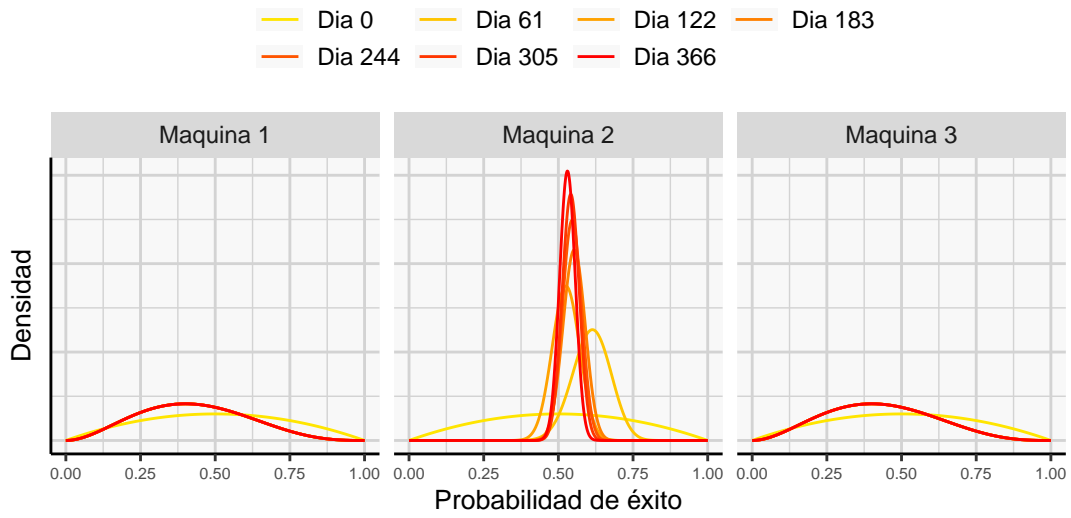
Estrategia 2: Greedy con tasa observada

La maquina que esta estrategia selecciona en un cierto dia es aquella que tenga la mayor probabilidad de éxito observada hasta el momento. Como al empezar el año no se tiene ninguna informacion sobre la probabilidad de éxito de cada maquina, se decide jugar cada una por separado hasta que todas tengan su primer éxito.

Se juega un año con esta estrategia y se obtienen los siguientes resultados:

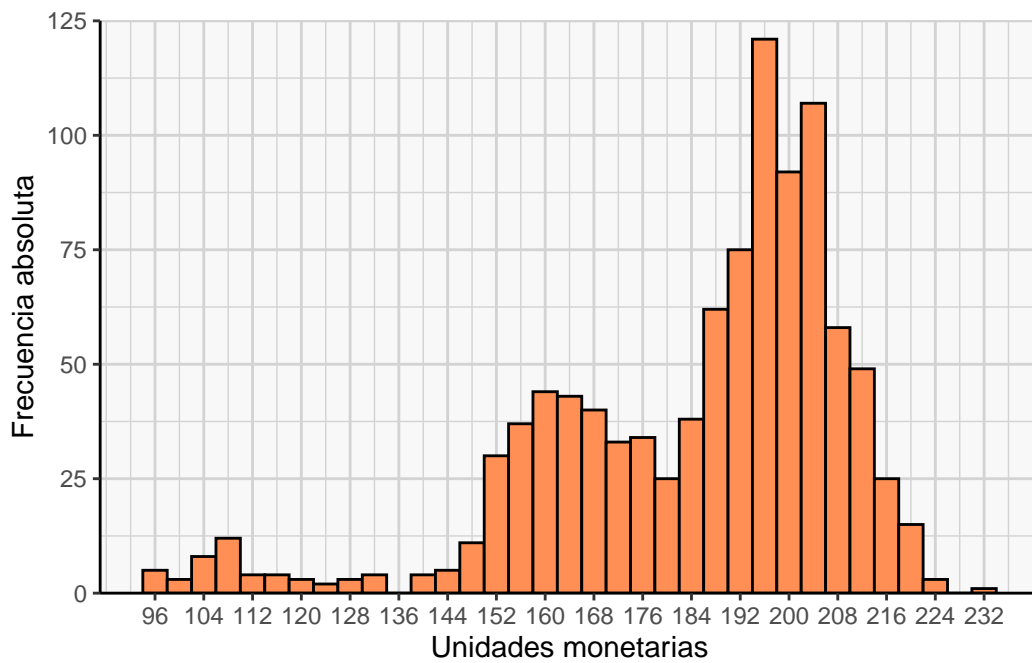


En este año se usa casi exclusivamente la maquina 2 y apenas se utilizan las otras dos. Esto pasa porque las probabilidades de exitos iniciales observadas son muy variables, por lo que la primera maquina en tener varios fracasos va a tener una probabilidad muy chica y se dejara de usar. Esto impacta a la creencia sobre la probabilidad de éxito de cada máquina de la siguiente manera.



La creencia sobre la probabilidad de exito de la maquina 2 va a ser muy precisa, al tener muchas muestras de ella, y las demas seran muy parecidas a la suposicion inicial.

Las simulaciones de los 1000 años arrojaron los siguientes resultados:



En promedio, siguiendo esta estrategia, se espera ganar 185.105 unidades monetarias en un año.

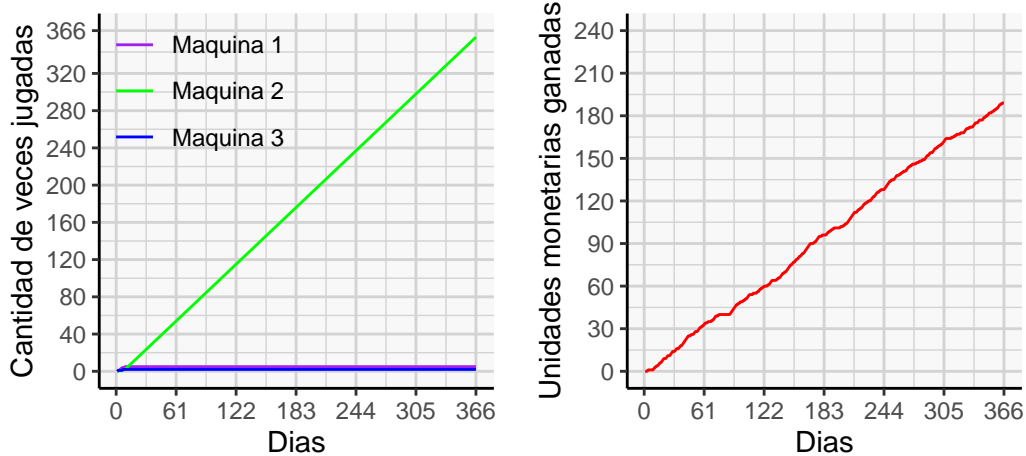
El 95% de las ganancias al final del año estuvieron entre 109 y 218.

Si bien es un metodo que en promedio gana una cantidad similar de dinero comparado a la situacion ideal, existe el riesgo de ganar mucho menos que el promedio, ya que en general la estrategia suele usar una misma maquina en casi todo el año y hay una probabilidad no despreciable de elegir la peor de ellas.

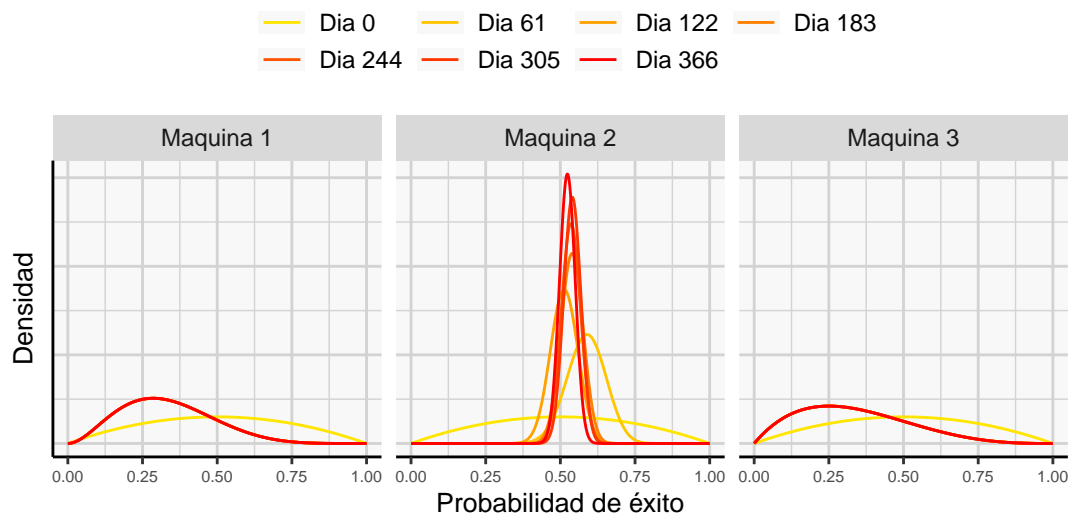
Estrategia 3: Greedy con probabilidad a posteriori

En esta estrategia, la maquina a elegir en cada dia es aquella que, hasta el momento, tenga la mayor esperanza entre las distribuciones a posteriori de las probabilidades de éxito.

Se juega un año con esta estrategia y se obtienen los siguientes resultados:

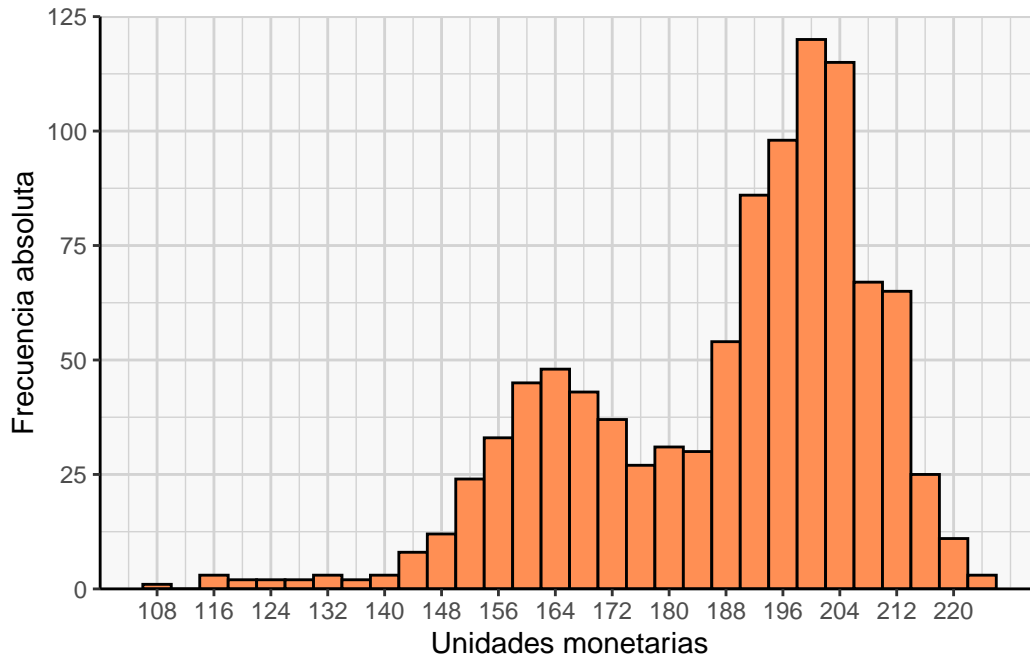


La seleccion de las maquinas en el año simulado es notablemente similar a la estrategia anterior. La unica diferencia es que las maquinas 1 y 3 se usaron un poco mas al principio del año.



Este metodo genera muchas observaciones de una maquina e ignora las demas, debido a su parecido con la estrategia anterior, por ende, la precision sobre la creencia de la probabilidad de exito sera grande para una maquina y chica en las otras.

Las simulaciones de los 1000 años arrojaron los siguientes resultados:



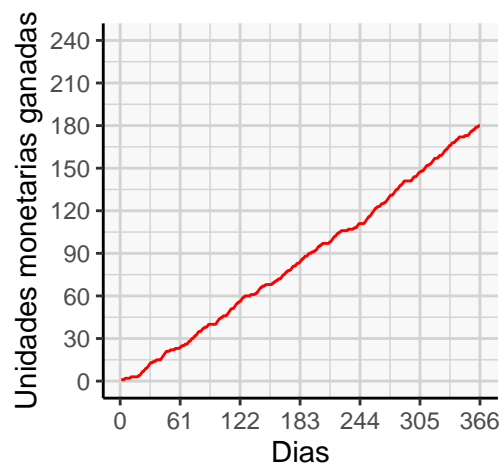
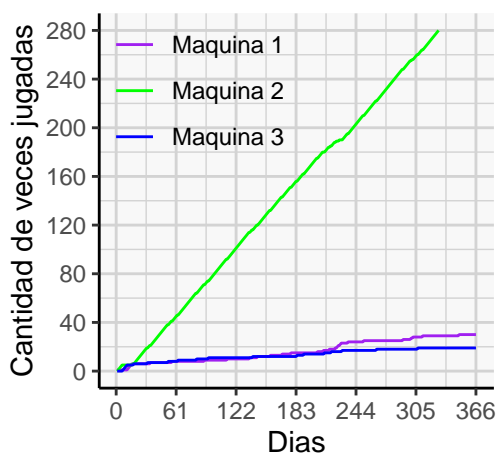
En promedio, siguiendo esta estrategia, se espera ganar 188.532 unidades monetarias en un año.

El 95% de las ganancias al final del año estuvieron entre 145 y 216.

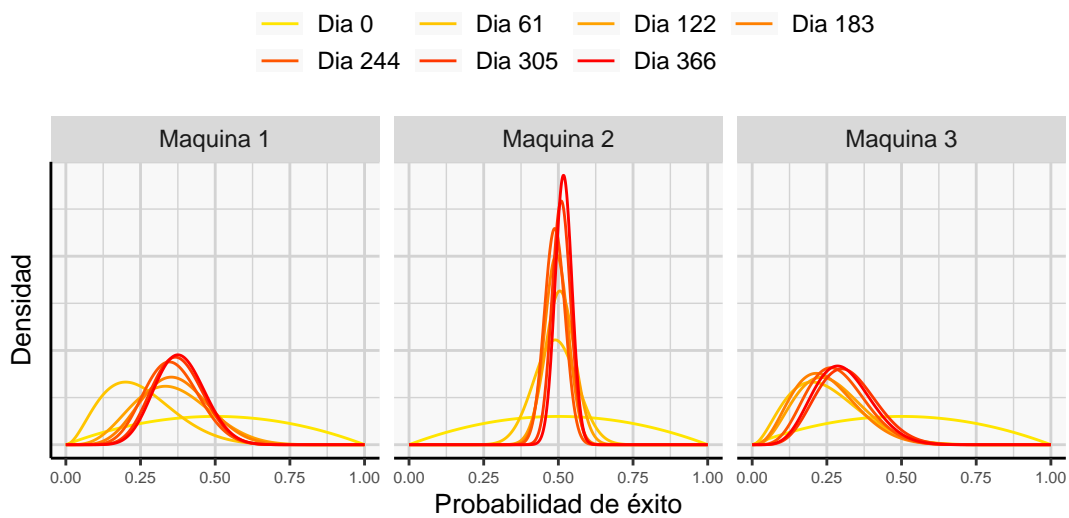
Esta estrategia sera un poco mejor que la anterior debido a que la cola izquierda del histograma es menos densa. Esto se debe a que al principio de cada año se tiene una mejor aproximacion de las probabilidades de éxito al haber utilizado la informacion de la distribucion a priori.

Estrategia 4: ϵ -greedy (con tasa observada)

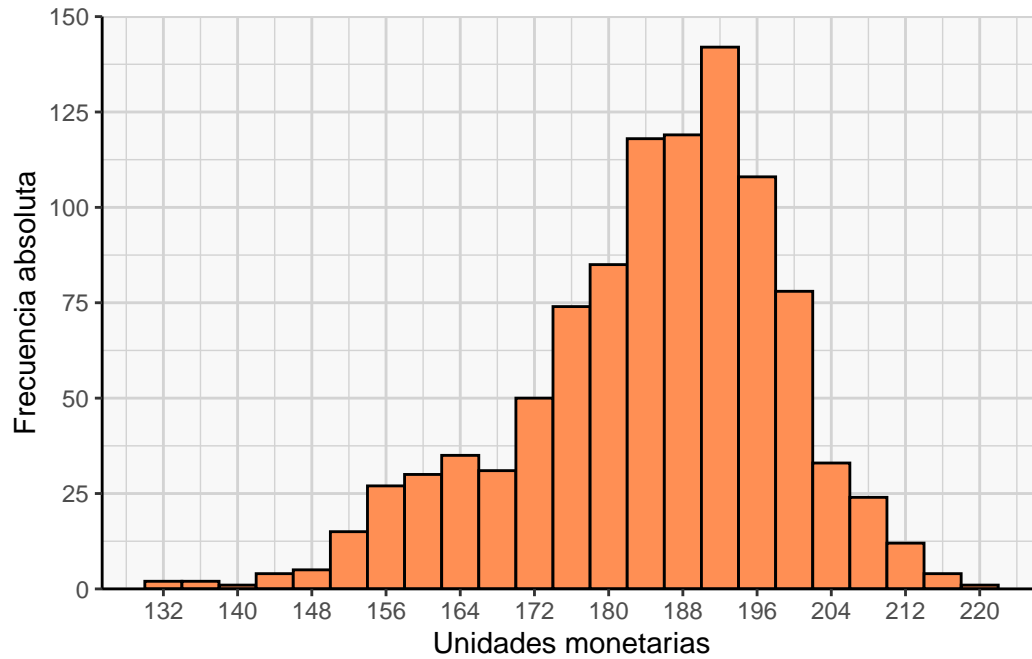
Se juega un año con esta estrategia y se obtienen los siguientes resultados:



Posteriors:

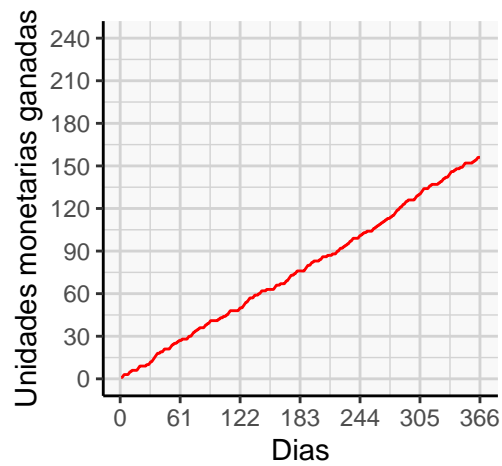
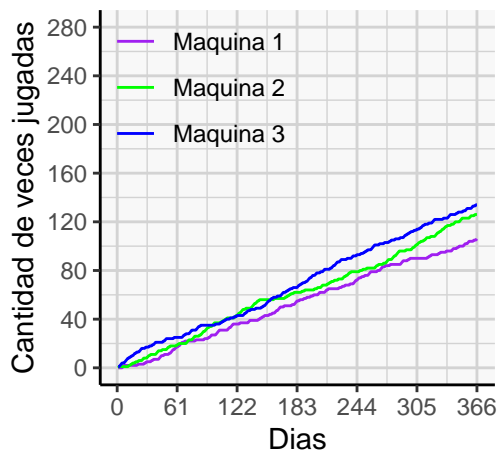


1000 años:

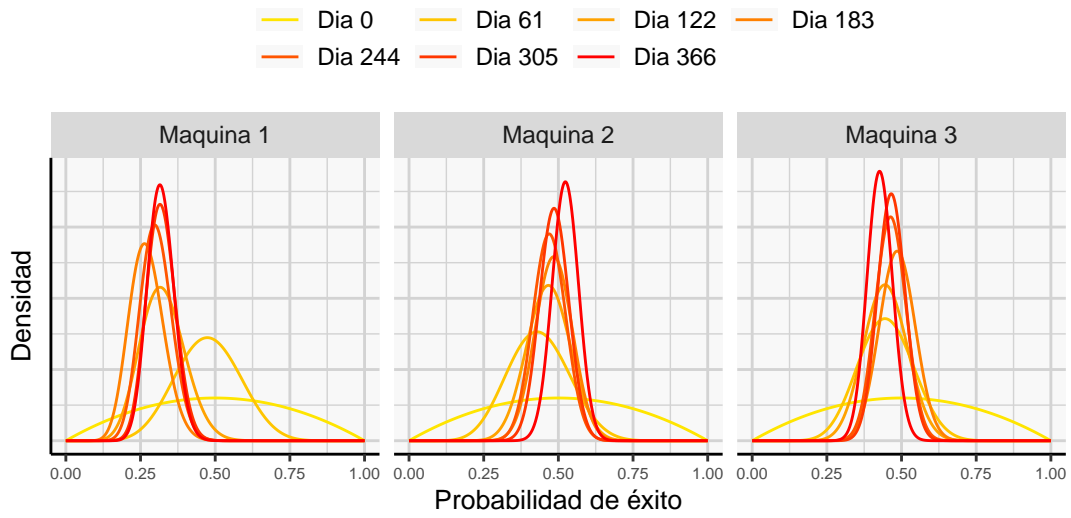


Estrategia 5: Softmax

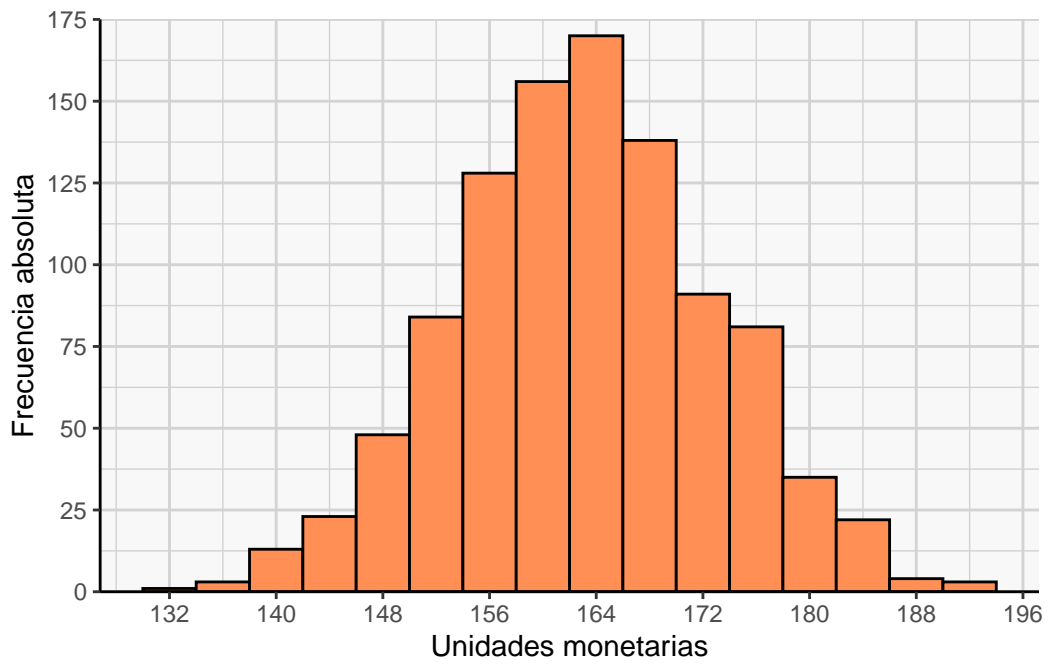
Se juega un año con esta estrategia y se obtienen los siguientes resultados:



Posteriors:

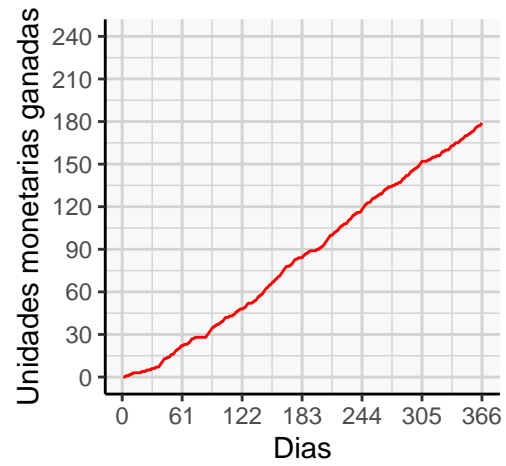
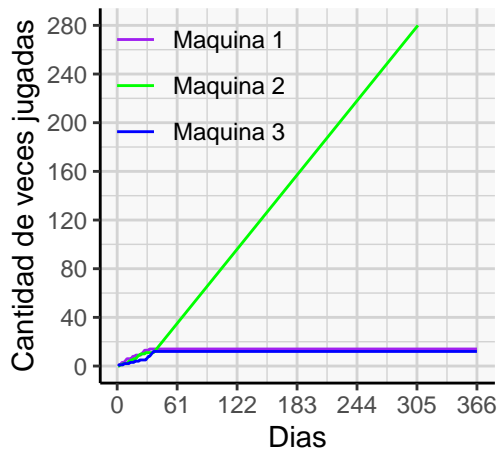


1000 años:

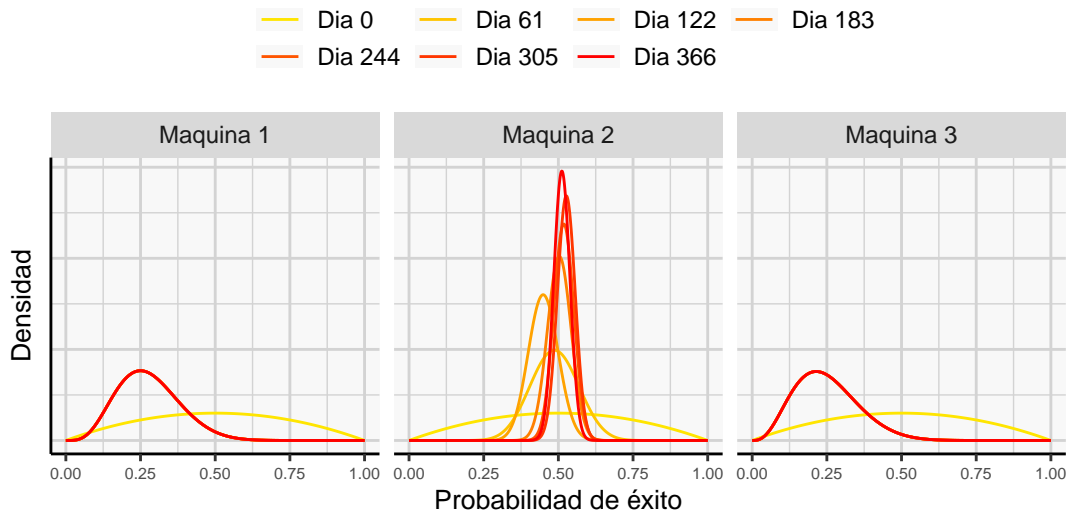


Estrategia 6: Upper-bound

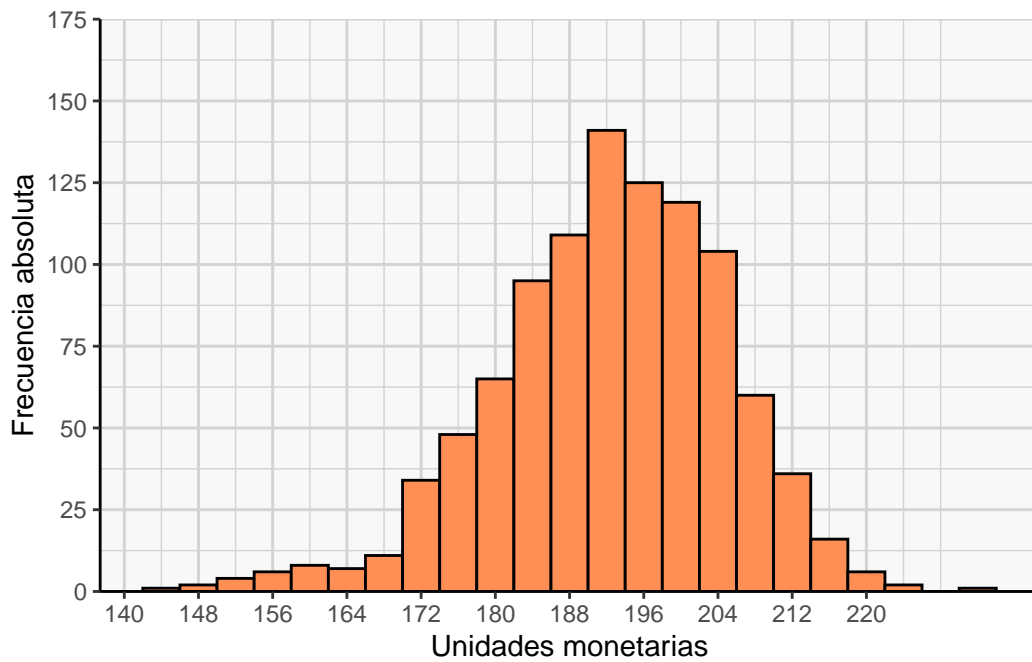
Se juega un año con esta estrategia y se obtienen los siguientes resultados:



Posteriors:



1000 años:



Estrategia 7: Thompson sampling