

Trabajo practico 1

The Multiarmed Bandit

Integrantes: Tomas Anderson, Alejo Vaschetti, Manuel Moresi

Profesores: Tomas Capretto, Ignacio Evangelista

Fecha de entrega: 17/04/2024

Introducción

To do

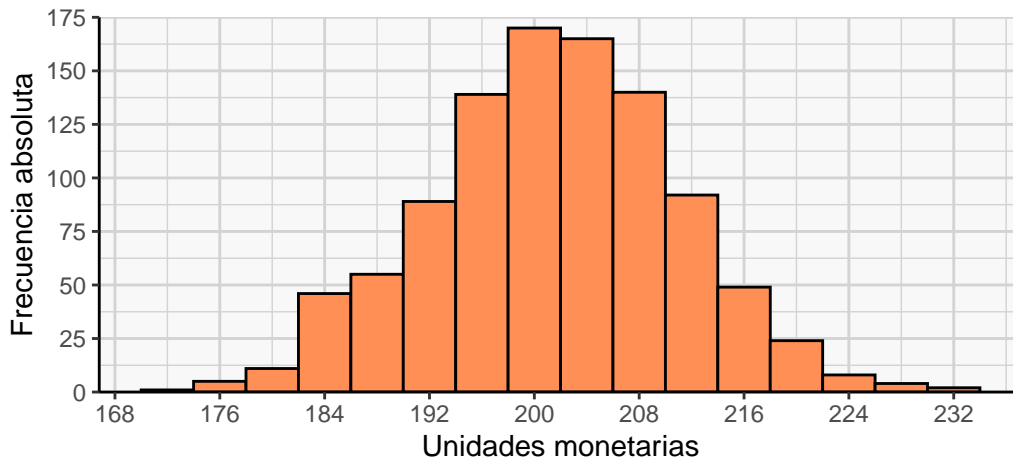
Desarrollo

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 modo de operación una persona que juegue un año se espera que va a ganar alrededor de 201.986 unidades monetarias.

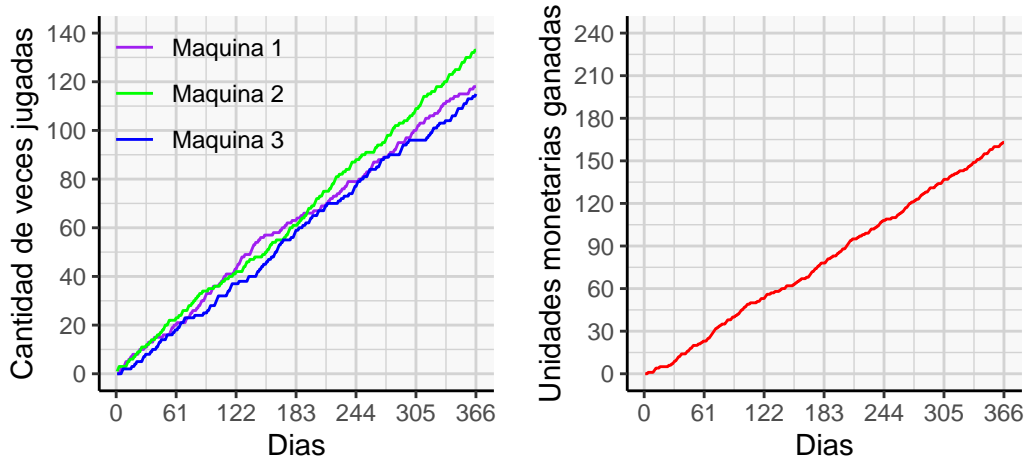
En el 95% de los años simulados las ganancias totales se encontraron entre 184 y 219 unidades monetarias. Dicho intervalo, y los que serán presentados en las próximas estrategias, se encuentran entre el 2,5% y el 97,5% de las ganancias totales del método ordenadas de menor a mayor.

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.

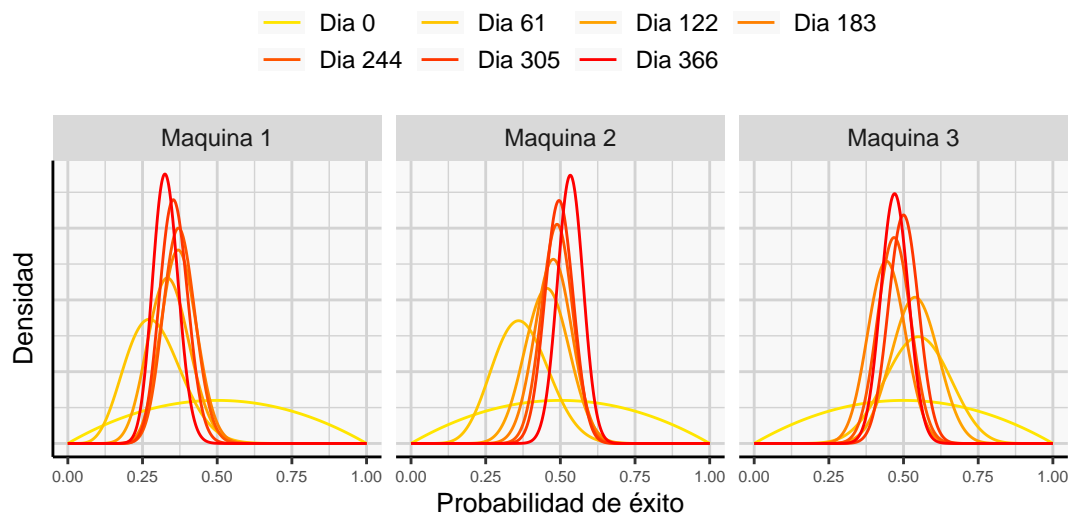
Estrategia 1: Completamente al azar

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 mas basica, ya que no requiere ningun tipo de informacion previa ni evoluciona en el tiempo.

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

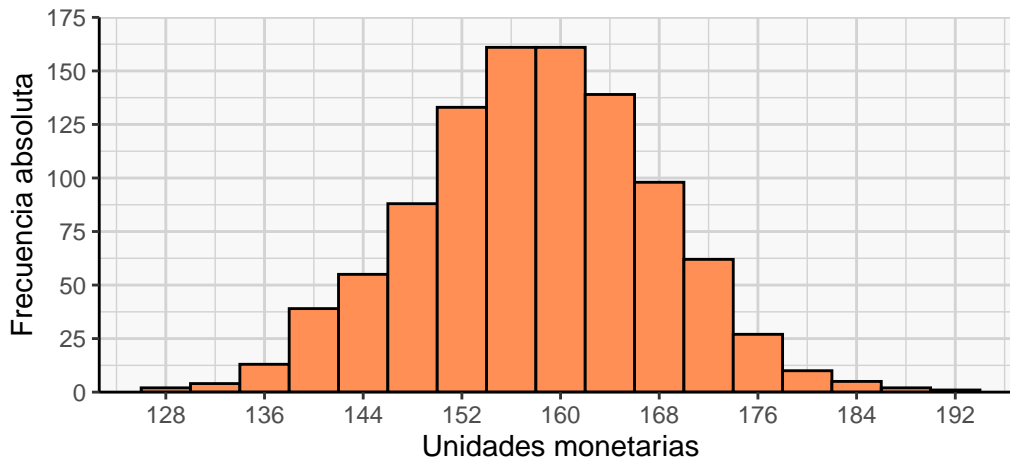


La cantidad de veces que se juegan las 3 máquinas son similares, y la diferencia entre ellas se debe puramente al azar. Este hecho impacta en la evolución de la creencia sobre la probabilidad de éxito de cada máquina, como se ve a continuación.



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 158.5 unidades monetarias en un año.

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

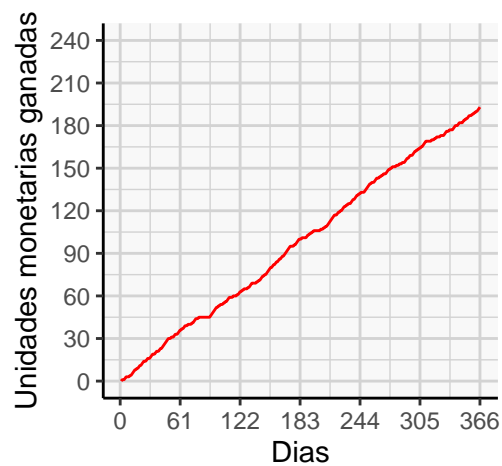
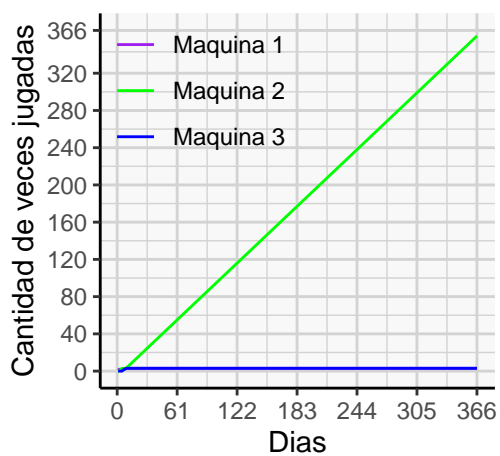
	Maquina 1	Maquina 2	Maquina 3
Media	122.04	122.27	121.69
Mediana	122.00	122.00	122.00

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.

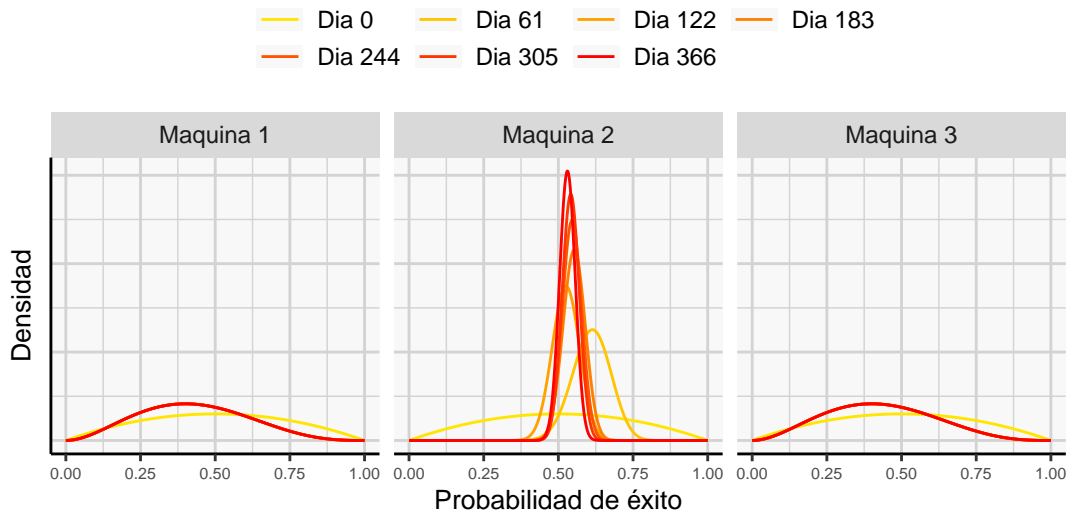
Estrategia 2: Greedy con tasa observada

La maquina que esta estrategia selecciona en un cierto día 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. Si las probabilidades de éxito coinciden en algunas maquinas se elige una de ellas aleatoriamente.

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

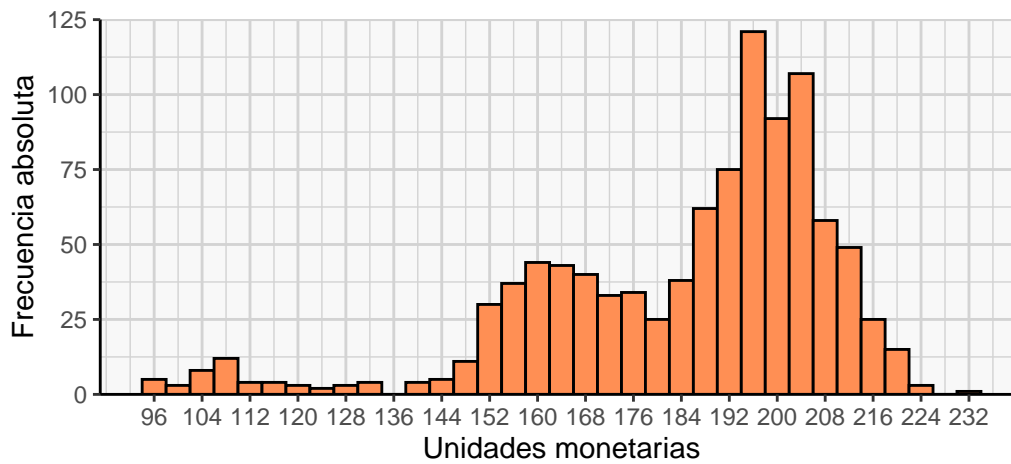


En este año se usa casi exclusivamente la maquina 2 y apenas se utilizan las otras dos. Esto pasa porque las probabilidades de exitos iniciales observadas son muy variables, por lo que la primera maquina en tener varios fracasos va a tener una probabilidad muy chica y se dejara de usar. Esto impacta 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 185.1 unidades monetarias en un año.

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

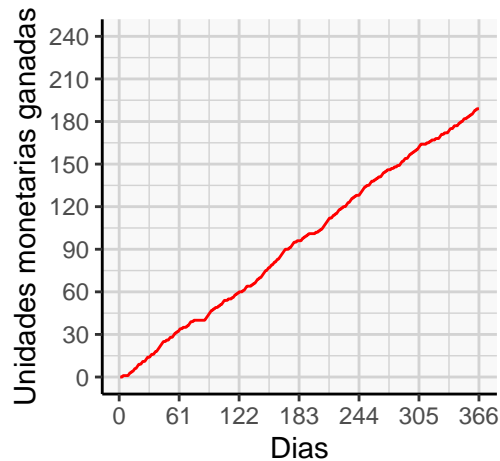
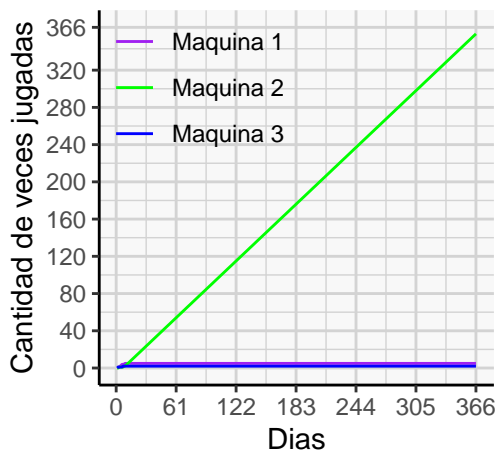
	Maquina 1	Maquina 2	Maquina 3
Media	23.37	238.18	104.45
Mediana	4.00	352.50	5.00

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.

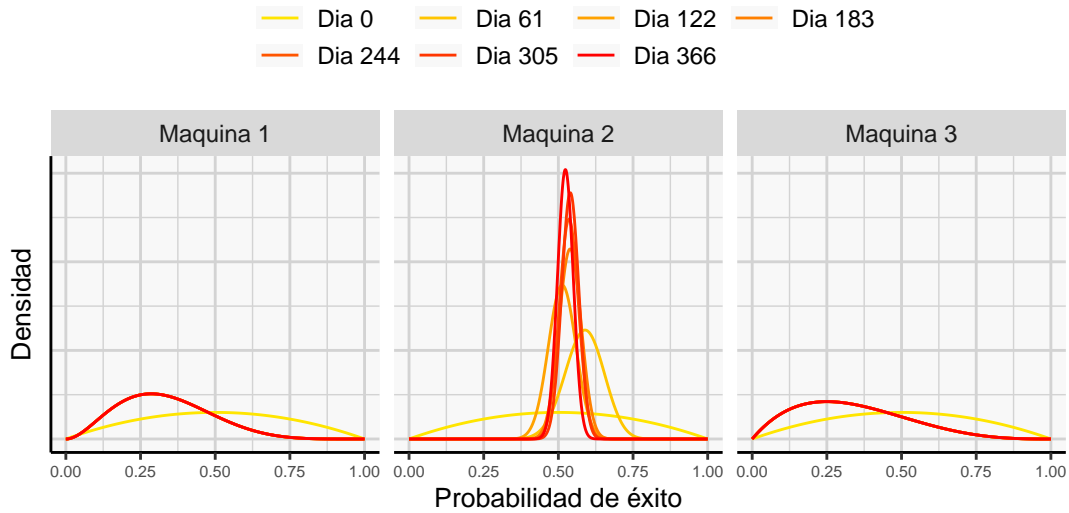
Estrategia 3: Greedy con probabilidad a posteriori

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. Si las maquinas tienen esperanzas iguales entonces se elige una de ellas aleatoriamente.

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

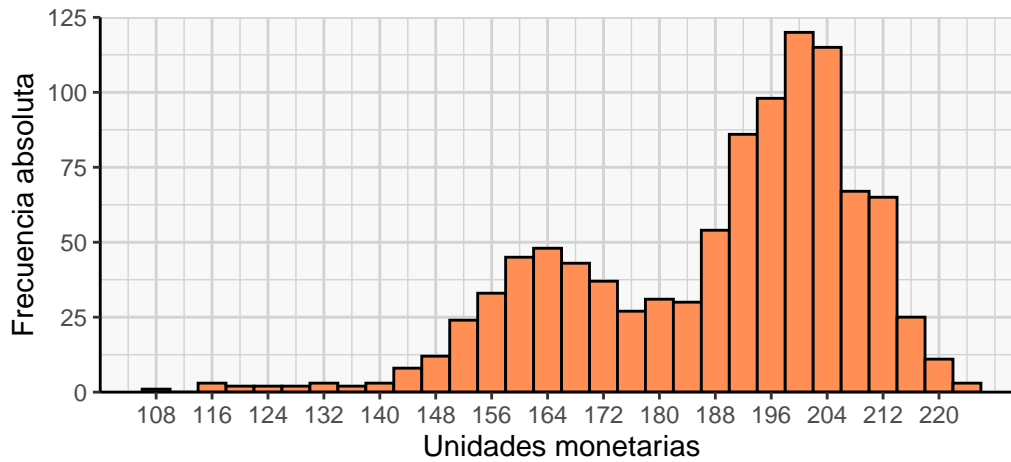


La seleccion de las maquinas en el año simulado es notablemente similar a la estrategia anterior. La unica diferencia es que las maquinas 1 y 3 se usaron un poco mas al principio del año, aunque sea dificil de ver.



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 las probabilidades de exito sera grande para una maquina y chica en las otras.

Las simulaciones de los 1000 años arrojaron los siguientes resultados:



	Maquina 1	Maquina 2	Maquina 3
Media	10.005	245.412	110.583
Mediana	2.000	355.000	5.000

En promedio, siguiendo esta estrategia, se espera ganar 188.53 unidades monetarias en un año.

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

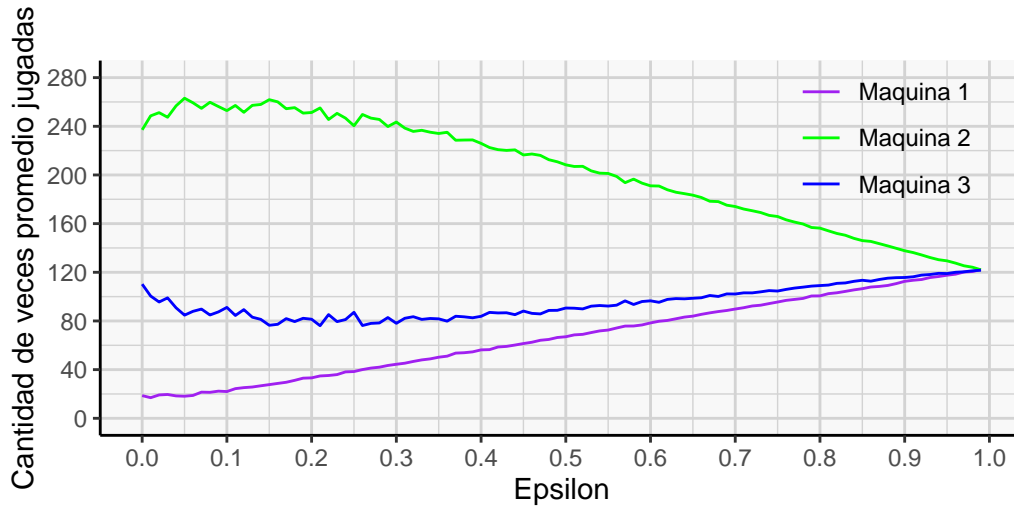
	Maquina 1	Maquina 2	Maquina 3
Media	10.01	245.41	110.58
Mediana	2.00	355.00	5.00

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 sus distribuciones a priori.

Estrategia 4: -greedy (con tasa observada)

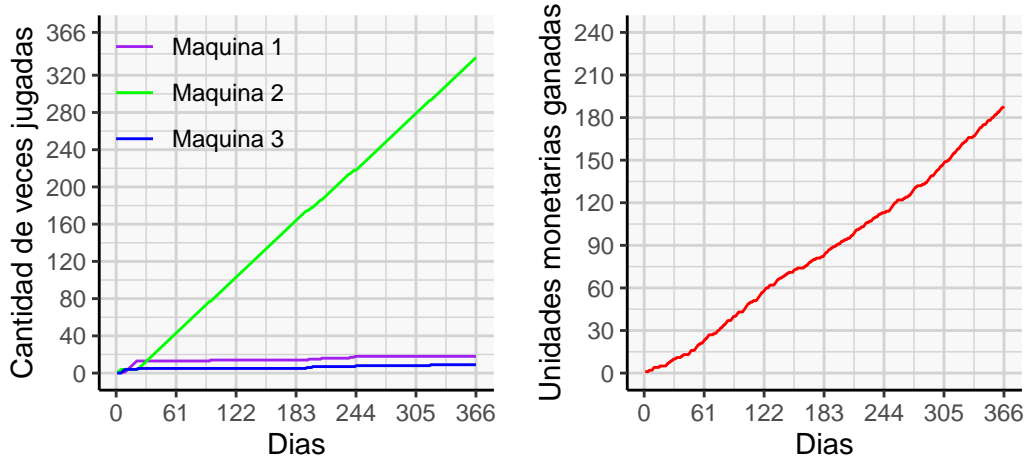
Esta estrategia es una combinacion de otras dos anteriores, la greedy con tasa observada y la completamente al azar. Se le elige utilizar la primera mencionada con una probabilidad de ϵ y

la segunda con una probabilidad de $1 - \epsilon$. Mientras mas cerca este de 1, se parecera mas al completamente al azar y mientras mas cerca este del 0 a greedy con tasa observada. Debido a que uno decide que valor asignarle a ϵ , se realiza un analisis previa para obtener el mejor ϵ .

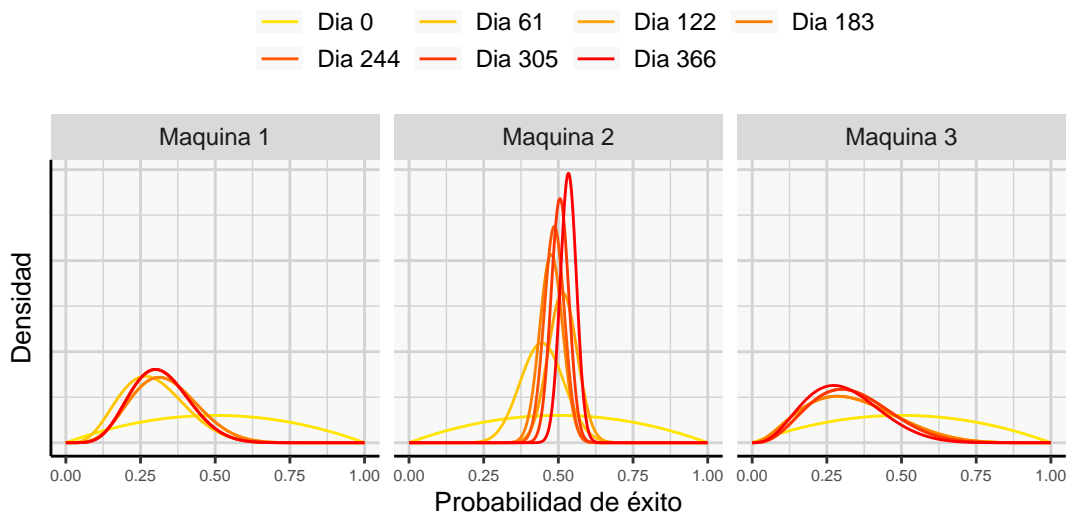


Queremos que las veces que se use la mejor maquina en promedio sea maxima, por lo que elegimos el ϵ que satisfaga esta condicion. Este valor sera 0.05.

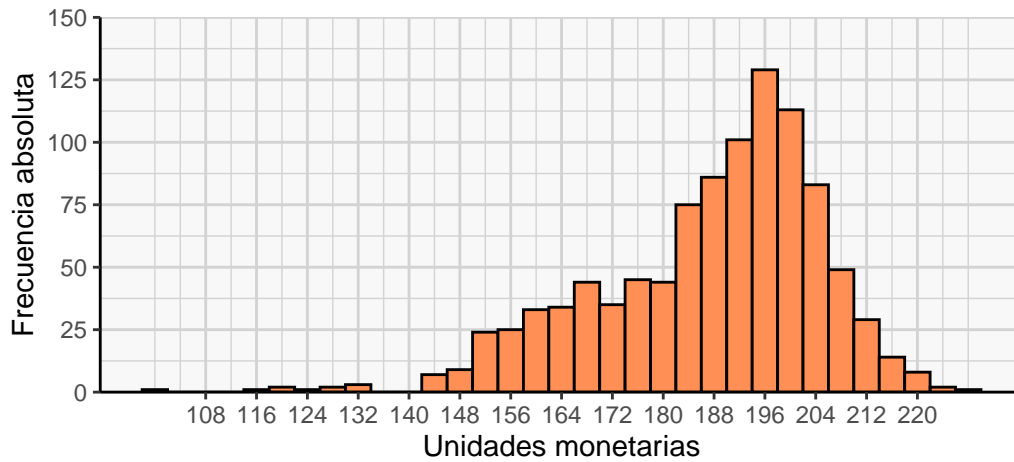
Se juega un año con esta estrategia y se obtienen los siguientes resultados:



Posteriors:



1000 años:



	Maquina 1	Maquina 2	Maquina 3
Media	18.98	256.66	90.36
Mediana	11.00	335.00	15.00

Estrategia 5: Softmax

En esta estrategia, para elegir la maquina a jugar primero se obtiene la probabilidad de exito observada π_i de cada maquina y se les aplica la siguiente formula:

$$\Pr(i) = \frac{e^{\pi_i/\tau}}{\sum_{j=1}^3 e^{\pi_j/\tau}}$$

Siendo τ un parametro de “temperatura” que controla el grado de exploracion. Luego se elige la maquina i con probabilidad $\Pr(i)$.

Se observa en la siguiente tabla las tasas de las probabilidades de exito reales de las maquinas para ver que les ocurre cuando τ varia.

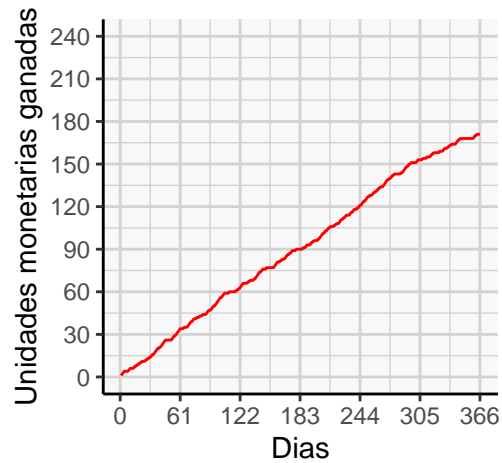
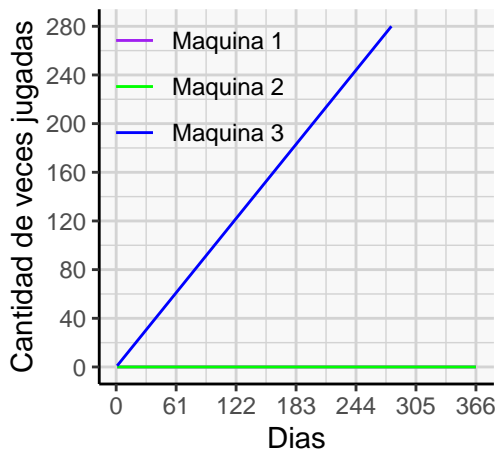
	Maquina 2 / Maquina 1	Maquina 2 / Maquina 3
Originales	1.83	1.22
Temperatura = 1/25	557.77	13.12

Temperatura = 1	1.28	1.11
Temperatura = 25	1.01	1.00

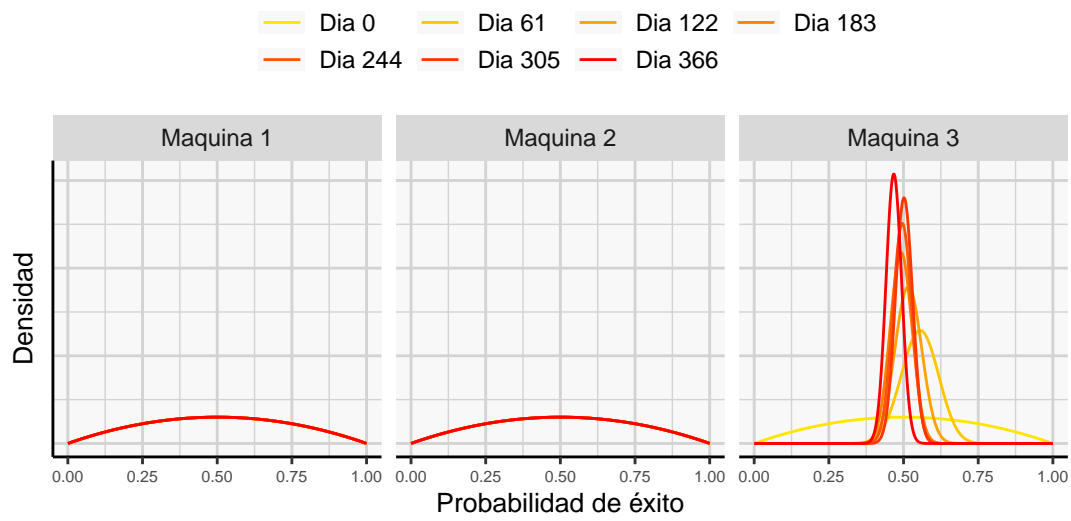
	Maquina 1	Maquina 2	Maquina 3
Media	10.005	245.412	110.583
Mediana	2.000	355.000	5.000

Cuando la temperatura crece, la diferencias entre las probabilidades de exito de cada maquina disminuyen mientras que las diferencias aumentan cuando la temperatura disminuye. Cuando la temperatura es alta, se exploran todas las maquinas, mientras que cuando es baja, se explota la que hasta el momento sea la que tenga la mayor probabilidad de exito.

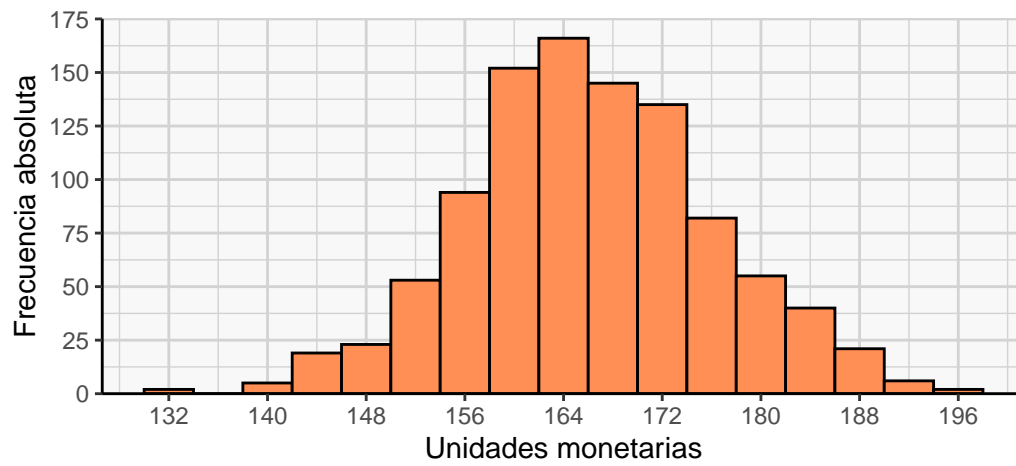
Se juega un año con esta estrategia y se obtienen los siguientes resultados:



Posteriors:



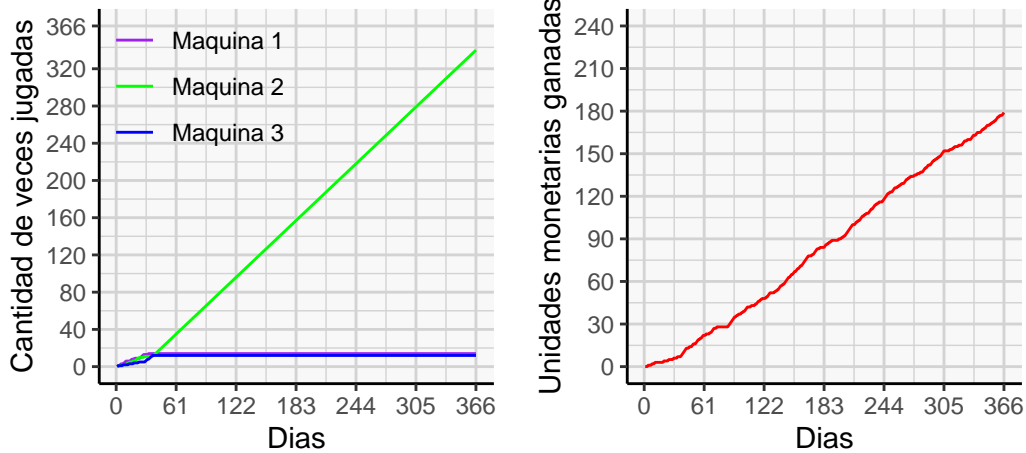
1000 años:



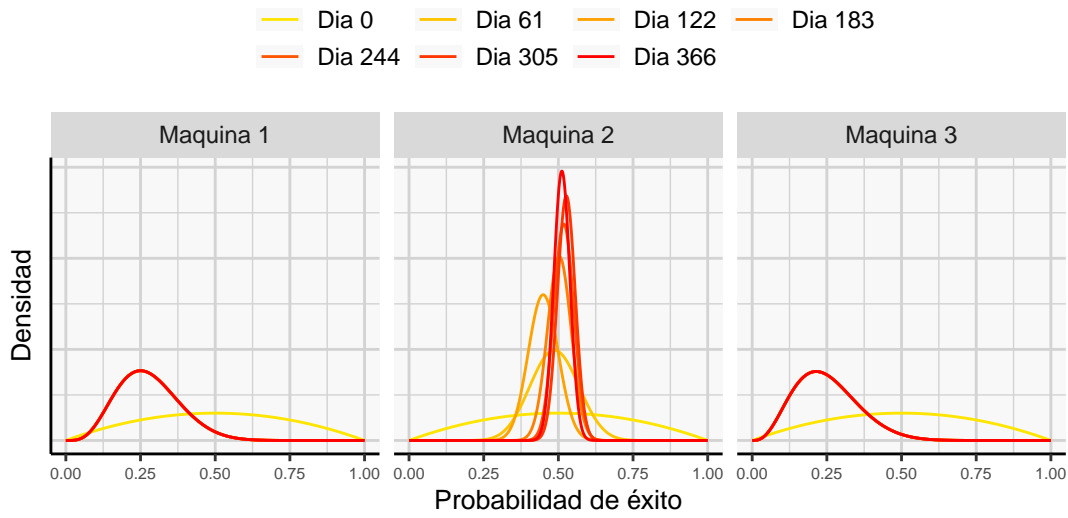
Estrategia 6: Upper-bound

Sera seleccionada la maquina que tenga el mayor extremo derecho de un intervalo de credibilidad. En este caso, el extremo derecho del intervalo usado es aquel que acumula el 97.5% de la densidad. Si los extremos derechos de los intervalos son iguales para algunas maquinas se elige una de ellas aleatoriamente.

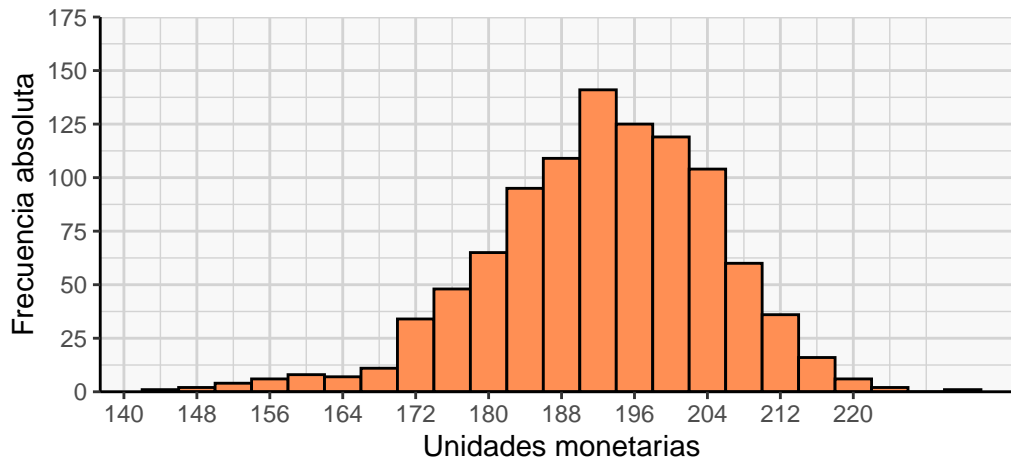
Se juega un año con esta estrategia y se obtienen los siguientes resultados:



Al comienzo del año se juegan todas las maquinas en una proporcion parecida, pero al cabo de un mes y un par de dias el metodo elige utilizar exclusivamente la maquina 2.



Como varias de las estrategias anteriores, generalmente esta tambien luego de un cierto numero de dias procede a jugar en una sola maquina. Pero para poder observar su diferencia con las demas tecnicas se realizaron simulaciones de 1000 años, las cuales arrojaron los siguientes resultados:



En promedio, siguiendo esta estrategia, se espera ganar 192.9 unidades monetarias en un año.

El 95% de las ganancias al final del año estuvieron entre 164 y 214.

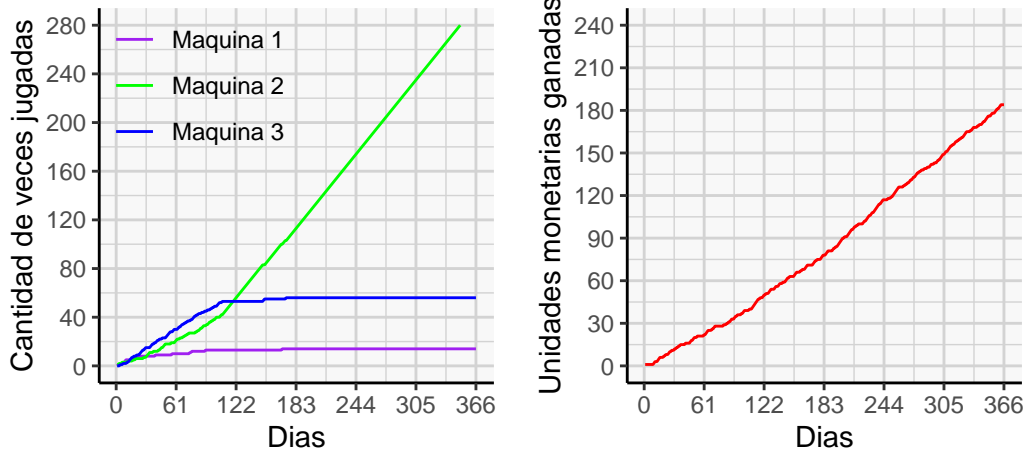
	Maquina 1	Maquina 2	Maquina 3
Media	13.95	295.61	56.44
Mediana	11.00	320.00	32.00

Se puede observar que el histograma de esta estrategia es mucho mas simetrico que el de las tecnicas anteriores que se le parecen. Esto se debe a que cada año la maquina con la que mas veces se suele quedar el metodo es la 2, es decir, la que mayor probabilidad de exito tiene.

Estrategia 7: Thompson sampling

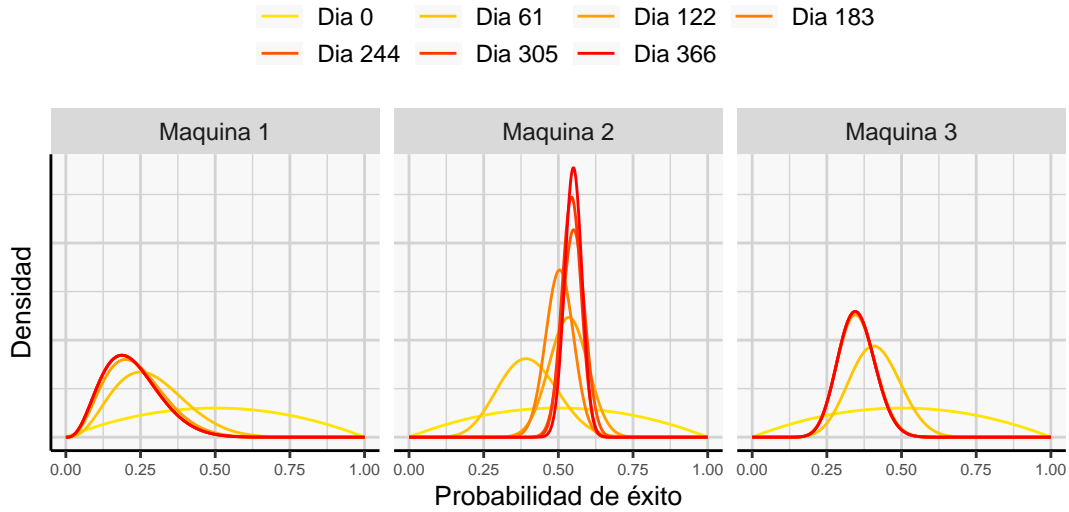
Para elegir que maquina se usa en el dia, se saca una muestra de la distribucion a porteriori de las probabilidades de éxito de cada una de ellas y se selecciona aquella con el maximo valor muestral.

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



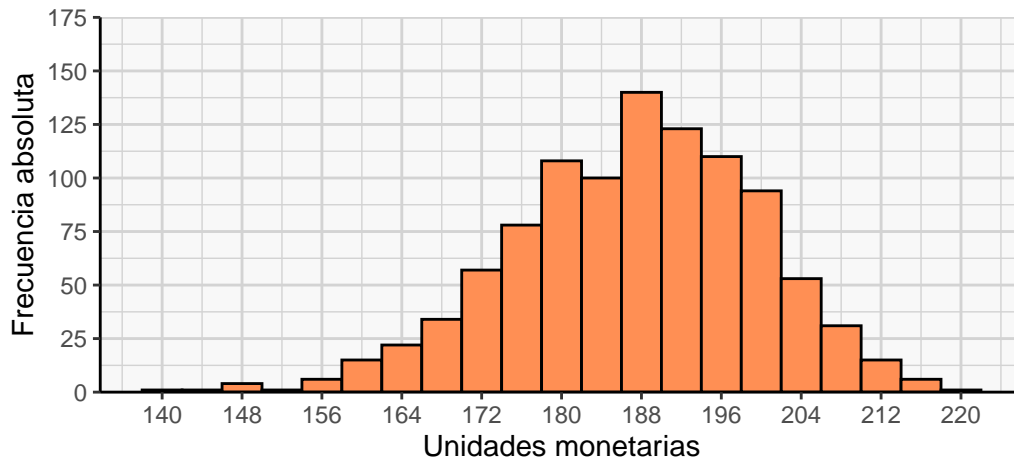
En este caso, en los primero 3 o 4 meses, la estrategia usa todas las maquinas. Luego de este periodo, utiliza mayormente la segunda dejando de utilizar las demas aproximadamente

a mitad de año. Se parece a las demas estrategias con la diferencia de usar muchas mas veces las maquinas con las que no se termina quedando el metodo.



Generalmente esta estragia, como la mayoria de las demas, luego de un cierto numero de dias procede a jugar con una sola maquina.

Para observar como varian las ganancias usando este metodo se simularon 1000 años, las cuales arrojaron los siguientes resultados:



En promedio, siguiendo esta estrategia, se espera ganar 192.9 unidades monetarias en un año.

El 95% de las ganancias al final del año estuvieron entre 164 y 214.

	Maquina 1	Maquina 2	Maquina 3
Media	22.99	269.43	73.58
Mediana	19.00	286.00	57.00

Esta estrategia, como la anterior, tiene un histograma bastante simetrico. Es mejor que los otros metodos parecidos pero es peor que upper bound ya que tiene un promedio de ganancias mas chico y ambos extremos de los intervalos son menores.

Hacer: e, temperatura, caratula, comparacion cuantas veces se elige cada maquina