

Introducción a procesos estocásticos, simulación y sampleo

Rodrigo Maranzana

Proceso estocástico

- Sucesión de variables aleatorias que evolucionan en función de un parámetro u otra variable.
- La evolución depende de una distribución de probabilidad.

Ejemplos:

Cotización del dólar MEP en argentina



Fuente: <https://www.rava.com/perfil/DOLAR%20MEP>

Falla de servicio en un servidor cloud AWS

Event	Status	Region/AZ	Start time	Last update time
Datapipeline operational issue	Open	us-east-1	March 1, 2017 at 6:58:56 AM UTC...	March 1, 2017 at 9:17:52 AM UTC...
Elasticbeanstalk operational issue	Closed	us-east-1	March 1, 2017 at 6:43:23 AM UTC...	March 1, 2017 at 9:17:21 AM UTC...
Workmail operational issue	Closed	us-east-1	March 1, 2017 at 6:39:21 AM UTC...	March 1, 2017 at 9:17:21 AM UTC...
ECR operational issue	Closed	us-east-1	March 1, 2017 at 6:58:49 AM UTC...	March 1, 2017 at 9:16:21 AM UTC...
Elastictranscoder operational issue	Closed	us-east-1	March 1, 2017 at 7:58:04 AM UTC...	March 1, 2017 at 9:15:51 AM UTC...
ACM operational issue	Closed	us-east-1	March 1, 2017 at 7:02:45 AM UTC...	March 1, 2017 at 9:15:50 AM UTC...
S3 operational issue	Closed	-	March 1, 2017 at 6:36:33 AM UTC...	March 1, 2017 at 9:12:49 AM UTC...
Mobilehub operational issue	Closed	us-east-1	March 1, 2017 at 6:56:50 AM UTC...	March 1, 2017 at 9:06:17 AM UTC...
Firehose operational issue	Closed	us-east-1	March 1, 2017 at 6:49:46 AM UTC...	March 1, 2017 at 9:06:17 AM UTC...
KMS operational issue	Closed	us-east-1	March 1, 2017 at 6:46:40 AM UTC...	March 1, 2017 at 9:04:17 AM UTC...
Redshift operational issue	Closed	us-east-1	March 1, 2017 at 6:44:27 AM UTC...	March 1, 2017 at 9:04:16 AM UTC...
Elasticfilesystem operational issue	Closed	us-east-1	March 1, 2017 at 6:40:12 AM UTC...	March 1, 2017 at 9:04:16 AM UTC...
Glacier operational issue	Closed	us-east-1	March 1, 2017 at 7:02:06 AM UTC...	March 1, 2017 at 9:02:16 AM UTC...
Cloudformation operational issue	Open	us-east-1	March 1, 2017 at 6:49:11 AM UTC...	March 1, 2017 at 9:02:16 AM UTC...
Elasticmapreduce operational issue	Open	us-east-1	March 1, 2017 at 6:49:47 AM UTC...	March 1, 2017 at 9:00:15 AM UTC...
Workspaces operational issue	Open	us-east-1	March 1, 2017 at 6:50:49 AM UTC...	March 1, 2017 at 8:58:15 AM UTC...
Storagegateway operational issue	Open	us-east-1	March 1, 2017 at 6:48:21 AM UTC...	March 1, 2017 at 8:58:15 AM UTC...

Fuente: https://www.theregister.com/2017/03/01/aws_s3_outage/

Clasificación según variable y parámetro

Discreta

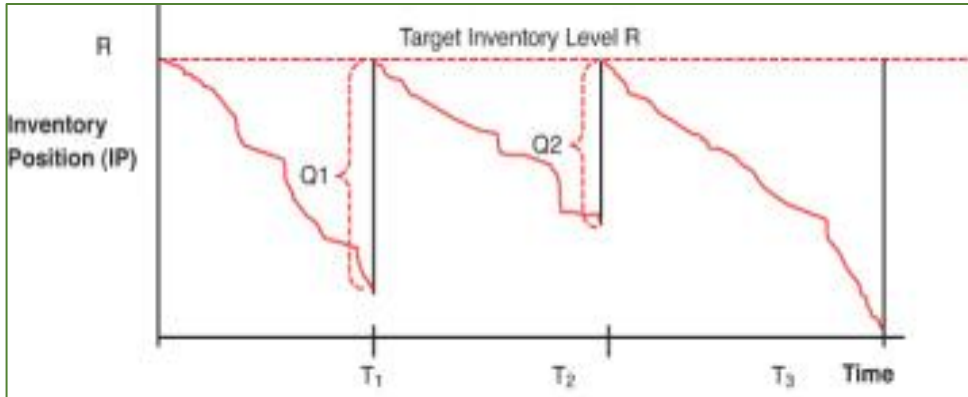
Variable

Continua

Evolución anual del rating crediticio de una institución

AAA	AAA	AAA	AAA
AA	AA	AA	AA
A	A	A	A
BBB	BBB	BBB	BBB
BB	BB	BB	BB
B	B	B	B
CCC	CCC	CCC	CCC
Default	Default	Default	Default

Sistema de control de inventarios de tiempo fijo



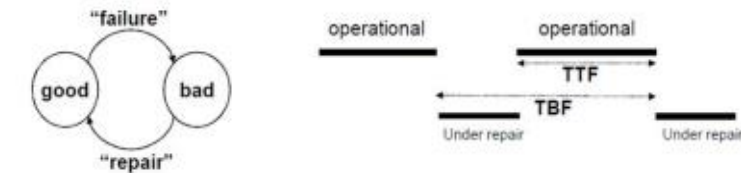
Fuente: <https://www.informit.com/articles/article.aspx?p=2167438&seqNum=7>

Estado de falla y reparación de una máquina

Failures with Repair



Time between failures: time to repair + time to next failure



Fuente: https://cdnc.itec.kit.edu/downloads/RC1_WS_2011_lecture5.pdf

Dinámica de stocks en la bolsa



Fuente: <https://finance.yahoo.com/quote/AAPL/>

Discreto

Parámetro

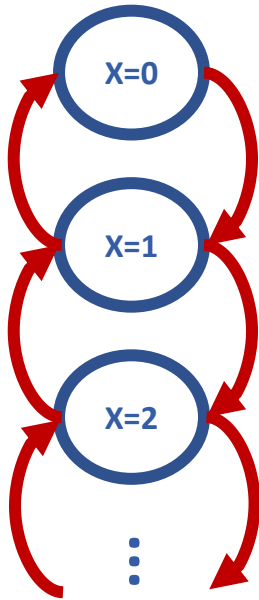
Continuo

Conceptos: Estado

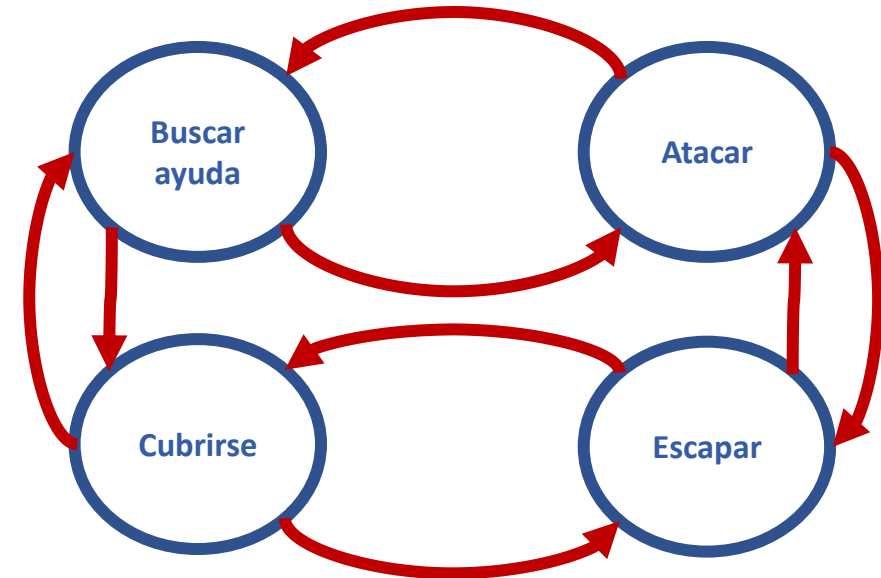
- Estado: configuración de variables y parámetros que describen completamente un sistema en un momento determinado.
- Espacio de estados: set de todos los estados posibles de un sistema.
- Similar al concepto de espacio muestral, pero no son lo mismo.
Recordemos: espacio muestral es el set de resultados de un experimento.

Ejemplos de estados

Cantidad de gente esperando en una fila:



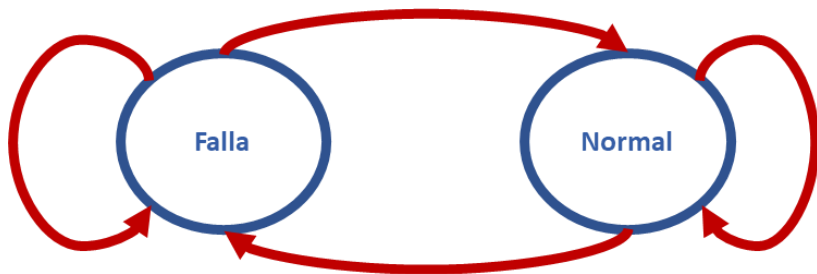
Comportamiento de una inteligencia artificial en un juego:



**Caso de autómata finito: set de estados finitos.*

Transición de estados

$P(\text{Normal} | \langle \text{historia de estados} \rangle)$

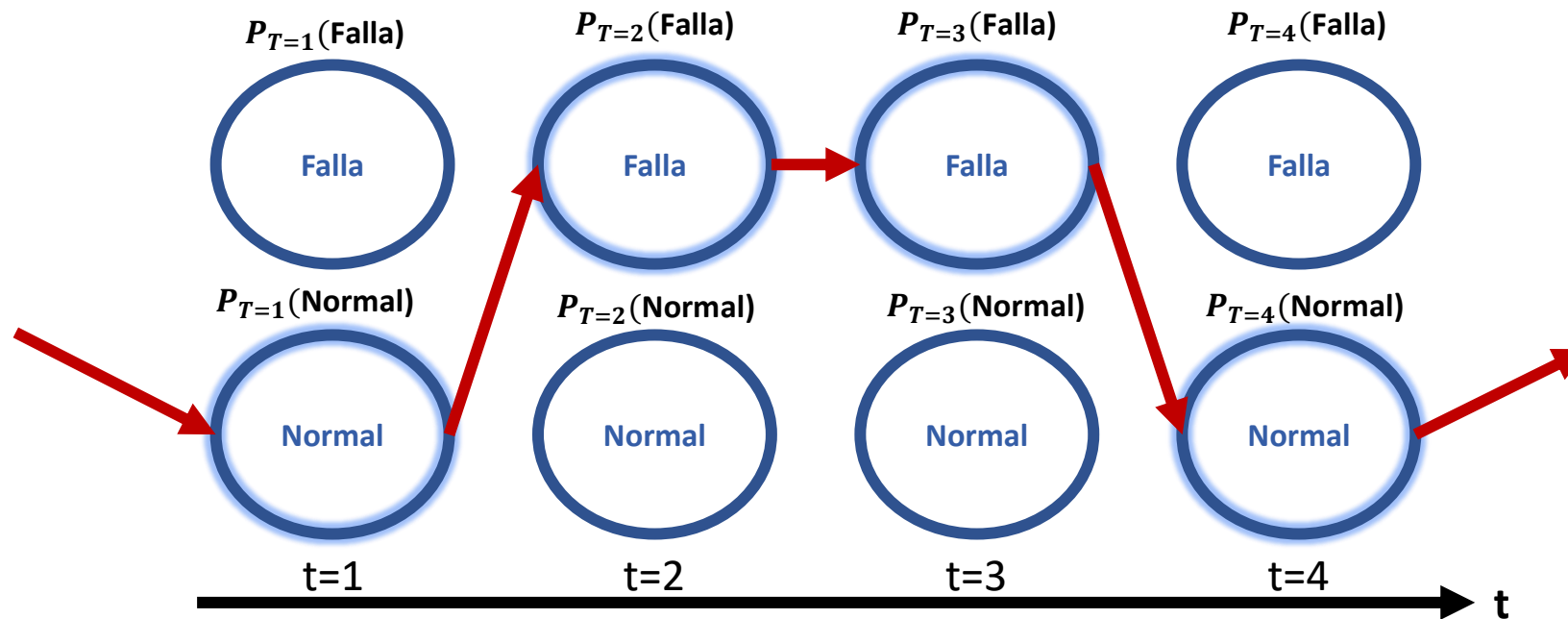


$P(\text{Falla} | \langle \text{historia de estados} \rangle)$

(Ej: Modelo de falla con parámetro discreto)

$P(\text{Estado} | \langle \text{historia de estados} \rangle)$: Probabilidad de transición, de cambiar a un estado, dada la historia de estados recorrida.

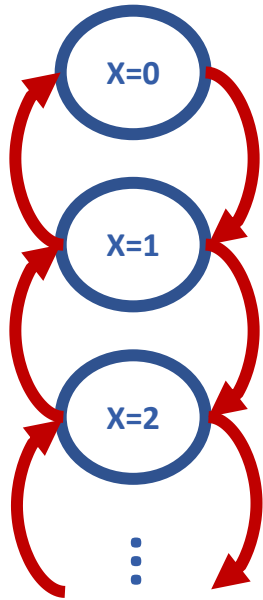
$P_{T=t}(\text{Estado})$: probabilidad de estar en un estado en un tiempo t



Clasificación según memoria

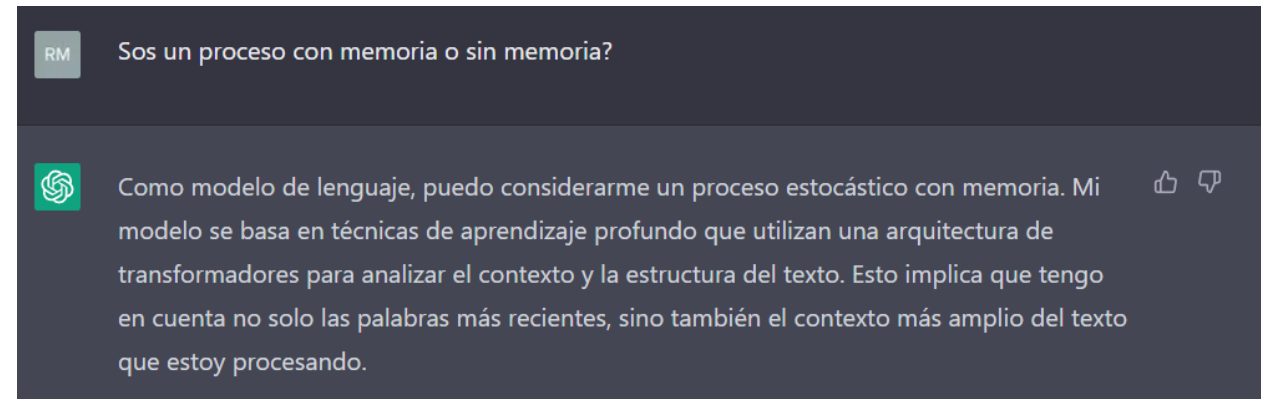
Procesos sin memoria: el resultado futuro del proceso solamente depende del estado presente.

Ej: cantidad de personas en una fila en una ventana de tiempo estable.



Procesos con memoria: el resultado futuro del proceso depende de uno o más estados de la secuencia pasada.

Ej: Large Language Models como ChatGPT



Fuente: <https://openai.com/blog/chatgpt>

Simulación

Evolución en el tiempo de un modelo que intenta replicar resultados de la realidad.

El objetivo es tomar una decisión en base a predicciones sobre el comportamiento de un sistema.

Simulación: conceptos

- Modelo: conjunto de características y comportamientos de un sistema.
- Cronómetro: sistema que permite la generación de pasos en el tiempo de simulación.
- Camino: Historia completa de una línea de tiempo de simulación.
- Paso: Movimiento unitario en la dirección del tiempo de simulación. En uno o varios caminos.
- Método de sampleo de V.A.: cálculo de un valor de la variable aleatoria siguiendo su distribución asociada.
- Generador de números aleatorios.
- Esperanza (de un evento en un tiempo dado): cálculo del valor medio transversal a todos los caminos en un punto dado de tiempo de simulación.

Generador de números aleatorios

Es un soft o hardware que produce una secuencia de números sin patrón identificable. No se debe poder predecir la secuencia.

Existen dos tipos:

- Verdaderos generadores: recogen datos de fenómenos físicos ruidosos que no se pueden predecir.
- Pseudo aleatorios: algoritmos que imitan la aleatoriedad. Son los más utilizados en la práctica.

Herramientas útiles: histograma

- Histograma: gráfico de barras de una variable, en donde cada una representa la frecuencia de un valor representado.
 - Los valores del eje x se denominan “bins”, y son agrupadores de eventos del resultado de la variable.
 - Los valores del eje y son las frecuencias asociadas a los valores en cada bin.
 - En python podemos usarlo mediante la librería Matplotlib*

[*] https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.hist.html

Ejemplo de histograma

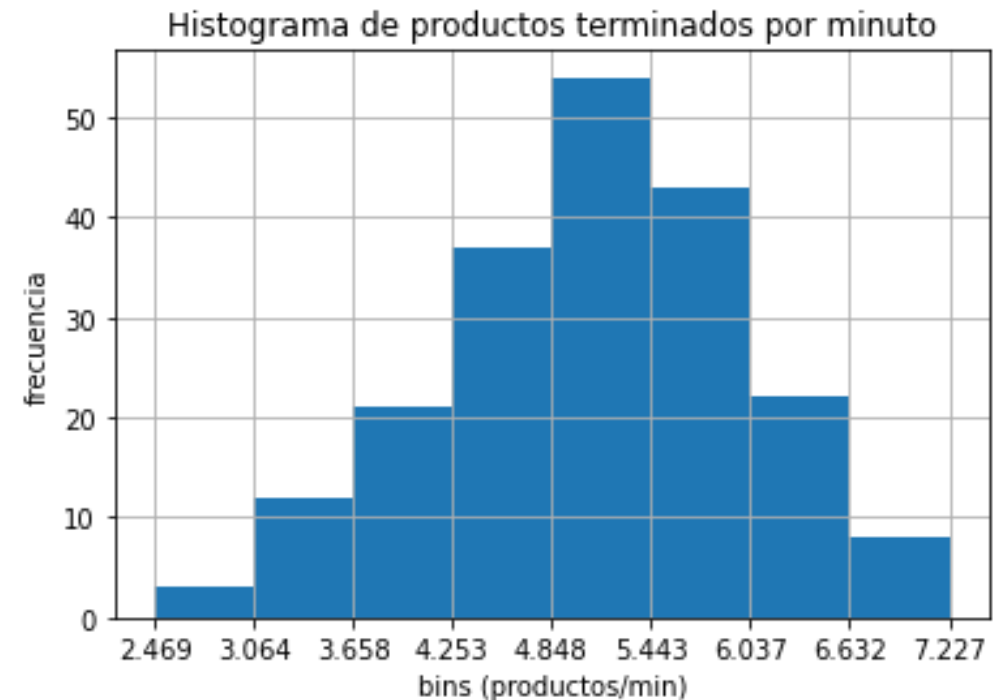
La última máquina de punta de línea procesa productos con una distribución que se considera normal según documentación.

Se recogen 200 datos de tiempo de despacho:

[3.72363418, 3.45195271, 5.00555358, 5.17057139, 4.83368523, 5.42215495, 4.02057953, 5.46945605, 5.24667385, 3.85746201, 5.491986 , 4.79417035, 5.09939141, 5.76297036, 6.59696642, 4.91440743, 6.637125 , 5.81895409, 4.50287724, 3.26092096, 4.30146374, 3.48649239, 4.43801592, 4.80342602, ...]

Usando un histograma de 8 bins:

- ¿Podríamos tener una primera aproximación al supuesto de la distribución normal?
- ¿Cuál podría ser la media?



Distribución empírica vs target

Cuando existe convergencia, al recoger mayor cantidad de datos de un proceso estocástico, nos acercamos a la distribución teórica.

Veamos el ejemplo anterior con 20, 100, 100.000 y datos:



Herramientas útiles: Histograma con KDE

Es un histograma solapado con una inferencia de la distribución de los datos, mediante un método llamado Kernel Density Estimation (KDE).

En python podemos usarlo mediante la librería Seaborn*



[*] <https://seaborn.pydata.org/generated/seaborn.histplot.html#seaborn.histplot>

Ejemplo inicial de simulación

En un juego de rol genérico tenemos dos dados balanceados de 6 caras cada uno.

Cada vez que tiramos los dados, **sumamos el resultado** y tenemos las siguientes reglas:

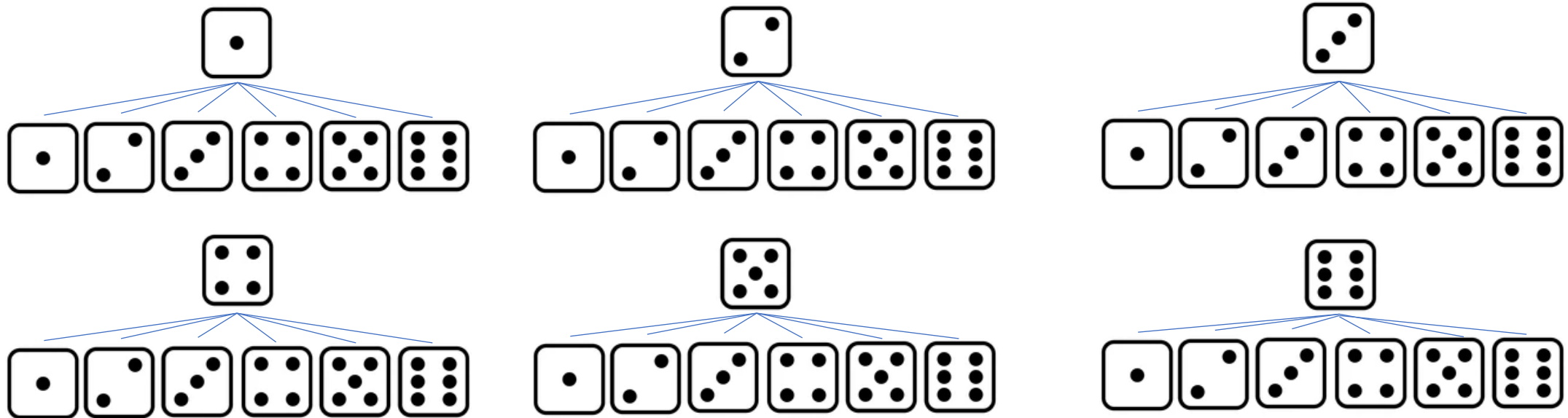


Fuente: <https://company.wizards.com/en>

1. Si la suma es **11 o 12**, golpeamos un **crítico de 150** veces el valor de la suma de los dados.
2. Si la suma es **un número par**, excepto punto 1, nuestro **golpe es bloqueado**.
3. Si la suma es **un número impar**, excepto punto 1, nuestro **golpe es normal, por 100** veces el valor de la suma de los dados.

Conceptos

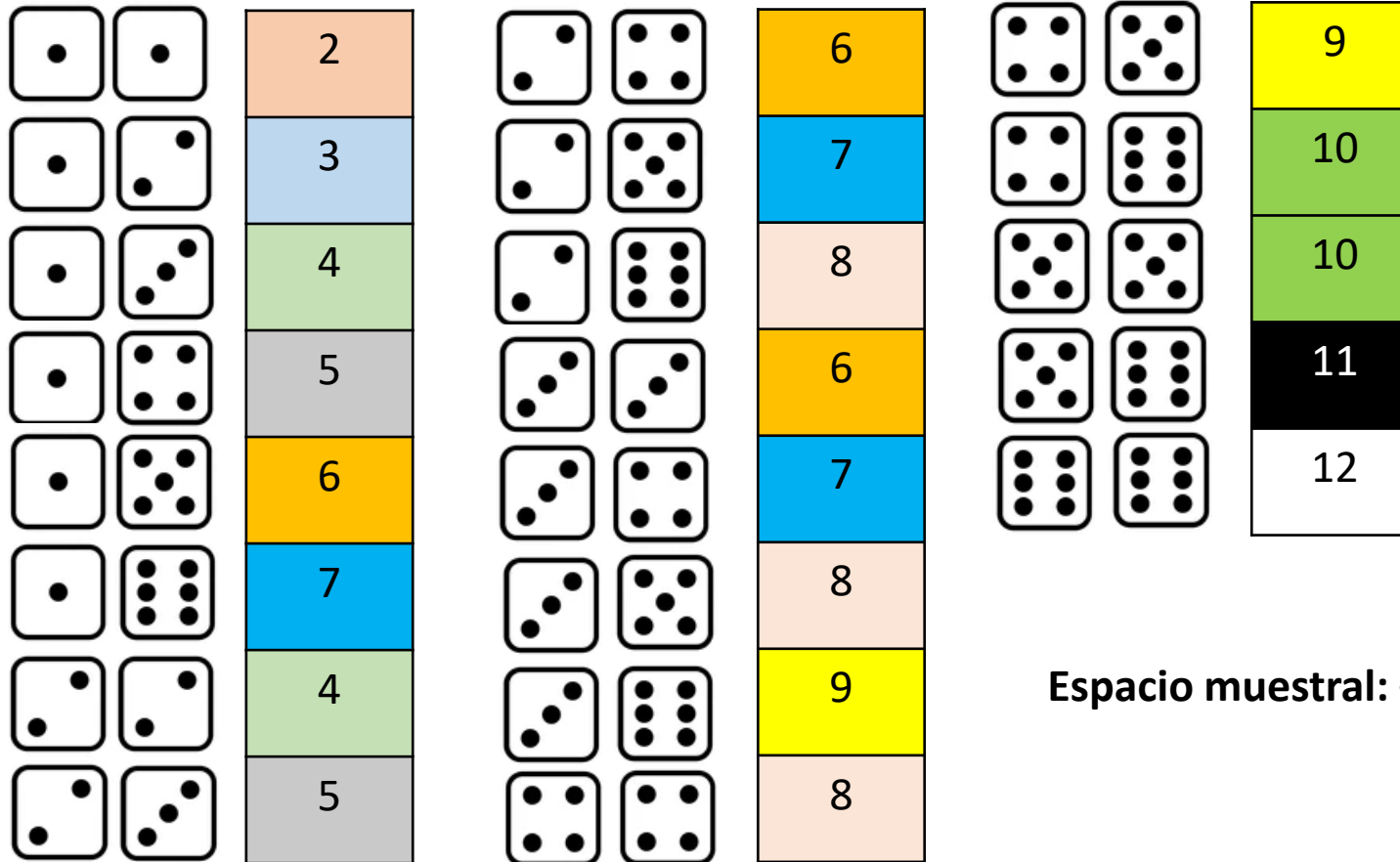
- Espacio muestral de los resultados de los dados, sin sumar.



$\{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), (3,1) \dots \}$

Conceptos

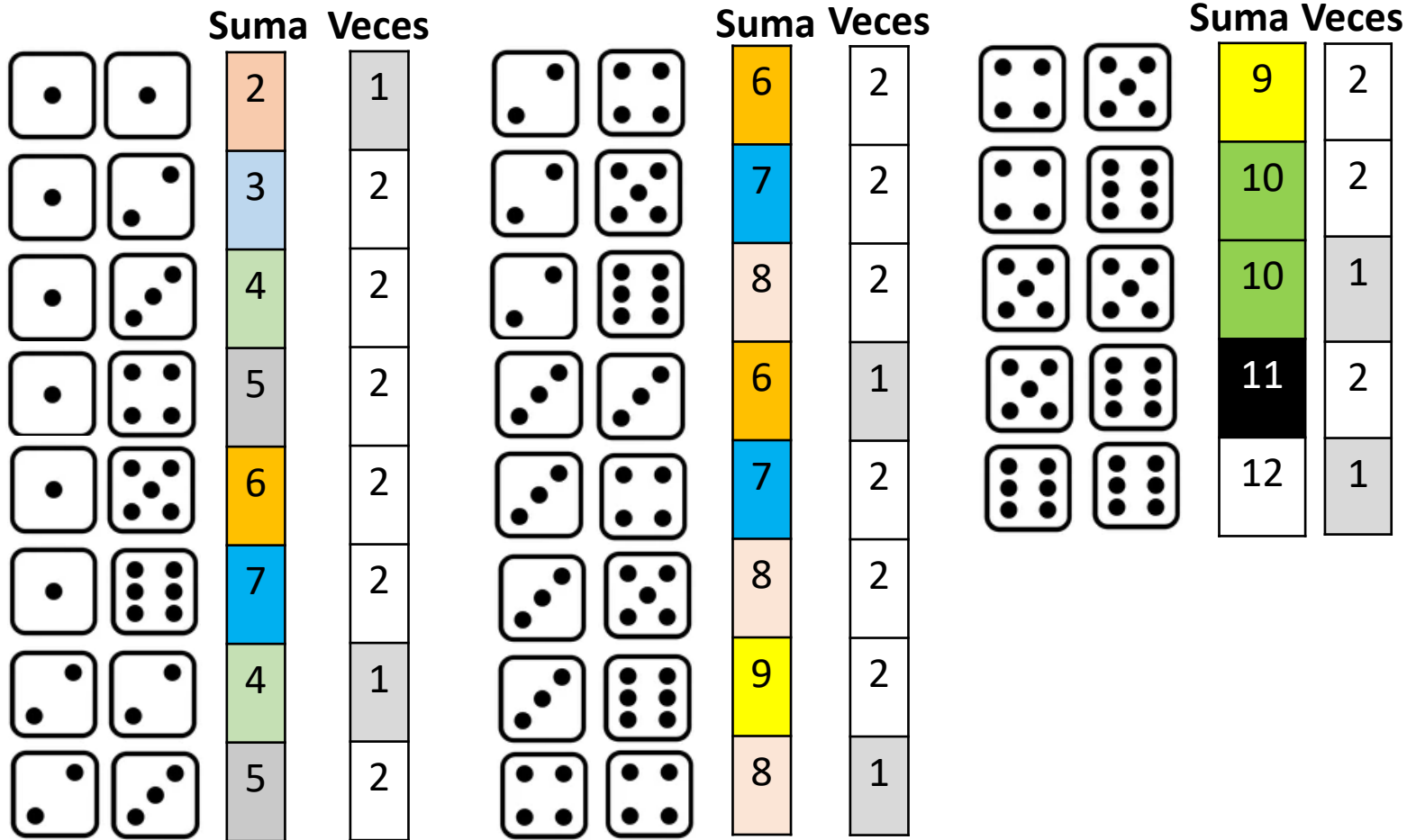
- Espacio muestral de la suma de los dados.



Espacio muestral: {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}

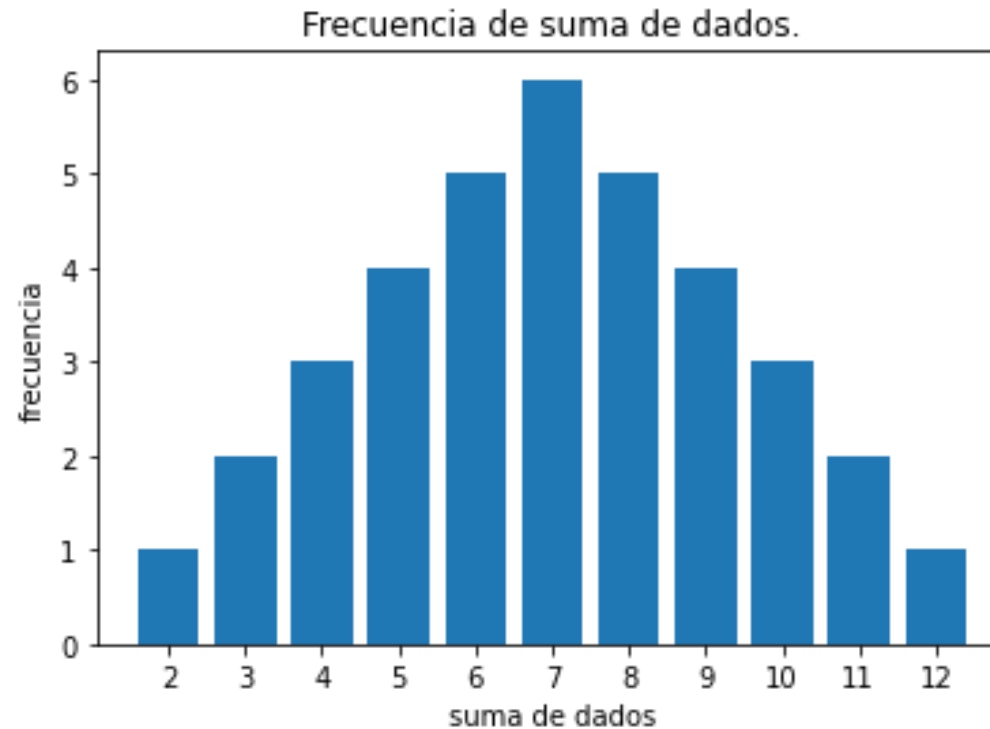
Conceptos

- Frecuencia de aparición de cada posibilidad.



Suma	Frecuencia
2	1
3	2
4	3
5	4
6	5
7	6
8	5
9	4
10	3
11	2
12	1

Histograma según resultado analítico



Simulación de suma: primer método

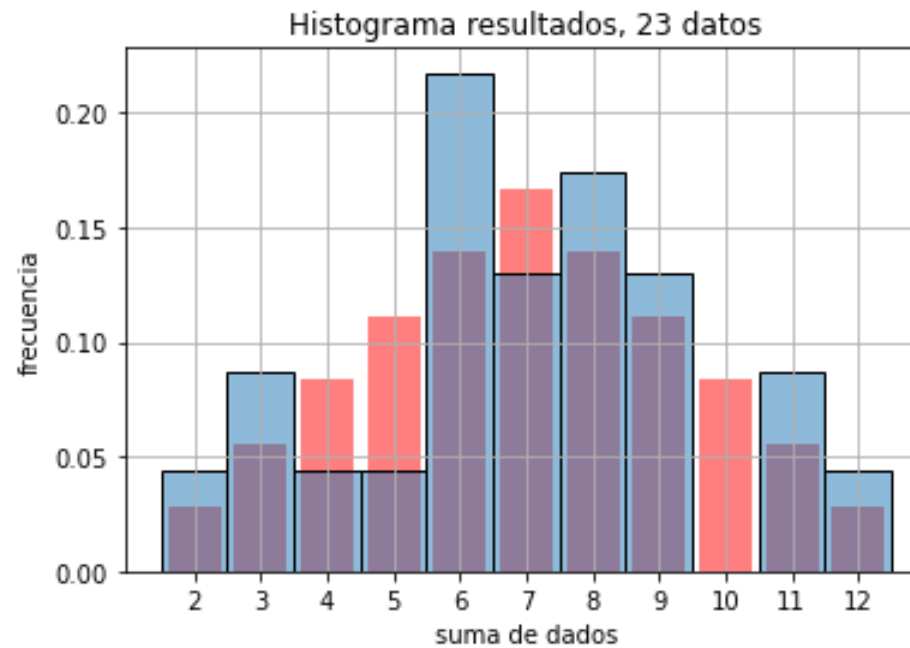
- Una primera aproximación es simular cada dado por separado, y sumar su resultado.
- La simulación implica obtener el resultado de una **Distribución Uniforme**. Las caras son equiprobables.
- En python lo podemos hacer mediante Numpy Random Integer*, generando números aleatorios enteros entre 1 y 6.

[*] <https://numpy.org/doc/stable/reference/random/generated/numpy.random.randint.html>

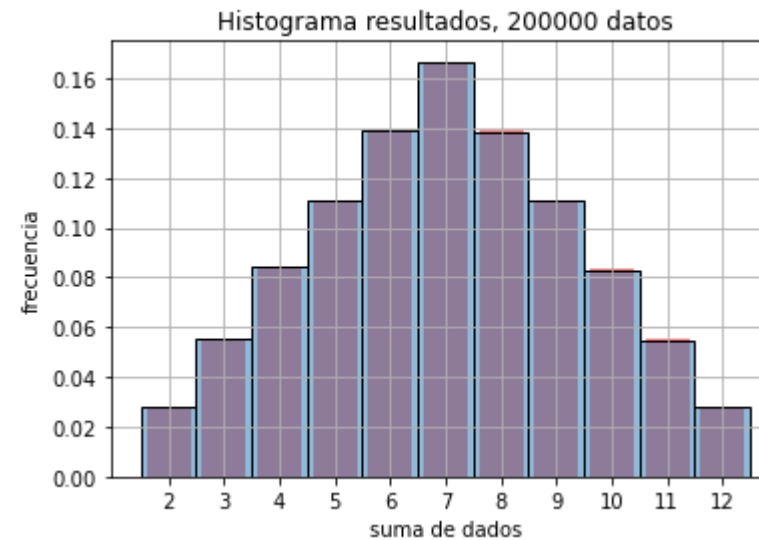
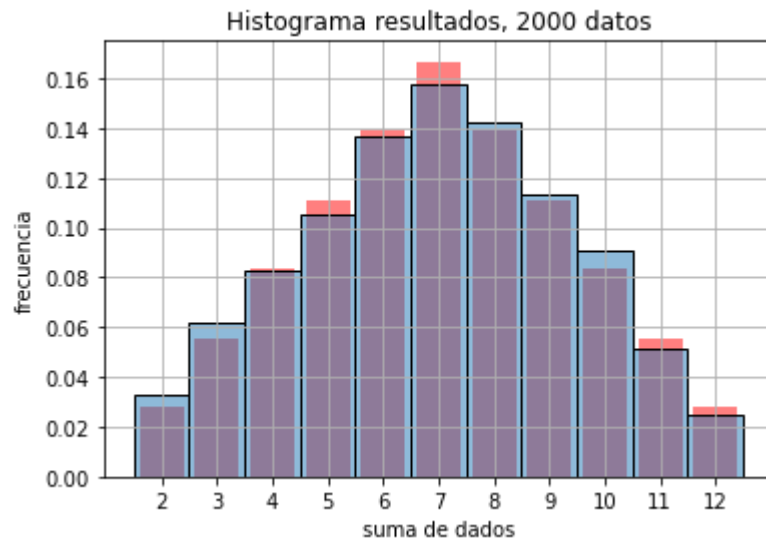
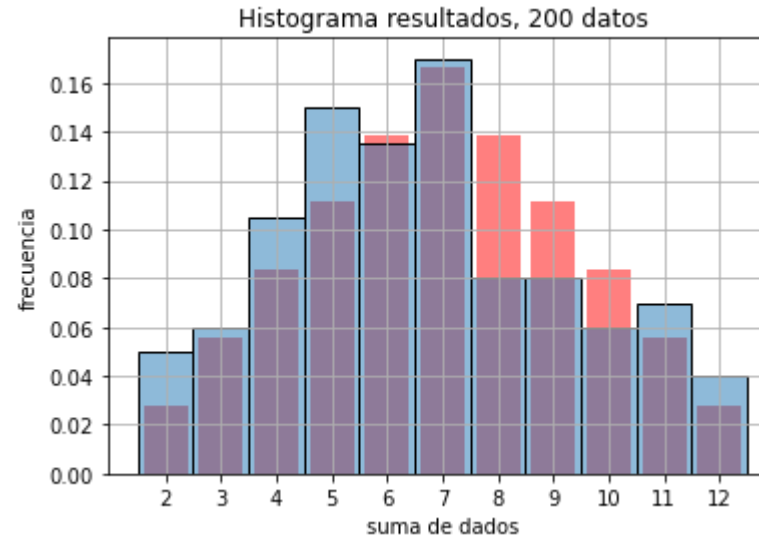
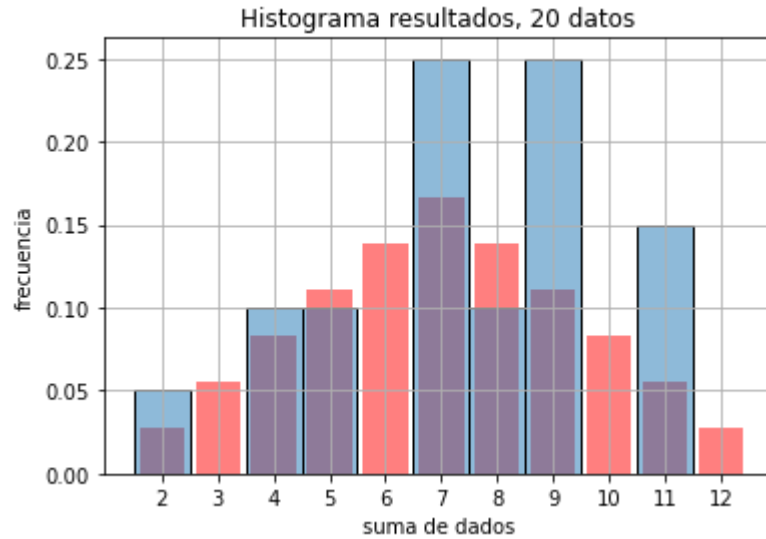
Simulación de suma: primer método

Simulamos 23 jugadas, comparamos el **resultado analítico** del histograma con el **empírico**.

	Dado1	Dado 2	Suma
0	3	2	5
1	2	5	7
2	5	5	10
3	4	1	5
4	4	3	7
5	6	4	10
6	6	1	7
7	6	1	7
8	4	6	10
9	5	3	8
10	3	1	4
11	5	6	11
12	5	5	10
13	5	2	7
14	6	2	8
15	2	4	6
16	2	6	8
17	1	1	2
18	6	3	9
19	1	4	5
20	1	6	7
21	3	1	4
22	1	2	3



Simulación de suma: primer método



Simulación de suma: segundo método

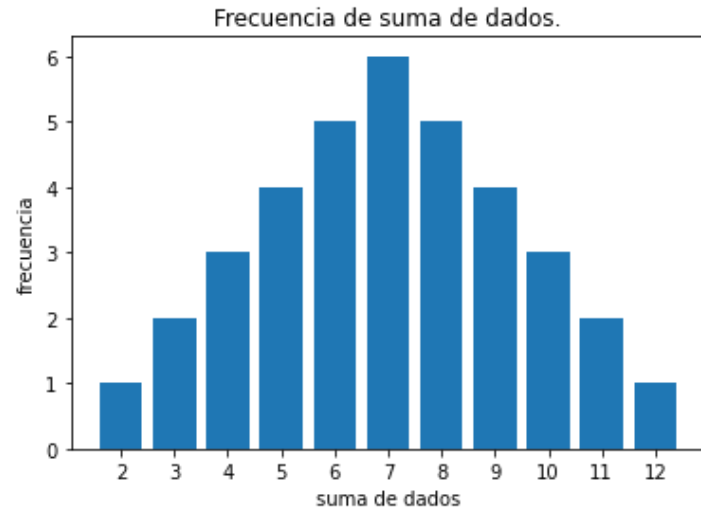
Sabiendo la distribución analítica, podemos simular directamente sobre ese histograma.

Conocemos las frecuencias de cada una de las sumas, las usamos para generar la probabilidad de obtener cada resultado.

Recordemos: hay 36 posibilidades en el espacio muestral.

Dividimos cada valor del histograma por 36, y tenemos la probabilidad.

Simulación de suma: segundo método



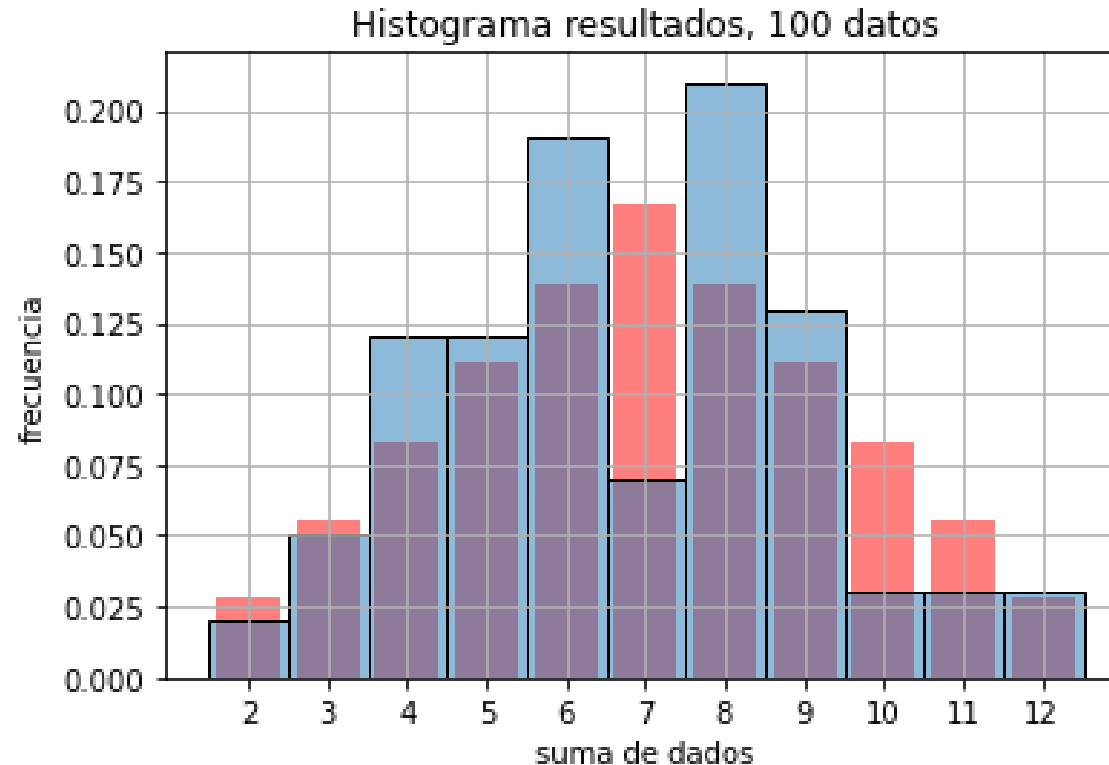
Para ejecutar este método en Python, usamos Numpy Random Choice* indicando en el parámetro “p” las probabilidades de cada valor.

Suma	Frecuencia	Probabilidad
2	1	0.03
3	2	0.06
4	3	0.08
5	4	0.11
6	5	0.14
7	6	0.17
8	5	0.14
9	4	0.11
10	3	0.08
11	2	0.06
12	1	0.03

[*] <https://numpy.org/doc/stable/reference/random/generated/numpy.random.choice.html>

Simulación de suma: segundo método

- Obtenemos los mismos resultados:



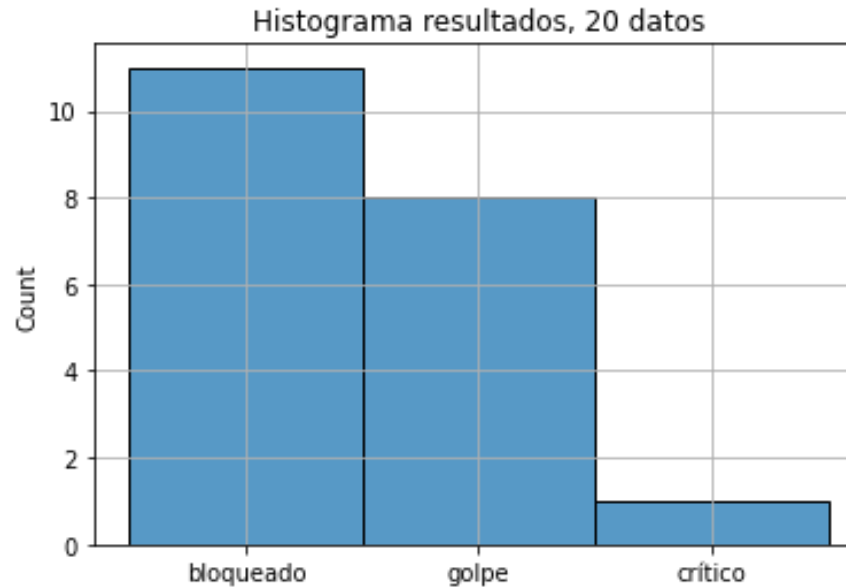
Simulación del juego de rol

- Volviendo al juego de rol teníamos las siguientes reglas:

1. Si la suma es **11 o 12**, golpeamos un **crítico de 150** veces el valor de la suma de los dados.
2. Si la suma es **un número par**, excepto punto 1, nuestro **golpe es bloqueado**.
3. Si la suma es **un número impar**, excepto punto 1, nuestro **golpe es normal, por 100** veces el valor de la suma de los dados.

Simulación de juego de rol

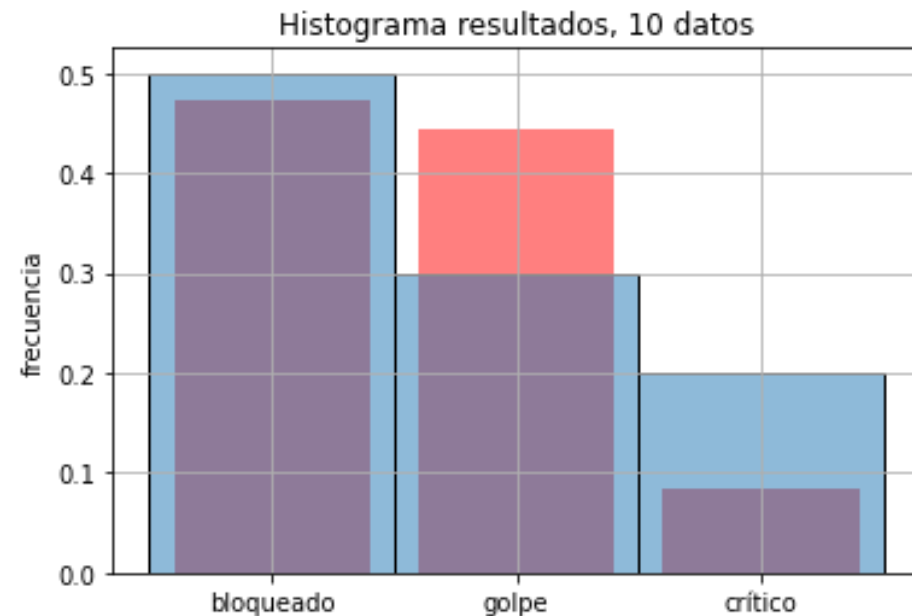
	Suma	Hits	Mensajes
0	4	0	bloqueado
1	7	700	golpe
2	7	700	golpe
3	6	0	bloqueado
4	4	0	bloqueado
5	7	700	golpe
6	3	300	golpe
7	9	900	golpe
8	6	0	bloqueado
9	6	0	bloqueado
10	11	1650	crítico
11	8	0	bloqueado
12	6	0	bloqueado
13	7	700	golpe



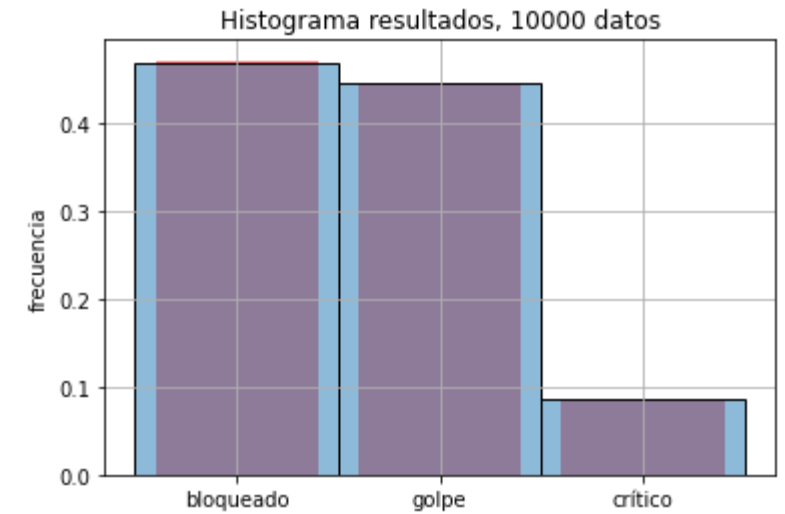
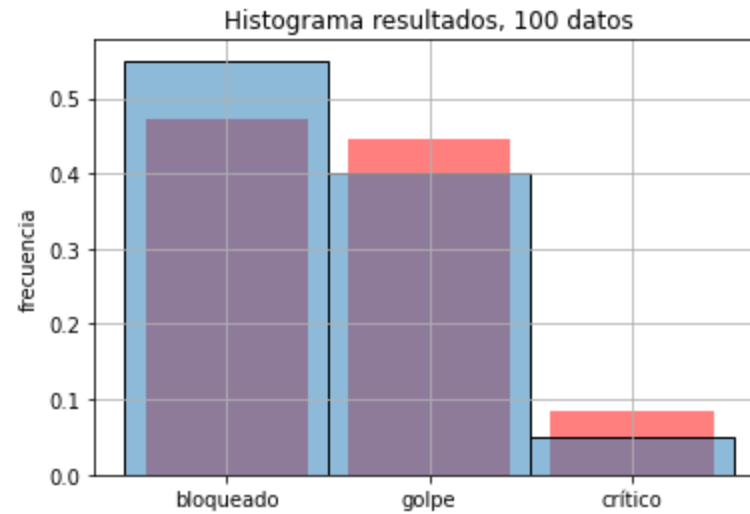
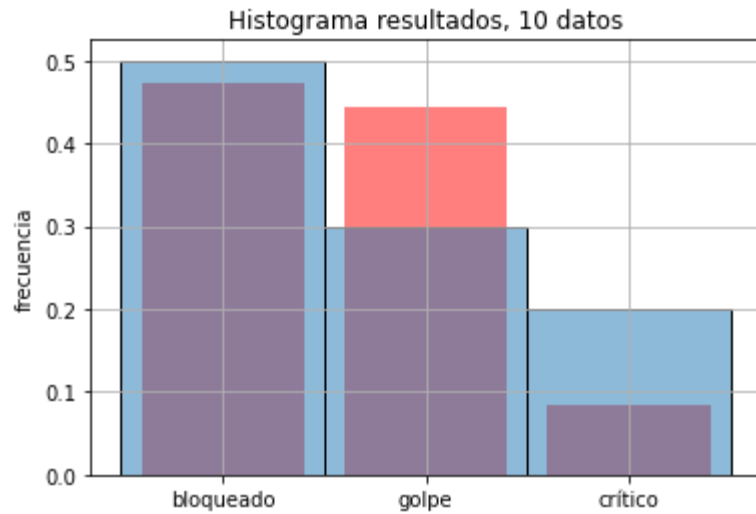
¿Podemos calcular la analítica?

Suma	Frecuencia	Decisión
2	1	bloqueo
3	2	golpe
4	3	bloqueo
5	4	golpe
6	5	bloqueo
7	6	golpe
8	5	bloqueo
9	4	golpe
10	3	bloqueo
11	2	crítico
12	1	crítico

Decisión	Frecuencia	Probabilidad
bloqueo	17	0.47
golpe	16	0.44
crítico	3	0.08



Resultados hasta convergencia con analítica



Simulación sobre los hits acumulados



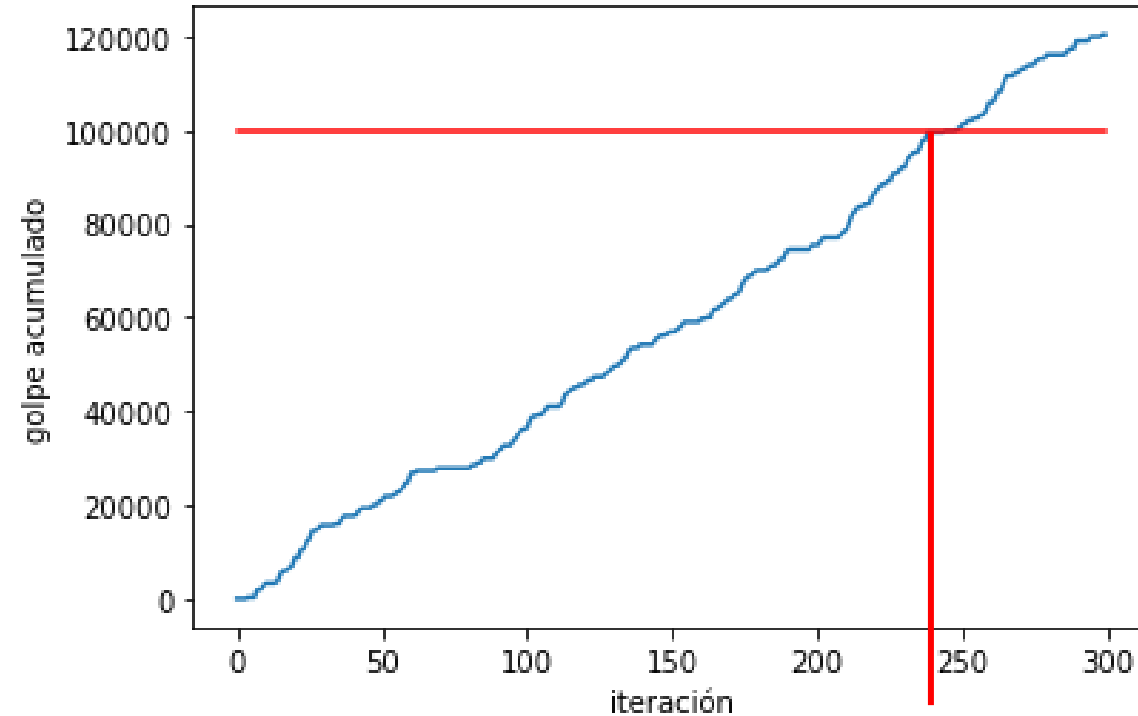
Dark Wizard

- 100.000 puntos de vida

¿Cuánto tardamos en matarlo?

Simulación sobre los hits acumulados

	Suma	Hits	Mensajes	Golpe acumulado
0	6	0	bloqueado	0
1	4	0	bloqueado	0
2	6	0	bloqueado	0
3	6	0	bloqueado	0
4	3	300	golpe	300
5	8	0	bloqueado	300
6	6	0	bloqueado	300
7	11	1650	crítico	1950
8	4	0	bloqueado	1950
9	9	900	golpe	2850
10	5	500	golpe	3350
11	10	0	bloqueado	3350
12	4	0	bloqueado	3350
13	4	0	bloqueado	3350
14	5	500	golpe	3850
15	11	1650	crítico	5500



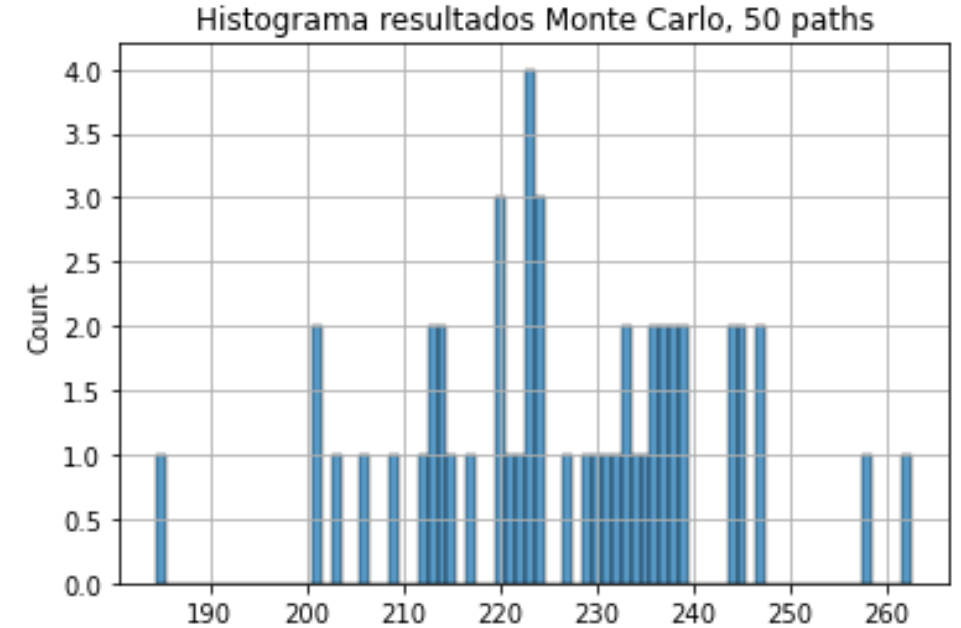
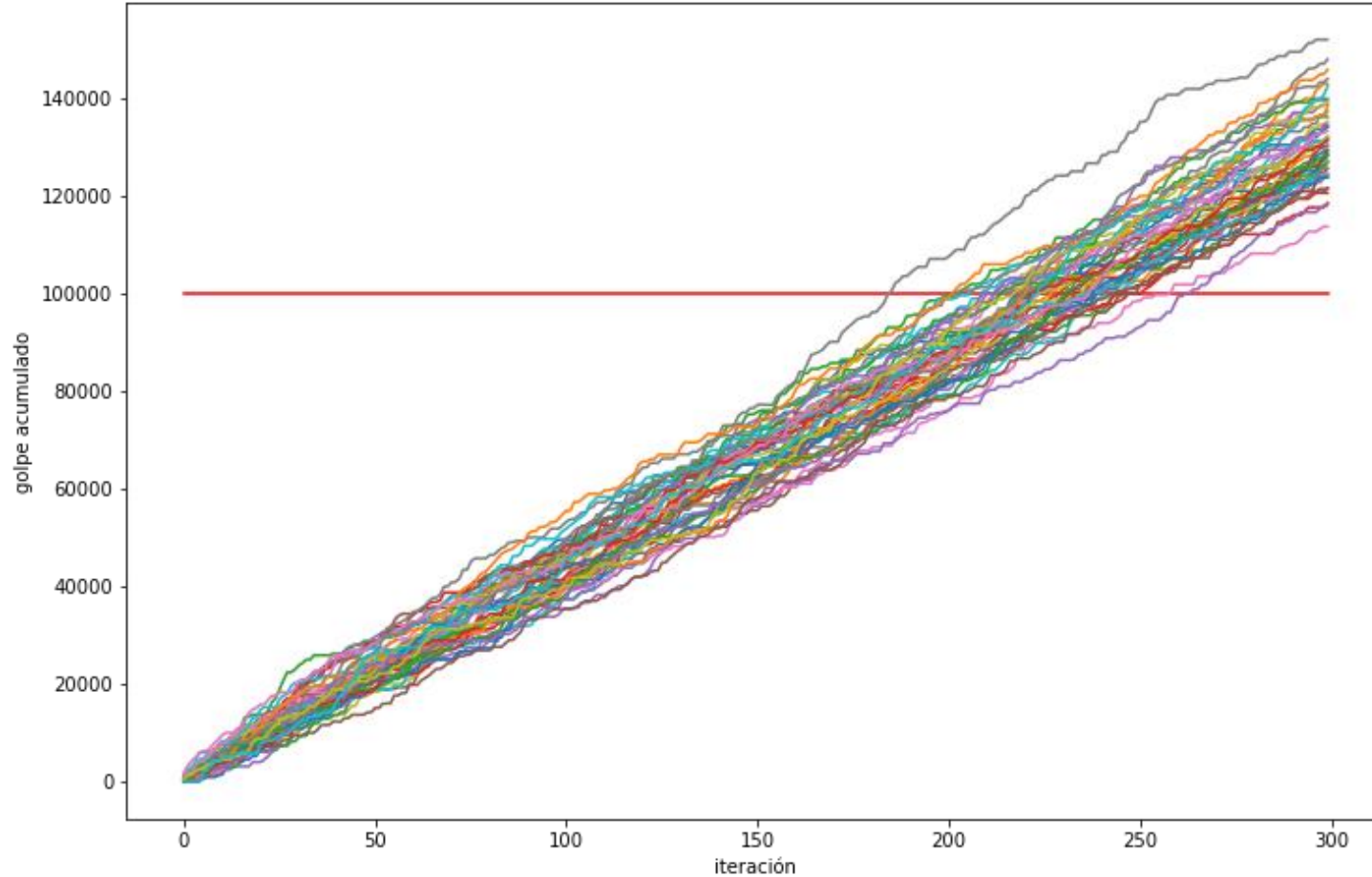
248 golpes (bloqueo o no)

Pero este valor es 1 camino de simulación. Cada vez que lo repetimos el resultado es diferente.

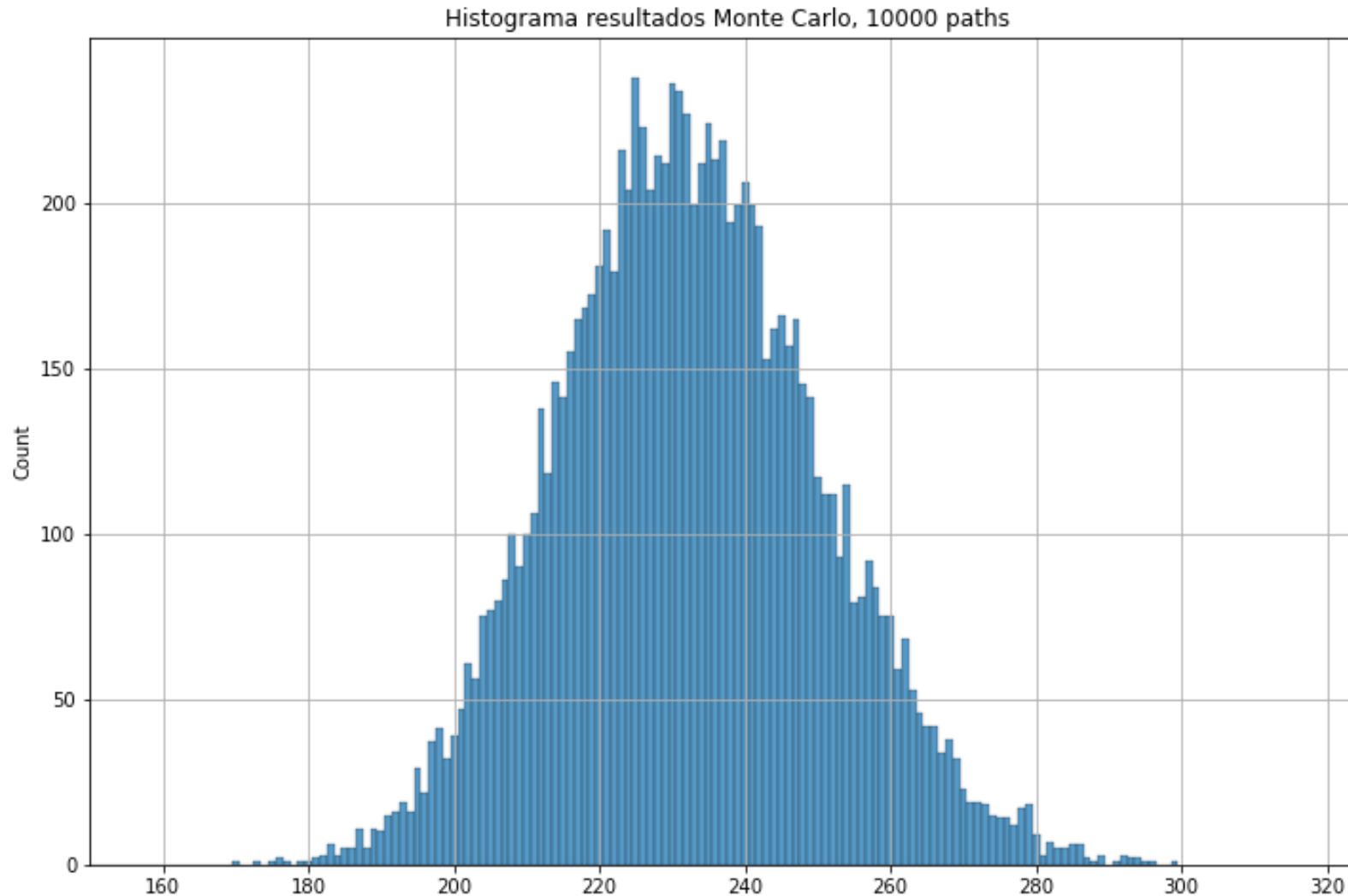
Método de Monte Carlo

- Ejecutamos N caminos, cada uno con una simulación de la variable aleatoria.
- Cada camino llegará al objetivo (matar a Dark Wizard), en un número de golpes totales distintos.

Ejemplo con 50 iteraciones



Aumento de paths para lograr convergencia



Vemos gráficamente que existe convergencia, y además es una distribución simétrica.

Calculamos la esperanza de la cantidad de golpes.

$$E[\text{golpes}] = 232.27$$