

75.26 - Simulación

Primer Cuatrimestre de 2021

**Trabajo Práctico 2**

| Fernando Sinisi | 99 139 | fsinisi@fi.uba.ar |
| --- | --- | --- |
| Tito Lasanta | 99 162 | tilasanta@fi.uba.ar |
| Santiago Almejún | 99 441 | salmejun@fi.uba.ar |
| Esteban Djeordjian | 100 701 | edjeordjian@fi.uba.ar |

**Índice**

[Ejercicio 1](#_uab5py3zjy6l) 2

[Ejercicio 2](#_s1pxmawhyr9u) 6

[Ejercicio 3](#_ermb8076dbe5) 10

[Ejercicio 4](#_tjwjx45at7qg) 18

[Anexo: código fuente](#_91o5m5pfh7sa) 22

# Ejercicio 1

**Proceso de Poisson**

Un proceso Poisson, con una tasa de arribos λ, es una secuencia de eventos en la cual cada uno de los intervalos de tiempo entre dos eventos consecutivos es una variable aleatoria idéntica e independiente con una distribución de probabilidades exponencial con parámetro λ . Sabemos que μ = 1/λ.

En este ejercicio, se pide simular un proceso de Poisson, con medias μ = 1, 2 y 4 segundos donde cada evento es una consulta a un servicio que usa una base de datos.

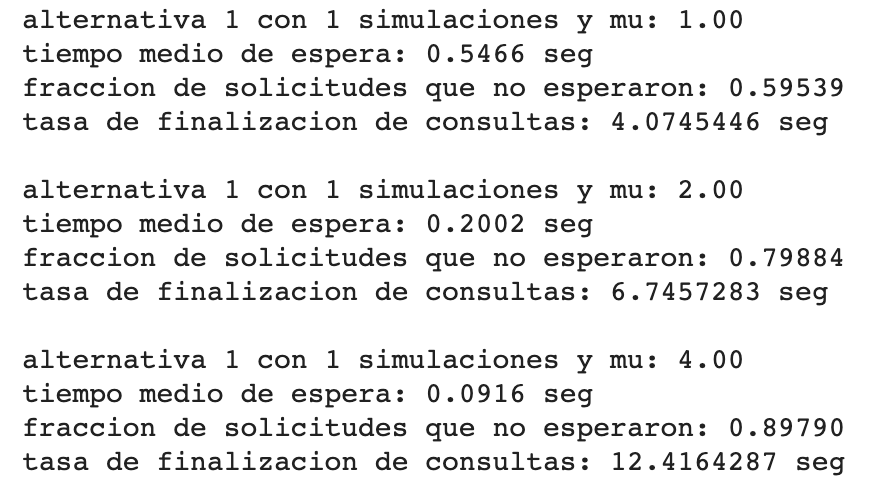
Como tenemos 2 alternativas de modelado para la base de datos y considerando que cada una de ellas producen un tiempo de demora para resolver una solicitud, tenemos que decidir cual alternativa es mejor evaluando el tiempo medio de espera, siendo este el tiempo promedio que una solicitud espera a ser procesada cuando la anterior aún no finalizó, la fracción/porcentaje de solicitudes que no esperaron ser procesadas y la tasa de finalización de consultas por segundo.

**Alternativa 1**

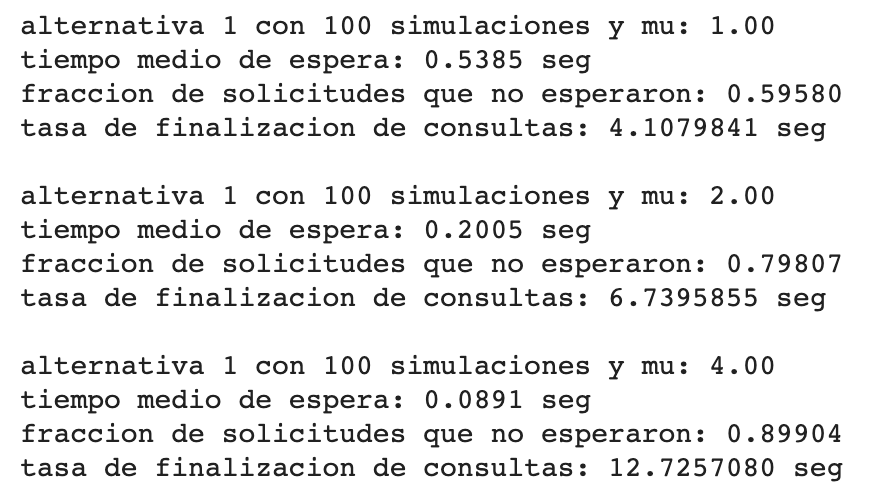
En este caso se utilizan 2 bases de datos distribuidas para resolver las consultas, la mismas se resuelven por la base “A” con probabilidad p = 0.6 y por la base “B” con probabilidad q = 1 - p = 0.4 . Cada base tiene su propia media de tiempo de demora, siendo μa = 0.7 y μb = 0.95.

Para esta alternativa planteamos que cada consulta del proceso de poisson principal se dividan en 2 procesos de poisson distintos, una para la base “A” y otro para la base “B” de manera de calcular los tiempos de demora y espera separados de acuerdo a sus respectivas tasas y luego volver a juntarlos.

Con esta alternativa obtuvimos para una sola simulación los siguientes resultados:



Como podemos observar, a medida que aumentó la media de los tiempos de llegada disminuyen los tiempos de espera y aumenta la tasa de finalización de consulta. Por otro lado, se pide en el enunciado se ejecute la simulación anterior 100 veces, obteniendo los siguientes resultados.

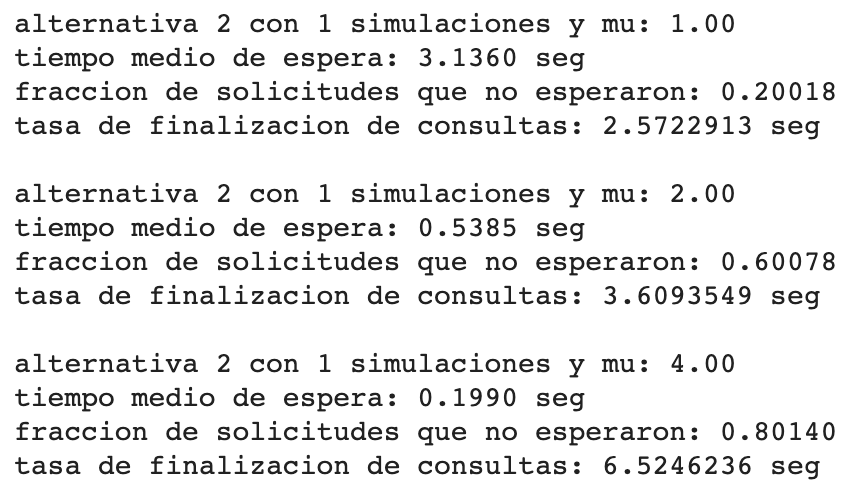


**Alternativa 2**

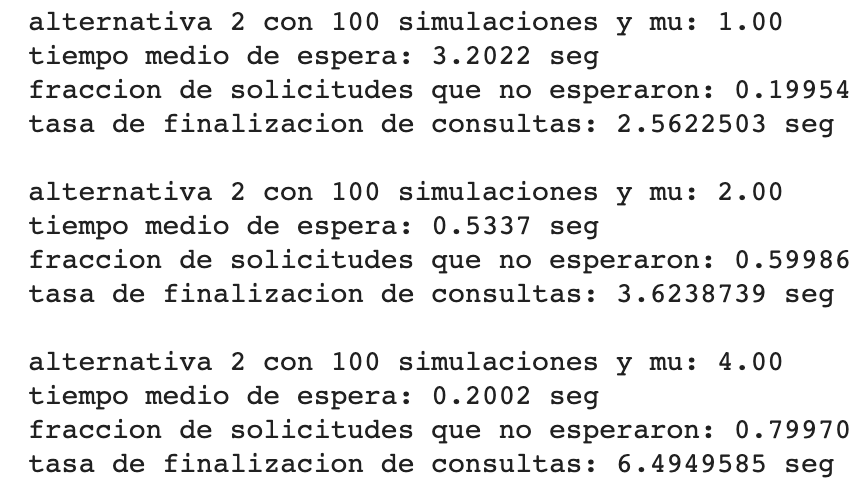
En este caso se utiliza una única base de datos para resolver las consultas, teniendo una media de tiempo de demora de μ = 0.8.

Para esta alternativa se calcularon directamente los tiempos de demora y espera sabiendo que todas las consultas llegaban a la misma base de datos.

De esta manera obtuvimos para una sola simulación los siguientes resultados:



Finalmente repetimos esta simulación 100 veces obteniendo los siguiente resultados:



**Conclusión**

Como conclusión podemos decir que si tuviéramos que elegir entre ambas alternativas optariamos por la primera (la que tiene 2 bases distribuidas) ya que para las variantes de μ simuladas obtiene mejores resultados: mayor tasa de finalización de consultas por segundo, mayor porcentaje de solicitudes que no esperaron y menor tiempo medio de espera por solicitud. Esto se ve reflejado en los siguientes gráficos donde verde es la alternativa 1 y rojo es la alternativa 2:

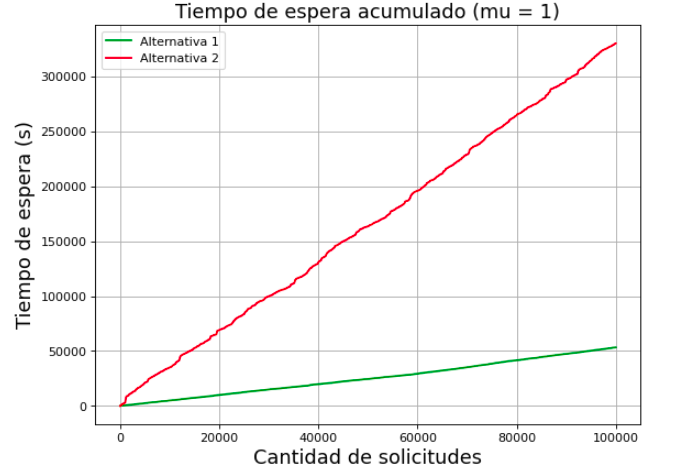


Fig. 1: tiempo de espera acumulado de las solicitudes para ambas alternativas y mu = 1.

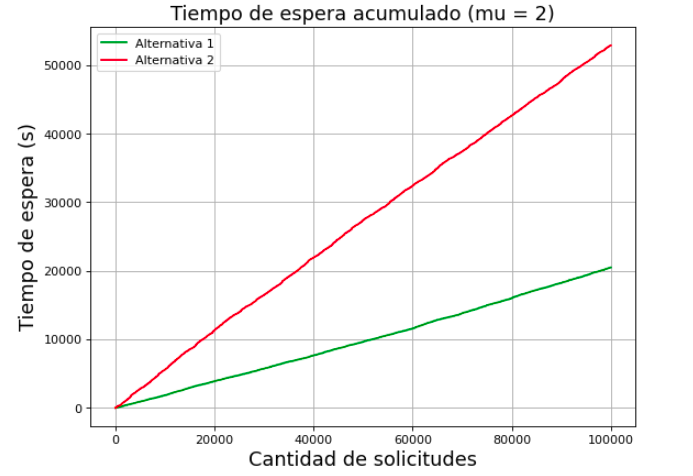


Fig. 2: tiempo de espera acumulado de las solicitudes para ambas alternativas y mu = 2.

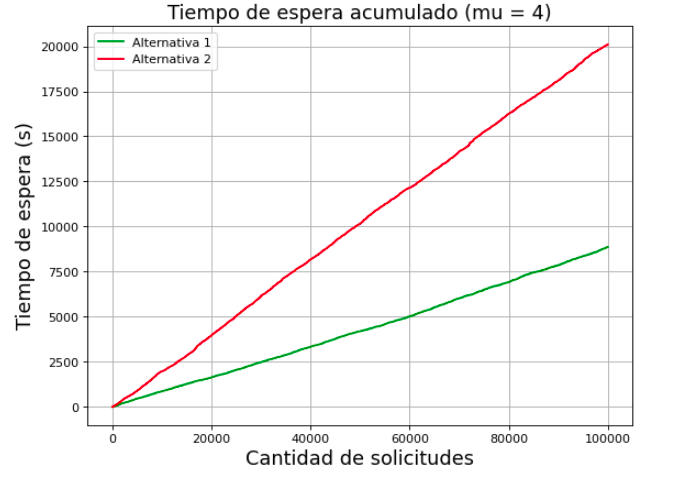


Fig. 3: tiempo de espera acumulado de las solicitudes para ambas alternativas y mu = 4.

# Ejercicio 2

A continuación recreamos y simulamos los modelos planteados por el paper *A Markov Chain Model for Covid-19 Survival Analysis* . Para esto, lo que se realizó fue tomar las probabilidades propuestas en el paper para cada modelo y crear una matriz de transición. Con esta matriz, se comenzó con un conjunto de individuos sanos y se simuló mediante exponer a estos individuos a las probabilidades de transición del modelo. A continuación se ve para cada modelo como evolucionan las poblaciones.

Se puede notar que la diferencia en población no modifica la curva, si no que solo los valores en el eje y.

**Modelo eficiente**

Para 10 individuos

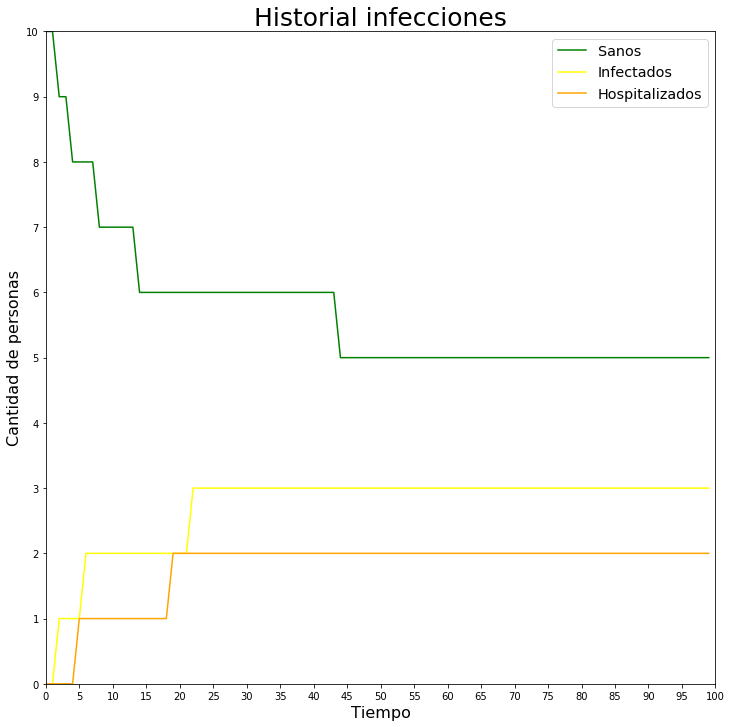


Fig. 4: evolución del modelo eficiente con 10 individuos.

Para 10 000 individuos

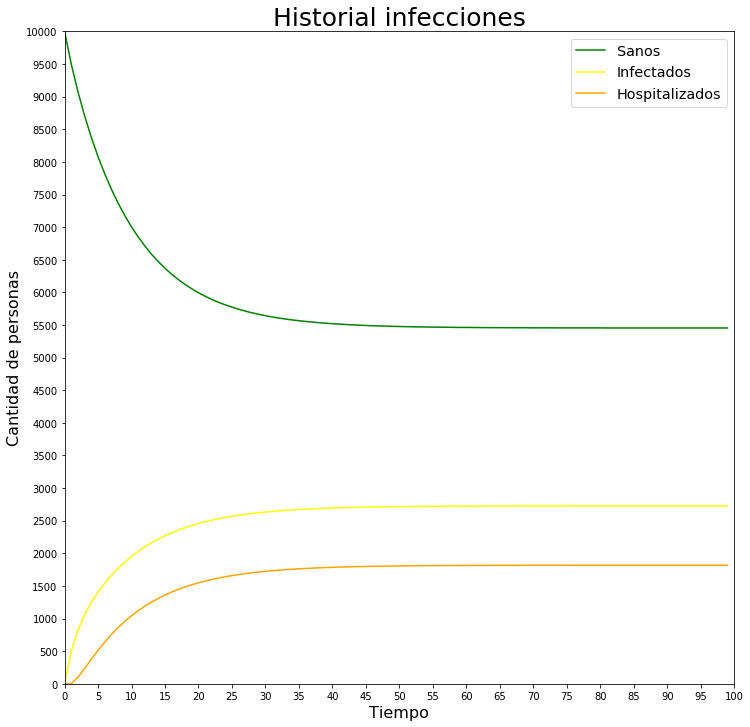


Fig. 5: evolución del modelo eficiente para 10 000 individuos.

Para el modelo eficiente el paper obtiene analiticamente que el estado de equilibrio de la población es (0,545455 0,272727 0,181818), Al simular el modelo iterativamente obtuvimos (0,5454 0,2727 0,1818). Como era de esperarse, los resultados de la simulación coinciden con los presentados en la publicación.

**Modelo ineficiente**

Para 10 individuos

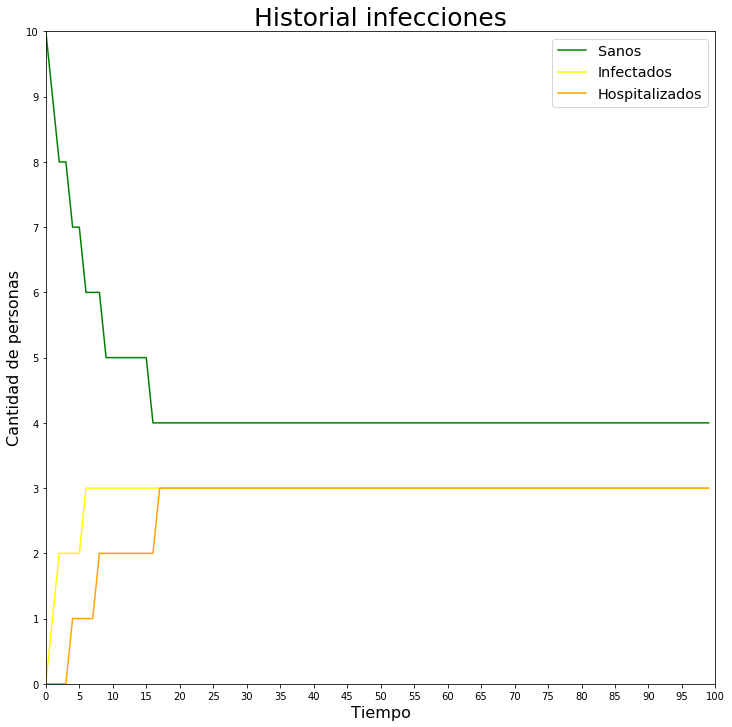


Fig. 6: evolución del modelo ineficiente para 10 individuos.

Para 10 000 individuos

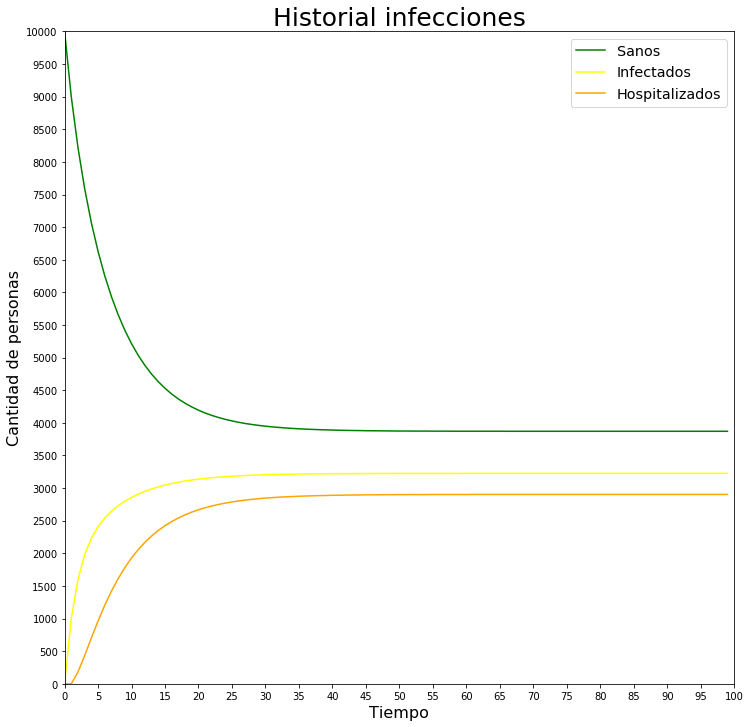


Fig. 7: evolución del modelo ineficiente para 10 000 individuos.

Para el modelo eficiente el paper obtiene analiticamente que el estado de equilibrio de la población es (0,387097 0,322581 0,290323), Al simular el modelo iterativamente obtuvimos (0,3870 0,3225 0,2903). Nuevamente los resultados coinciden.

**Modelo complejo**

Para 10 individuos

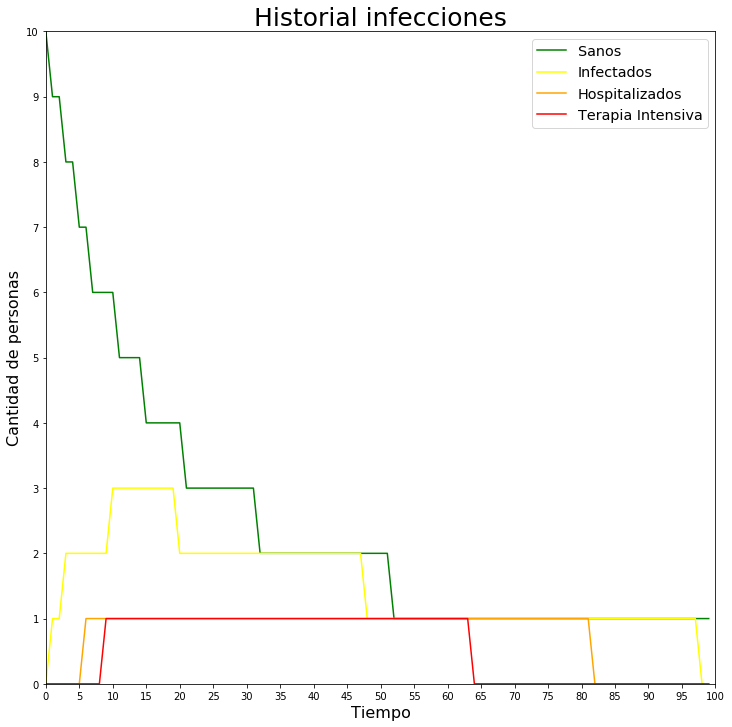


Fig. 8: evolución del modelo complejo para 10 individuos.

Para 10 000 individuos

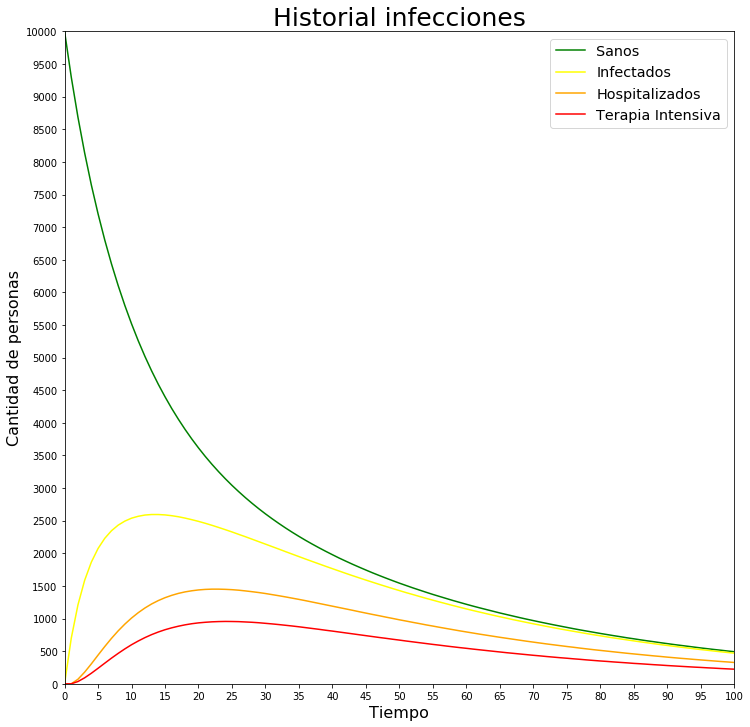


Fig. 9: evolución del modelo complejo para 10 000 individuos.

Para el modelo complejo encontramos una equivalencia entre el análisis y la simulación. Según el estudio analítico le toma a un individuo sano en promedio 59.52 días en fallecer, mientras que en nuestra simulación al realizar operaciones para obtener este valor, obtenemos también que le toma a un individuo 59.52 días en promedio en fallecer.

**Conclusión**

Como se mostró anteriormente, los resultados analíticos obtenidos por el estudio concuerdan con lo simulado. Sin embargo, hay cosas que se ven de forma más veloz solo luego de haber realizado la simulación. Un ejemplo de esto es que el mayor cambio entre iteraciones se concentra en las primeras instancias, donde se pasa rápidamente de un estado en donde todos están sanos a un estado donde un gran porcentaje de individuos está infectado. Esto nos lleva a dudar de la validez de algunos números obtenidos. En particular, el promedio que tarda un individuo en fallecer desde el estado de sano en total es 59 dias, pero si cortamos la simulación antes de que todos los individuos fallecen y calculamos en esa instancia el promedio de días transcurridos hasta fallecer, obtenemos que el 84% de los individuos fallecieron y que solo tomaron en promedio 44 días. Esto nos recuerda, como los mismos autores del estudio aclaran, que todos los modelos están mal, solo que algunos son útiles. Finalmente, cabe aclarar que una muy buena manera de constatar la veracidad y validez de un estudio analítico es mediante una recreación simulada.

# Ejercicio 3

**Modelo epidemiológico**

El modelo SIR para el estudio de epidemias está definido por las siguientes ecuaciones:

(1)

(2)

(3)

donde:

* y son las cantidades de personas sanas, infectadas y recuperadas;
* es la cantidad de personas;
* , y son las respectivas variaciones en función del tiempo; y
* y son las tasas de transmisión y recuperación (que son datos del enunciado).

De acuerdo a los datos del enunciado, y tomando un valor de , se consideran como valores iniciales , y .

Para realizar una simulación utilizando este modelo, basta tomar un método de integración numérica, y utilizarlo para obtener de forma discretizada las funciones solución se las ecuaciones diferenciales anteriores, (dentro de un intervalo acotado). Para esto, se considera el método de integración numérica Runge-Kutta de orden 4 (para las 3 variables) con un paso de 0,1, tomado en días, durante 150 días, entendiendo a este valor ( al igual que el de ) como representativo para la simulación.

A continuación, se muestran los gráficos asociados a las curvas que describen la evolución de la epidemia.

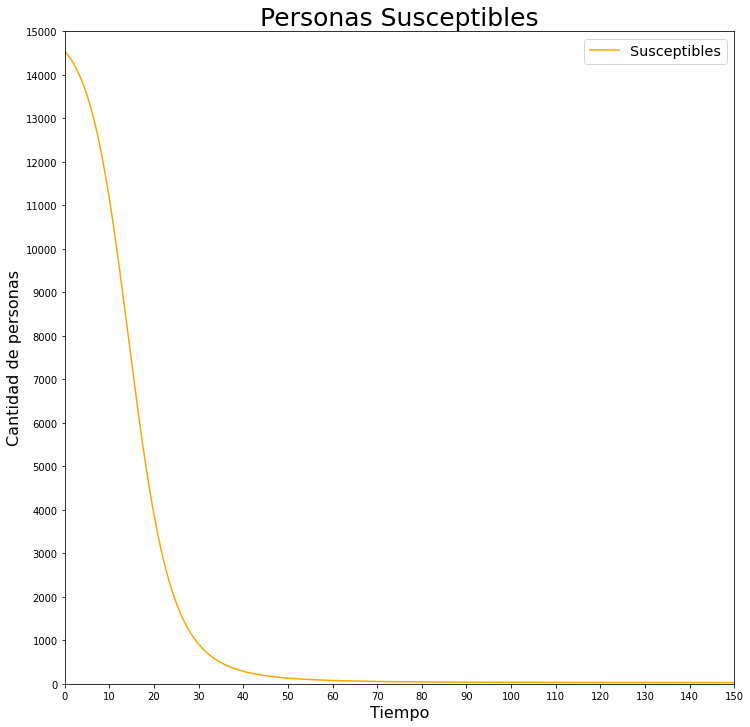


Fig. 10: cantidad de personas susceptibles de enfermarse en función del tiempo.

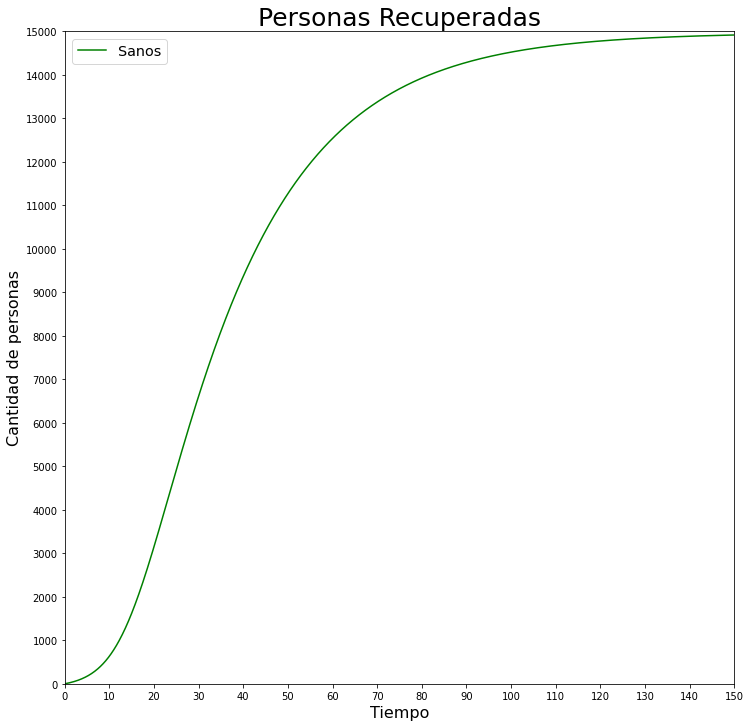


Fig. 11: cantidad de personas recuperadas en función del tiempo.

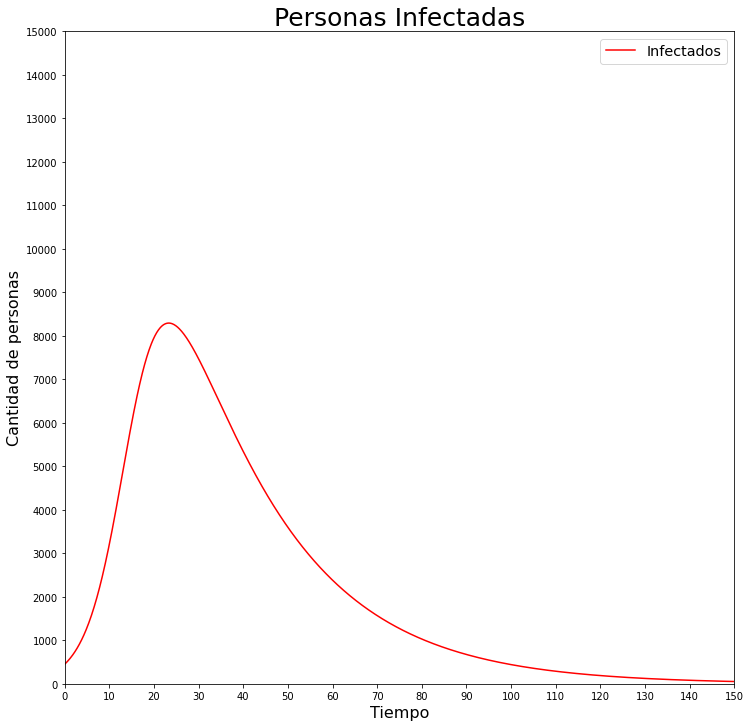


Fig. 12: cantidad de personas infectadas en función del tiempo.

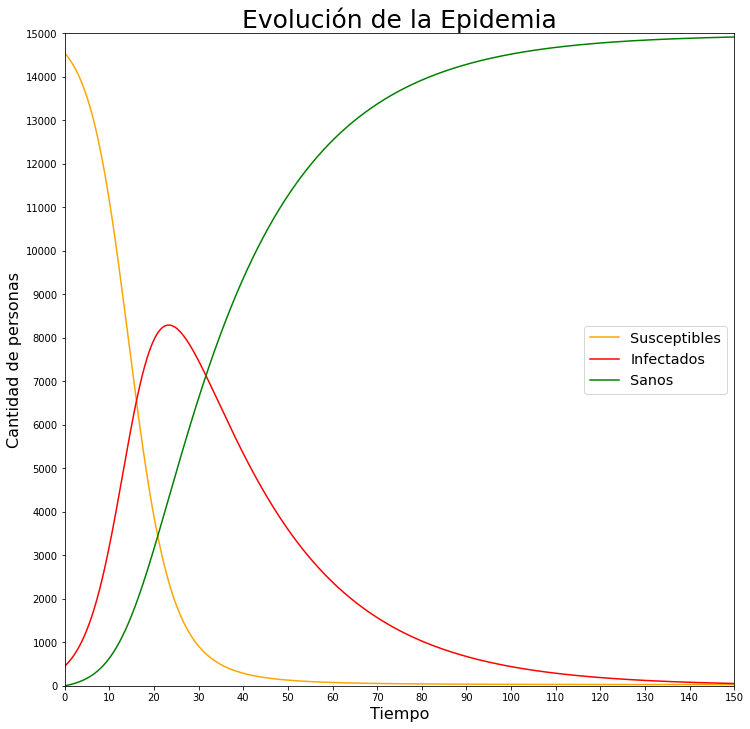


Fig. 13: superposición de las figuras 10, 11 y 12.

Dada la hipótesis del modelo de que los agentes no fallecen por la enfermedad, la tasa de personas susceptibles de contagio decrece de forma monótona, así como aumenta de forma análoga la cantidad de recuperados. Por otra parte, el entrecruzamiento de las curvas de personas recuperadas y susceptibles en este caso es cercano al momento correspondiente del pico de infectados. El mismo se puede "predecir" a partir de la condición de pendiente de recta tangente nula sobre la derivada de la cantidad de infectados:

De (2):

Esto es decir que cuando la cantidad de personas susceptibles de enfermarse sea aproximadamente 2400 (o bien, el 16% de N), entonces la cantidad de personas enfermas será máxima, lo cual será aproximadamente a los 23 días. El pico de infectados también es representativo de la cantidad máxima de personas que deberían ser asistidas de forma simultánea en el caso de que la enfermedad lo requiriera de forma explícita. Esto es decir que si la capacidad de atención médica de la región donde se encuentra la población (medida en cantidad de personas que se pueden atender simultáneamente) es inferior al valor pico de la curva de infectados, entonces no se podrá atender a toda la población que lo requiera durante un período de tiempo. Independientemente de que el modelo no considere la posibilidad de fallecimiento, esto último podría corresponderse con daños irreparables en la salud de aquellos personas que no pudieron ser atendidas, como secuelas de por vida.

**Consideración de la capacidad límite de atención**

Como el modelo se estabiliza cuando todas las personas se recuperan (y por ende, posteriormente a que todas hayan sido infectadas), se tiene que el área bajo la curva de infectados siempre es N. Esto es decir que si el sistema modelado muestra un pico pronunciado, entonces la epidemia terminará más rápido. En contraste, se puede intentar aumentar el tiempo de duración de la misma con el objetivo de no sobreexigir la capacidad del sistema de salud.

Como se está tratando con ecuaciones diferenciales (sin integraciones que permitan conocer su expresión analítica), no se puede plantear directamente una condición explícita contra el pico de los infectados, pero se puede considerar lo comentado en el párrafo anterior y la tasa de transmisión de la enfermedad , la cual corresponde a la "velocidad" con la cual los agentes se enferman. Esto sería coherente teniendo en cuenta que posiblemente sea una variable manipulable en la práctica, a partir de medidas de control sobre la movilización de personas, mientras que posiblemente esté más asociada a la naturaleza de la enfermedad, y no tanto a factores externos. Entonces, logrando disminuir la tasa de contagios se puede aumentar el tiempo que la epidemia tarda en terminar, logrando un pico de infectados menor. Si con una tasa de infección de 27% se llegó a un pico de aproximadamente un 56% de la población (8 300 personas de las 15 000 simuladas), haciendo una aproximación lineal de la dependencia entre y (máxima cantidad de personas infectadas en simultáneo), entonces para lograr:

la aproximación lineal queda:

donde se indica N en vez de N considerando que la aproximación lineal tendrá un error asociado, que puede ser amortiguado por el sobre cálculo (entendiendo 5% como el menor valor porcentual que tiene sentido discretizar para un uso corriente de cota de error, manteniendo el mismo valor que el pedido si se considera una sola cifra representativa).

Entonces, para mantenerse en un valor seguro, se propone reducir a 0,12 el valor de la tasa de infección. Vemos que esta aproximación parece haber sido adecuada:

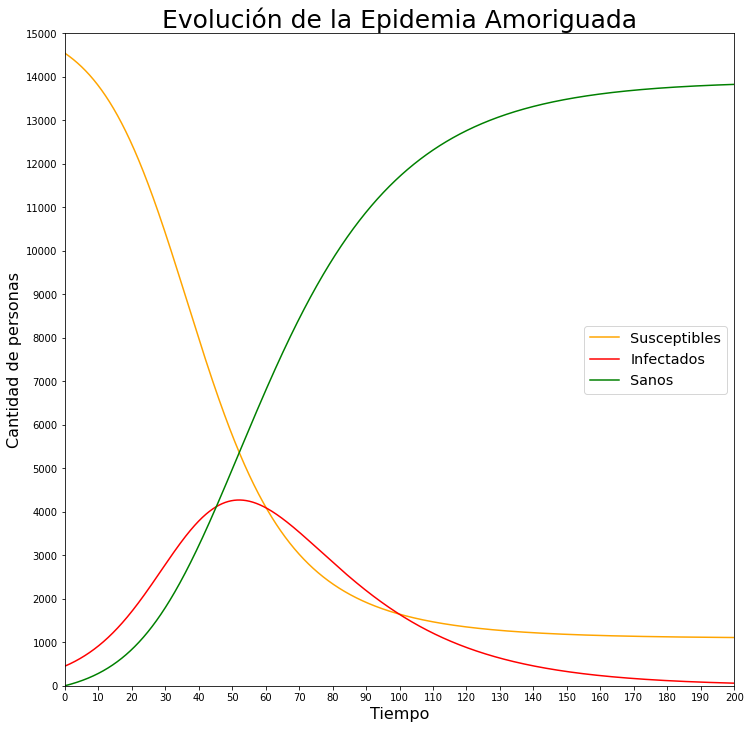


Fig. 14: evolución de la epidemia con una tasa de transmisión de menos de la mitad.

Afortunadamente, la aproximación lineal logra un pico de algo menos de 4 300 personas infectadas, lo cual ronda el 28% de la población. Es decir, no es exactamente lineal la relación entre la tasa de infectados y el pico de la epidemia, pero la cota de error con discretización de uso corriente resultó adecuada para la incerteza introducida.

**Modelo alternativo: duplicando la tasa de infección**

A partir de lo estudiado anteriormente, la situación recíproca se daría con una pandemia más acelerada, con un correspondiente pico de infectados más drástico (por ejemplo, si no hay ningún tipo de recaudo sobre la movilización de la población).

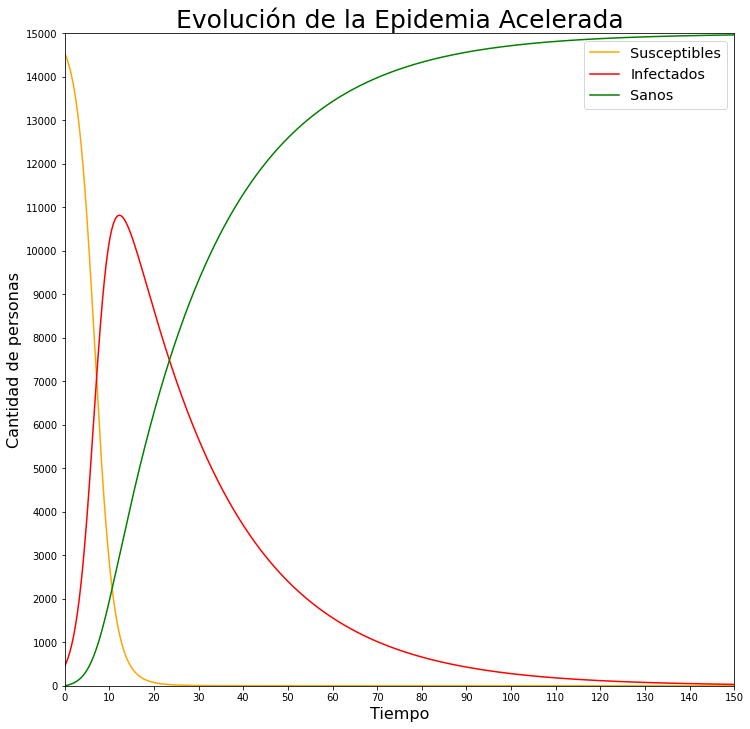


Fig. 15: evolución de la epidemia con una tasa de transmisión de duplicada.

El resultado es claro: en menos del tiempo dado por la duración promedio de la enfermedad:

(días)

se alcanza el peor momento de la epidemia, con alrededor de un 60% de la población infectada en menos de 2 semanas.

Se ve que el modelo alternativo propuesto no se asemeja a ninguno de los modelos analizados en el ejercicio anterior. La única curva representada en forma análoga en ambos ejercicios es la de susceptibles (identificada de forma análoga la cantidad de sanos), ya que los infectados en el ejercicio anterior están segmentados entre los que están hospitalizados y los que no. Se aprecia cómo la curva de personas susceptibles en el ejercicio anterior se estabiliza rápidamente en un valor no nulo tanto para el caso del sistema eficiente como para el ineficiente, cosa que no ocurre en este último modelo alternativo presentado. Con respecto a la curva de infectados, solo en el caso complejo del ejercicio anterior se aprecia un pico (máximo que no es el de estabilización), pero el mismo es mucho menos marcado, afectando a un porcentaje considerablemente menor de la población, y con un decrecimiento más moderado.

**Modelo alternativo: enfermedad de rápida recuperación**

Retomando el caso original, puede considerarse la variante de que las personas puedan recuperarse de la enfermedad en menos de una semana. Por proponer un valor, puede plantearse:

Para las condiciones dadas se da un caso interesante: la recuperación es tan rápida que menos de la mitad de las personas llega a infectarse:

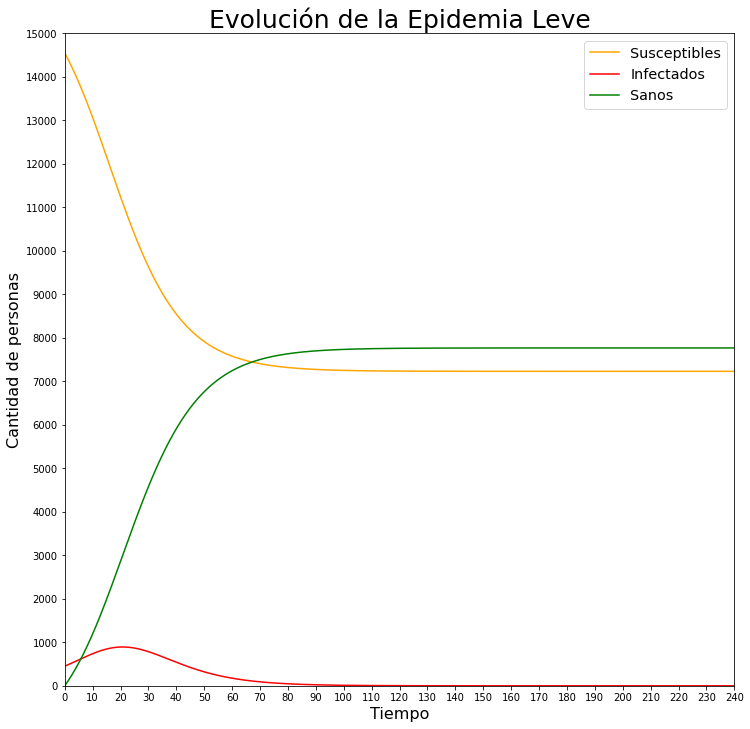


Fig. 16: evolución de la epidemia con recuperación doblemente rápida.

Con respecto a la comparación con el ejercicio anterior, en este caso sí se ve una estabilización algo más veloz de la curva de infectados, y la misma se da luego de un impacto acotado sobre la población, al igual que en el caso del modelo complejo del ejercicio anterior. Por otra parte, la forma en que se estabiliza la curva de susceptibles difiere de la forma en que evoluciona la correspondiente curva de sanos en el ejercicio anterior, considerando que en el gráfico se muestra el doble del rango temporal con respecto al ejercicio anterior

**Modelo alternativo: alto porcentaje de infectados iniciales**

Puede considerarse que ocurre si en la situación inicial se inicia con un valor considerablemente más alto de personas enfermas. Se propone un porcentaje inicial del 30% de personas infectadas.

Lo interesante de este caso es que el gráfico queda similar al modelo inicial, pero "desplazado hacia la izquierda": si se compara con la primera evolución, se ve que las curvas tienen las mismas características, siendo el nuevo gráfico un truncamiento del anterior en apariencia. Esto recuerda a la "falta de memoria" de los procesos de Markov. En este caso, se estableció como valor inicial un porcentaje de infectados alcanzado previamente por el modelo anterior, y lo que se evidencia (con respecto de la parte izquierda de la curva de infectados) es que el desenvolvimiento es similar al restante del modelo original luego de alcanzar dicho porcentaje.

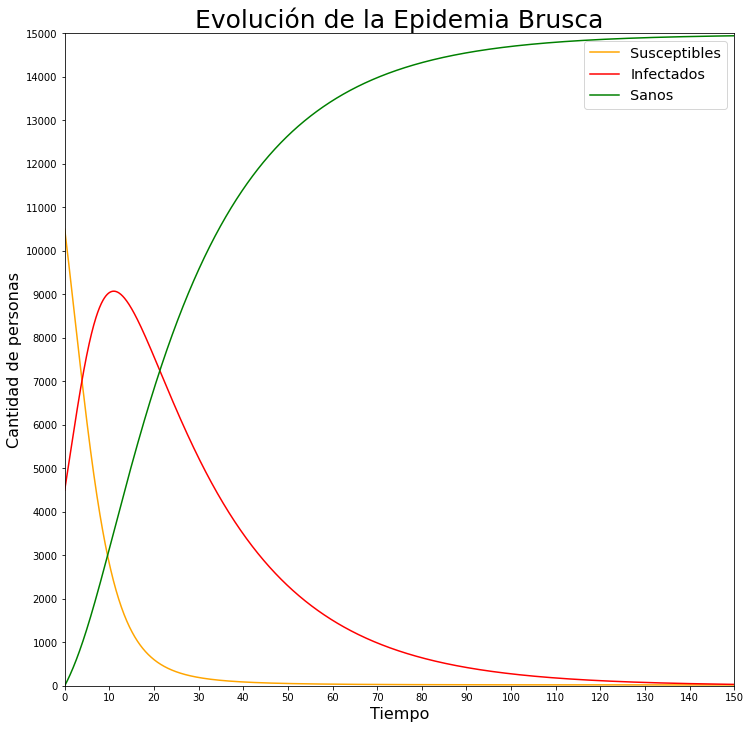
****

Fig. 17: evolución de la epidemia con el 30% de la población infectada inicialmente.

En sintonía plena con el primer modelo alternativo presentado, no se ven semejanzas con respecto a ninguno de los tres sistemas analizados en el ejercicio anterior.

**Conclusión**

Aunque el modelo SIR sea una clara simplificación para estudiar la evolución de una epidemia, el mismo es útil para analizar ciertos aspectos de la misma a partir de los coeficientes de las tasas de infección y recuperación. Las simulaciones hechas permiten entender, como poco a nivel cualitativo, el impacto que tiene la duración de la enfermedad en un agente o la tasa de infección en la fluctuación de la cantidad de infectados. En este sentido, las simulaciones sobre modelos simples permiten tener indicadores de granularidad gruesa para entender el impacto de ciertas condiciones en una población amenazada por una enfermedad contagiosa.

Por otro lado, es interesante ver como un modelo de integración numérica de ecuaciones diferenciales de orden cuatro puede lograr soluciones aproximadas precisas a pesar de no estar utilizando un paso demasiado bajo. Esto se puede visualizar en la sinuosidad de las curvas, y da la pauta de que la familia de modelos Runge-Kutta en órdenes altos puede ser aplicada para problemas de interés, a diferencia de otros métodos explícitos o implícitos de menor orden, como el conocido método de Euler.

# Ejercicio 4

Para este ejercicio se pide realizar la simulación del ejercicio 1 utilizando la librería SimPy. Para todo lo pertinente al enunciado, referirse a la sección ejercicio 1 de este documento. En esta sección se detalla la resolución del mismo en términos de la librería pedida.

**Componente activo: pedidos**

Los pedidos que llegan al servidor web son los componentes activos de la simulación. Éstos son representados con una función generadora a través de la cual se simula el tiempo de llegada según una variable aleatoria exponencial. Al llegar, realizan la query, interactuando con el otro componente principal: la base de datos.

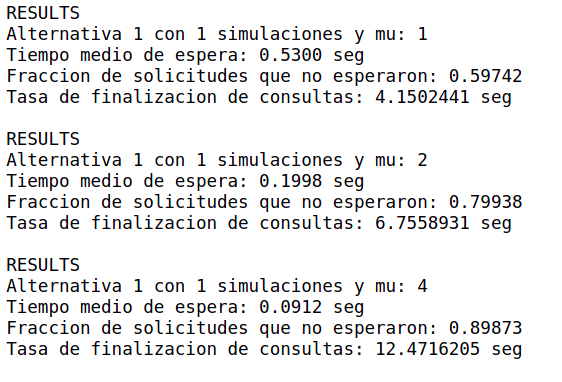
**Componente pasivo: base de datos**

La base de datos es un componente pasivo en la simulación. Es un recurso que tiene capacidad para atender 1 solo pedido a la vez. Todo pedido que intente utilizarla cuando está en uso esperará hasta que se encuentre disponible. La simulación del procesamiento del pedido se hace a través de un tiempo aleatorio exponencial.

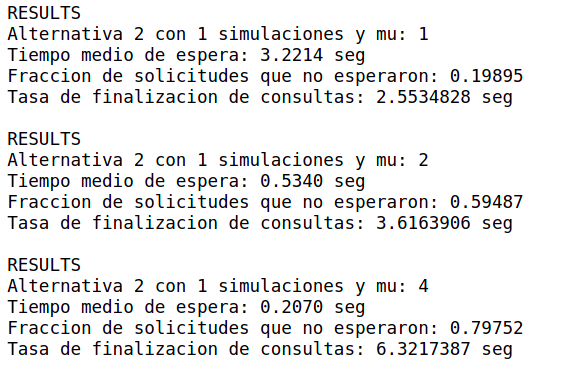
**Control**

Toda la simulación ocurre dentro de un ambiente o environment. Gracias a este elemento, es posible obtener los tiempos clave de los pedidos: generación, espera y procesamiento. Además, la simulación se realiza hasta que todos los 100.000 pedidos hayan sido procesados.

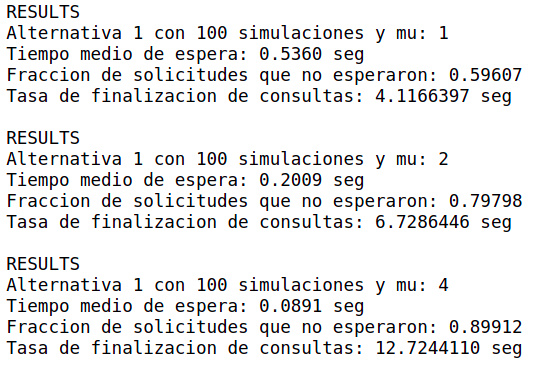
**Resultados: Alternativa 1; 1 Simulación**



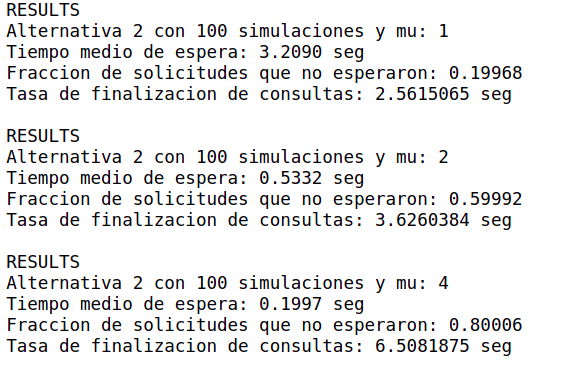
**Resultados: Alternativa 2; 1 Simulación**



**Resultados: Alternativa 1; 100 Simulaciones**



**Resultados: Alternativa 2; 100 Simulaciones**



**Comparaciones**

Los resultados son similares a los obtenidos en el ejercicio 1. Para ambas alternativas y para los tres valores de mu, los resultados coinciden fidedignamente tanto para una simulación como para cien. Esto tiene sentido, ya que la librería utilizada no debería tener influencia en los resultados. SimPy es simplemente una herramienta de simulación de eventos discretos que se presenta como una buena alternativa a la codificación de situaciones como esta desde cero.

**Conclusión**

Las conclusiones son las mismas que en el ejercicio 1. La mejor alternativa es la primera, la que cuenta con dos bases de datos distribuidas, debido a que ofrece un mejor servicio. Los gráficos obtenidos también son muy similares a los del ejercicio 1.

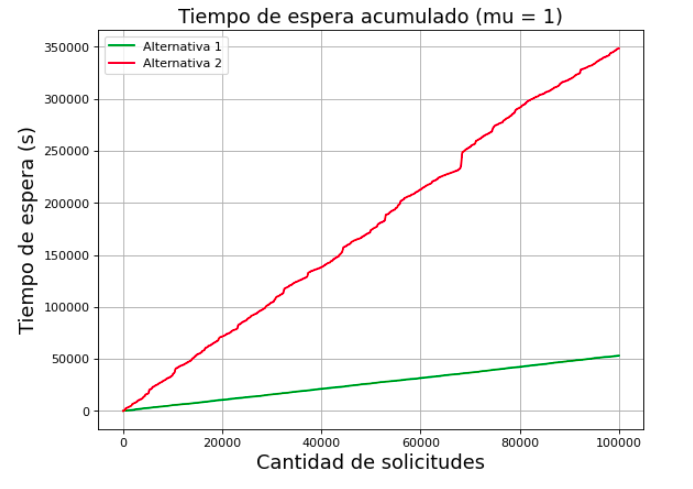


Fig. 18: tiempo de espera acumulado de las solicitudes para ambas alternativas y mu = 1

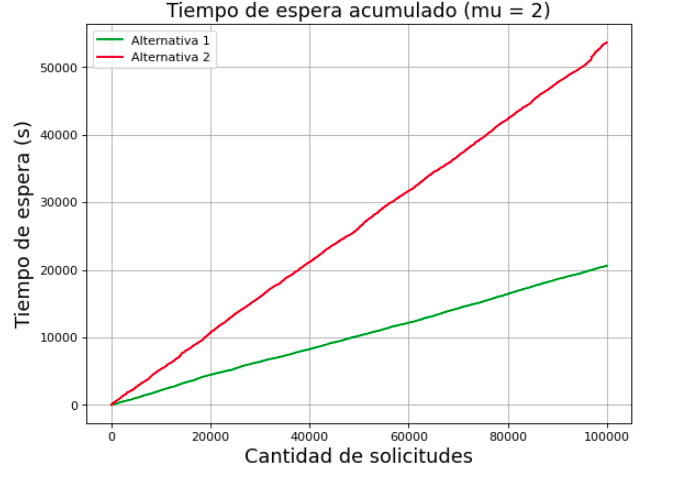


Fig. 19: tiempo de espera acumulado de las solicitudes para ambas alternativas y mu = 2.

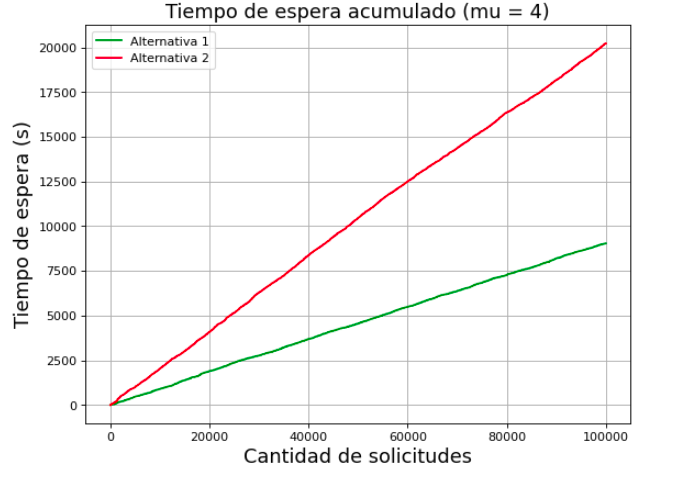


Fig. 20: tiempo de espera acumulado de las solicitudes para ambas alternativas y mu = 4.

# Anexo: código fuente

Se adjuntan los notebooks utilizados para lograr la solución descrita.