

Reporte: Modelo SIR y Simulación Estocástica

López Pérez Olin Tonatiuh

7/julio/2023

1 Introducción

Los modelos epidemiológicos son herramientas poderosas para comprender y predecir la propagación de enfermedades infecciosas en una población. Estos modelos, basados en principios matemáticos y estadísticos, permiten a los investigadores y expertos en salud analizar la dinámica de la enfermedad, evaluar el impacto de intervenciones y tomar decisiones informadas para controlar y mitigar brotes epidémicos.

Uno de los modelos más utilizados en epidemiología es el modelo SIR, que recibe su nombre de los tres compartimentos que representa: Susceptibles, Infectados y Recuperados. Este modelo fue propuesto por los matemáticos Kermack y McKendrick en 1927 como una extensión del modelo SI (Susceptibles-Infectados) originalmente desarrollado por Ross en 1911.

El modelo SIR se basa en supuestos simplificados pero realistas sobre la propagación de enfermedades infecciosas en una población. Considera que los individuos pueden estar en uno de los tres estados: susceptibles, que aún no han sido infectados y pueden contraer la enfermedad; infectados, que están actualmente infectados y pueden transmitir la enfermedad a los susceptibles; y recuperados, que se han recuperado de la enfermedad y son inmunes, al menos a corto plazo.

La transición entre estos compartimentos se modela mediante ecuaciones diferenciales que describen las tasas de cambio de cada grupo a lo largo del tiempo. En el modelo SIR, se asume que la tasa de transmisión de la enfermedad depende de la cantidad de individuos infectados y susceptibles en un momento dado, y que los individuos infectados se recuperan y adquieren inmunidad de manera irreversible.

A lo largo de los años, el modelo SIR ha sido ampliamente utilizado en el campo de la epidemiología para estudiar enfermedades infecciosas como la gripe, el sarampión, la tuberculosis y más recientemente, el COVID-19. Ha demostrado ser una herramienta valiosa para comprender la dinámica de la propagación de enfermedades, evaluar el impacto de intervenciones como la vacunación y la cuarentena, y predecir escenarios futuros.

Además del modelo SIR, existen otros modelos epidemiológicos que incorporan más compartimentos y consideran aspectos adicionales de la propagación de enfermedades. Por ejemplo, el modelo SEIR agrega un compartimento de expuestos, que representa a individuos infectados pero no contagiosos, y permite capturar el período de incubación de la enfermedad. Otros modelos más complejos incluyen compartimentos para individuos susceptibles vacunados, individuos asintomáticos, individuos en cuarentena, entre otros.

Estos modelos epidemiológicos han sido fundamentales en la comprensión de las dinámicas de las enfermedades infecciosas y han contribuido a la toma de decisiones informadas en salud pública. Sin embargo, es importante tener en cuenta que estos modelos se basan en supuestos simplificados y simplificaciones matemáticas, y que los resultados obtenidos pueden variar según los parámetros y condiciones específicas del brote epidémico.

En este trabajo, nos centraremos en el modelo SIR y utilizaremos la simulación estocástica para analizar su comportamiento en diferentes escenarios. A través de múltiples simulaciones, exploraremos las fluctuaciones y la incertidumbre asociadas a la propagación de la enfermedad y compararemos los resultados con las predicciones teóricas del modelo. El objetivo es obtener una comprensión más completa y realista de la dinámica epidémica y evaluar la validez del modelo SIR en diferentes contextos.

2 Obtención de las ecuaciones del modelo SIR

El modelo SIR se basa en un sistema de ecuaciones diferenciales que describen las tasas de cambio de los compartimentos de Susceptibles (S), Infectados (I) y Recuperados (R). Para obtener estas ecuaciones, consideramos las siguientes premisas:

1. La tasa de cambio de los Susceptibles es igual a la tasa a la cual los individuos se vuelven Infectados, restándoles la tasa a la cual se recuperan:

$$\frac{dS}{dt} = -\beta SI + \gamma R$$

2. La tasa de cambio de los Infectados es igual a la tasa a la cual los individuos se vuelven Infectados (provenientes de los Susceptibles) menos la tasa a la cual se recuperan:

$$\frac{dI}{dt} = \beta SI - \gamma I$$

3. La tasa de cambio de los Recuperados es igual a la tasa a la cual los individuos se recuperan (provenientes de los Infectados):

$$\frac{dR}{dt} = \gamma I$$

Donde: - β es el parámetro que representa la tasa de transmisión de la enfermedad. - γ es el parámetro que representa la tasa de recuperación de la enfermedad.

Estas ecuaciones describen cómo cambian las cantidades de individuos en cada compartimento a lo largo del tiempo, teniendo en cuenta la transmisión de la enfermedad de los Susceptibles a los Infectados y la recuperación de los Infectados.

3 Adimensionalización de las ecuaciones

Para facilitar el análisis y la interpretación de las ecuaciones del modelo SIR, es común realizar una adimensionalización de las mismas. Esto implica escalar las variables y los parámetros para eliminar unidades y simplificar las expresiones.

Consideremos las siguientes adimensionalizaciones:

1. Escalamos el tiempo dividiéndolo por una constante T :

$$\tau = \frac{t}{T}$$

2. Escalamos las cantidades de individuos en cada compartimento dividiéndolas por una constante N (la población total):

$$s = \frac{S}{N}, \quad i = \frac{I}{N}, \quad r = \frac{R}{N}$$

3. Escalamos los parámetros de tasa de transmisión y tasa de recuperación dividiéndolos por la constante T^{-1} :

$$\beta' = \frac{\beta}{T}, \quad \gamma' = \frac{\gamma}{T}$$

Sustituyendo estas adimensionalizaciones en las ecuaciones del modelo SIR, obtenemos:

$$\frac{ds}{d\tau} = -\beta' si + \gamma' r$$

$$\frac{di}{d\tau} = \beta' si - \gamma' i$$

$$\frac{dr}{d\tau} = \gamma' i$$

Para obtener una expresión para x , definimos $x = \frac{I}{N}$. Integrando la segunda ecuación, tenemos:

$$\int \frac{di}{\beta' si - \gamma' i} = \int d\tau$$

Esto nos lleva a:

$$\frac{1}{\beta'} \ln \left(\frac{s}{s_0} \right) - \frac{\gamma'}{\beta'} \ln \left(\frac{1-x}{1-x_0} \right) = \tau + C$$

Donde s_0 y x_0 son las condiciones iniciales de s y x respectivamente. Resolviendo para x , obtenemos:

$$x = 1 - (1 - x_0)e^{-\alpha\tau} - s$$

Reemplazando $\tau = \frac{t}{T}$, obtenemos finalmente:

$$x = 1 - (1 - x_0)e^{-\alpha t} - s$$

Esta expresión adimensionalizada para x nos permite analizar y representar el comportamiento de la proporción de individuos infectados en función del tiempo, considerando las condiciones iniciales y los parámetros del modelo.

4 Procesos de Poisson y simulación estocástica

En el modelo SIR, En algún momento no hay flujo de individuos entre los compartimentos, es decir, no hay transiciones entre susceptibles a infectados ni de infectados a recuperados. Sin embargo, en algún momento posterior, una transición puede ocurrir de manera aleatoria. Para modelar esta incertidumbre, se utiliza una distribución exponencial.

Se asume que la probabilidad de que un individuo pase de un compartimento a otro sigue una distribución exponencial. Esto implica que hay dos procesos de Poisson entrelazados en el modelo SIR. La distribución exponencial describe el tiempo entre eventos y se utiliza para representar los tiempos de transición entre los diferentes estados del modelo.

En este contexto, el proceso de Poisson es un proceso estocástico que describe eventos que ocurren de manera aleatoria en el tiempo, de tal manera que la probabilidad de que ocurra un evento en un intervalo específico es proporcional a la duración del intervalo. En el modelo SIR, se utiliza el proceso de Poisson para simular la ocurrencia de transiciones entre compartimentos.

La simulación se realiza en el caso discreto del modelo SIR, donde las poblaciones de susceptibles, infectados y recuperados se representan mediante números enteros. La idea es simular escenarios en los que las transferencias entre compartimentos ocurren de manera impredecible y en momentos variables, reflejando así de manera más realista la propagación de una epidemia.

Para determinar cuál será la siguiente transición a ocurrir, se utiliza el tiempo de transición más corto de las dos ecuaciones del modelo SIR. Esto significa que se selecciona el mínimo entre los tiempos esperados de transición de las ecuaciones. De esta manera, se determina el próximo evento de transición.

En el contexto de la simulación, se utiliza una variable exponencial cuyo parámetro es la suma de los parámetros individuales de las distribuciones exponenciales que representan las transiciones. Esta variable exponencial modela el tiempo mínimo de transición entre los estados del modelo.

La simulación estocástica permite cerrar los procesos y simular la propagación de la enfermedad en el modelo SIR de manera realista. Al considerar la aleatoriedad en las transiciones y los tiempos de ocurrencia, se pueden explorar escenarios más cercanos a la realidad y obtener una comprensión más completa de la propagación de una epidemia.

5 Metodología

En esta sección, describiremos la metodología utilizada para realizar las simulaciones del modelo SIR y analizar los resultados.

A lo largo de las realizaciones, se consideraron diferentes casos variando los parámetros del modelo. Se tomaron valores de α de 1.5, 2, 2.5 y 3 y se seleccionaron diferentes poblaciones N de 1000, 10000, 100000 y 1000000 individuos. Además, se fijó una población inicial infectada (x_0) con valores de 0.0001, 0.001, 0.01 y 0.1. Esto generó un total de 64 combinaciones, para cada una de las cuales se realizaron 300 corridas de simulación.

Durante la simulación, se obtuvo una familia de curvas, donde cada una representa una iteración. Sin embargo, el enfoque principal se centró en analizar el valor máximo de infectados y el punto en el que la pandemia finaliza, es decir, cuando el número de infectados alcanza su punto más bajo y tiende a cero.

Una vez completado el código de simulación, se procedió a cambiar los parámetros mencionados anteriormente (población inicial, población inicial infectada y parámetro α). De esta manera, se obtuvo una muestra de datos lo suficientemente grande y diversa para realizar comparaciones.

Se descartaron los datos de simulaciones que no cumplieran con los requisitos del problema, como aquellas en las que la pandemia no se extinguía o en las que los resultados no eran consistentes.

La razón de realizar un alto número de simulaciones (300 corridas para cada combinación) radica en la posibilidad de obtener el valor esperado de los R_{inf} (número básico de reproducción efectiva) calculados en dichas corridas, así como su varianza. Esto permitió analizar si el valor esperado obtenido de las simulaciones era estadísticamente igual o diferente al R_{inf} calculado mediante las ecuaciones diferenciales ordinarias (EDO) del modelo SIR.

Se calculó el sesgo de los resultados simulados al restar el máximo de los valores obtenidos en las EDO a los valores simulados. Luego, se utilizó la función `erf` (función de error) para realizar el ajuste de los resultados simulados y se graficaron las curvas. Además, se aplicó la prueba de Lilliefors para verificar si los datos seguían una distribución normal con un 95

Es importante mencionar que se utilizó la ecuación adimensional del modelo SIR para realizar las simulaciones, lo cual permitió obtener resultados comparables y coherentes.

6 Análisis de los resultados

Aquí se presentan y analizan en detalle los resultados obtenidos de las simulaciones. Se muestran las gráficas generadas, que representan la evolución de los compartimentos de susceptibles, infectados y recuperados a lo largo del tiempo. Se discute el comportamiento de las simulaciones en relación con el modelo SIR y se realiza una comparación con los valores teóricos. Además, se examinan casos particulares que cumplen con los requisitos deseados y se descartan aquellos que no cumplen con los criterios establecidos.

7 Simulación y Análisis Estadístico

En esta sección, se implementó el modelo SIR utilizando el lenguaje de programación MATLAB. El código utilizado se muestra a continuación tomando en cuenta un parámetro α de 1.5, una población de infectados de $x_0=0.1$ y una población inicial de 1000000, si queremos correr otra simulación, cambiamos estos parámetros:

```
l=300; % número de corridas
RI=zeros(1,1);
XM=zeros(1,1);
RM=zeros(1,1);
RRM=zeros(1,1);
RIN=zeros(1,1);
FInf=zeros(1,1);
SupInfAr=zeros(1,1);
SupInfAb=zeros(1,1);
DIF_Inf=zeros(1,1);
DIF_M=zeros(1,1);
dm=0;
dM=0;

for i=1:l
    % Parámetros iniciales
    a = 1.5;
    x0 = 0.1; % Población inicial infectada
    x = @(r)1-(1-x0)*exp(-a*r)-r;
    rinf = 1 + lambertw(0, -a*(1-x0)*exp(-a))/a;

    % Graficar curva teórica
    plot((0:0.01:rinf), x(0:0.01:rinf), 'r-');
    hold on;
    plot([0, 1], [0, 0], 'k-');

    N = 1000000; % Población total
    X = zeros(1, 3*N); X(1) = x0;
    R = zeros(1, 3*N); R(1) = 0;
    xm = x0;
    k = 1;
```

```

% Simulación del modelo SIR
while (X(k) > 0) && (k < 3*N)
    if (-log(rand(1,1))/a/X(k)/(1-X(k)-R(k)) < -log(rand(1,1))/X(k))
        X(k+1) = X(k) + 1/N;
        R(k+1) = R(k);
    else
        X(k+1) = X(k) - 1/N;
        R(k+1) = R(k) + 1/N;
    end
    if xm < X(k+1)
        xm = X(k+1);
        rm = R(k+1);
    end
    k = k + 1;
end

% Graficar curva simulada
plot(R(1:k), X(1:k), 'b.');
```

```

% Almacenar resultados
RI(i) = R(k); % Infinito_aproximado=RI;
XM(i) = xm;
RM(i) = rm;
RRM(i) = x(rm);
RIN(i) = rinf; % Infinito_teorico=RIN;
DIF_M(i) = RRM(i) - xm;
DIF_Inf(i) = rinf - R(k);
end
```

```

% Graficar distribución de R-infinito
figure
plot(sort(RM), [1:length(RM)]/(1+length(RM)), 'r.')
RI = sort(RI); % Valores de Infinito_aproximado ordenados de forma ascendente
mediaInf = mean(RI); % Promedio de los valores RInf calculados
desviacionInf = std(RI); % Desviación estándar de los valores RInf calculados
plot([min(RI):0.001:max(RI)], (erf(([min(RI):0.001:max(RI)]-mediaInf)/desviacionInf/sqrt(2)) + 1)/2, 'r-')
hold on;

for ll = 1:l-1
    plot([RI(ll), RI(ll+1)], [ll/l, ll/l], 'b-')
    plot(RI(ll), ll/l, 'b.')
end

plot(RI(1), 1, 'b.')
```

```

% Análisis estadístico
for i = 1:l
    FInf(i) = (1 + erf((RI(i)-mediaInf)/(desviacionInf*sqrt(2))))/2;
    SupInfAr(i) = abs(FInf(i)-(i/l)); % cálculo de |F(r)-k/n|
    SupInfAb(i) = abs(FInf(i)-((i-1)/l)); % cálculo de |F(r)-(k-1)/n|
end

MAX1 = max(SupInfAr); % |F(r)-k/n|
MAX2 = max(SupInfAb); % |F(r)-(k-1)/n|
maximadiferencia = max(MAX1, MAX2);
```

```

sesgo = erf(abs(mediaInf-rinf)*sqrt(1)/(desviacionInf/sqrt(2)));

% Mostrar resultados
disp("Valor R_Inf teórico");
disp(rinf);
disp("Promedio de los valores R Infinito calculados: ");
disp(mediaInf);
disp("Desviación de los valores R Infinito calculados: ");
disp(desviacionInf);
disp("F(Rk): ");
disp(maximadiferencia);
disp('Estadístico de prueba para Lilliefors: ');
disp(maximadiferencia*sqrt(1));
disp('Valor para comprobar si es Inssegado o no ');
disp(sesgo);

```

Se muestra la información relevante, como el valor teórico de R-infinito, el promedio y la desviación estándar de los valores simulados, el estadístico de prueba para Lilliefors y el valor de sesgo

Para la simulación de las corridas con el eje x como el tiempo y el eje y como la cantidad de infectados tenemos un ejemplo, donde lo que esta en azul es la simulación, y en rojo el valor teórico.

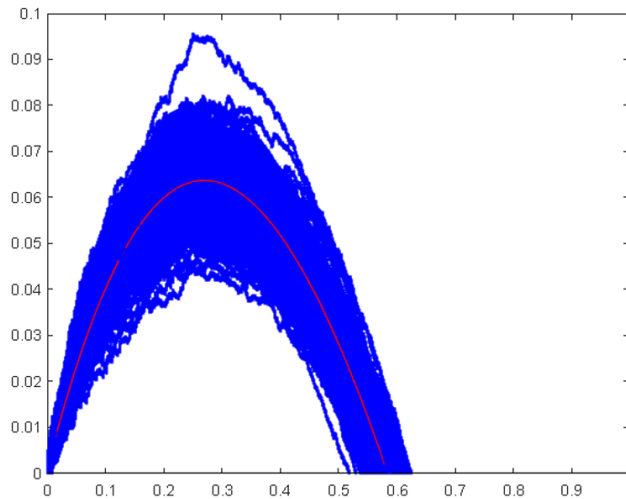


Figure 1: Caso $x_0=0.001$, $N=10000$, $\alpha=1.5$, 300 iteraciones.

8 Prueba de Lilliefors y análisis de las diferencias

En esta sección, se explica cómo se realizó la prueba de Lilliefors utilizando las diferencias entre los valores teóricos y los valores obtenidos en las simulaciones. Se detalla el proceso de cálculo de las diferencias y se utiliza la función `erf` en MATLAB para realizar la prueba. Se discuten los resultados de la prueba y se interpreta si los datos simulados siguen una distribución normal.

En la siguiente figura tenemos un ejemplo donde están los máximos de infectados de cada iteración y se verifica con la prueba de Lilliefors su distribución normal.

En este ejemplo en azul estala simulación y en rojo los datos teoricos.

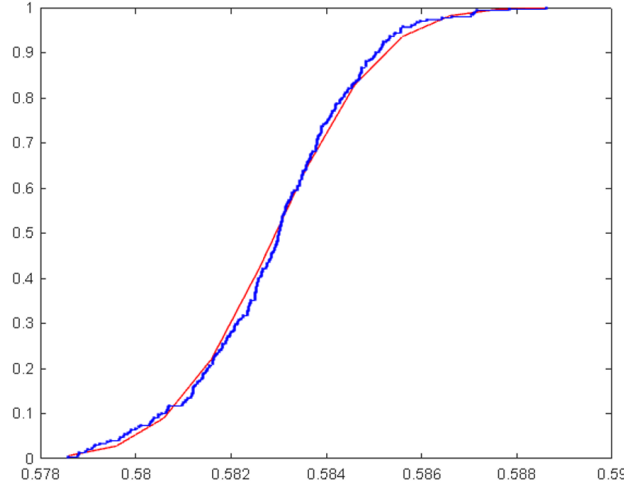


Figure 2: Caso $x_0=0.0001$, $N=1000000$, $\alpha=1.5$, 300 iteraciones.

No todas las gráficas tienen este comportamiento, y las que no lo tienen no cumplen con los requisitos del problema, esto casi siempre asociado a que no tiene sentido, puede haber 0.1 de población infectada en 1000 de población inicial, y el modelo no se ajusta de manera adecuada a casos como estos.

9 Análisis de Variabilidad y Sesgo

Realizaremos un análisis de la variabilidad y el sesgo en los resultados de la simulación del modelo SIR. El objetivo de este análisis es evaluar la calidad de la simulación y determinar si existen diferencias significativas entre los resultados teóricos y los resultados obtenidos en las simulaciones. Esto nos permitirá obtener una comprensión más precisa del comportamiento del modelo SIR en diferentes escenarios y obtener conclusiones más sólidas sobre su validez.

El análisis de variabilidad se enfoca en examinar las diferencias entre los valores teóricos y los valores simulados de (r^∞) y (r_∞) para diferentes combinaciones de los parámetros N y α . Buscaremos identificar patrones de variabilidad en los resultados y determinar cómo cambian a medida que varían estos parámetros. Además, evaluaremos el sesgo en los resultados al calcular el valor absoluto de las diferencias y comparar estos valores con umbrales específicos. El sesgo nos indicará si existen tendencias sistemáticas en los resultados simulados que difieren de los valores teóricos.

A continuación, presentamos el código utilizado para realizar la simulación del modelo SIR y generar los resultados:

```
N=[1,2,3,4];
x0=[-4,-3,-2,-1];
a=[1.5,2,2.5,3];
a1=[];
a2=[];
a3=[0.535,0.0003,0.0001,9E-05];
a4=[0.0000,1E-04,0.0E+00,2E-05];
b1=[];
b2=[0.0115,0.401,0.0002,1E-04];
b3=[0.00002809,0.00002809,0.00003136];
b4=[1.29881E-05,1.15586E-05,1.18832E-05];
c1=[0.00229441,0.00003481,0.00002704,0.00001369];
c2=[0.00004624,0.00000784,0.00000729,0.00000784];
c3=[0.00000289,0.00000289,0.00000324,0.00000196];
c4=[1.33102E-06,1.41396E-06,1.25261E-06,1.02435E-06];
d1=[0.00000289,0.00000324,0.00000324,0.00000144];
d2=[0.00000576,9.16825E-07,7.79053E-07,7.9299E-07];
d3=[2.8395E-07,3.07914E-07,2.96437E-07,2.20618E-07];
```

```

d4=[1.28644E-07,1.34564E-07,1.18811E-07,1.1837E-07];
figure
plot(a,a3,a,a4);
title(['log_{10}(x0)=',num2str(x0(1))])
xlabel('Alfa')
ylabel('(r^{\hat{\infty}})-(r_{\infty})')
legend({'N=100,000','N=1,000,000'},'Location','northeast')
figure
plot(a,b2,a,b3,a,b4);
title(['log_{10}(x0)=',num2str(x0(2))])
xlabel('Alfa')
ylabel('(r^{\hat{\infty}})-(r_{\infty})')
legend({'N=10,000','N=100,000','N=1,000,000'},'Location','northeast')
figure
plot(a,c1,a,c2,a,c3,a,c4);
title(['log_{10}(x0)=',num2str(x0(3))])
xlabel('Alfa')
ylabel('(r^{\hat{\infty}})-(r_{\infty})')
legend({'N=1,000','N=10,000','N=100,000','N=1,000,000'},'Location','northeast')
figure
plot(a,d1,a,d2,a,d3,a,d4);
title(['log_{10}(x0)=',num2str(x0(4))])
xlabel('Alfa')
ylabel('(r^{\hat{\infty}})-(r_{\infty})')
legend({'N=1,000','N=10,000','N=100,000','N=1,000,000'},'Location','northeast')

```

A partir de los resultados obtenidos, podemos observar la variabilidad en los valores de $(r^{\hat{\infty}})$ y (r_{∞}) para diferentes valores de N y α . También podemos evaluar el sesgo al comparar estos valores con los resultados teóricos. A continuación, presentaremos las posibles conclusiones que podemos obtener a partir de este análisis.

Dado que el modelo SIR es un modelo epidemiológico que representa la propagación de una enfermedad en una población, estas gráficas pueden proporcionar información sobre la variabilidad de los resultados del modelo en función de los parámetros.

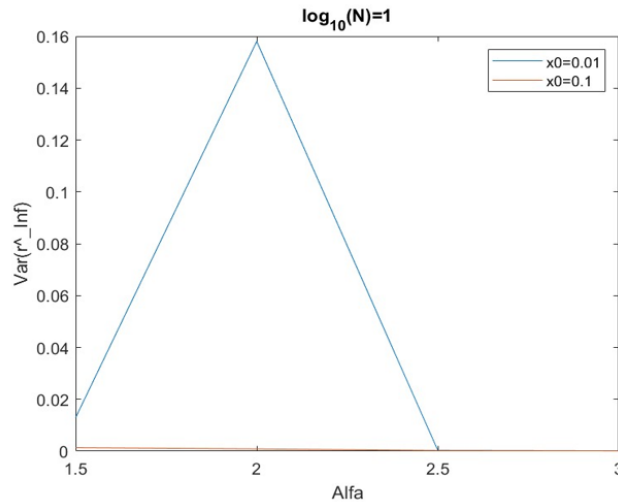


Figure 3: Picos.

Picos en las gráficas: Los picos pueden indicar que hay valores específicos de x_0 o α que generan resultados extremos en términos de variabilidad. Estos valores pueden corresponder a puntos críticos o umbrales donde el sistema exhibe comportamientos significativamente diferentes.

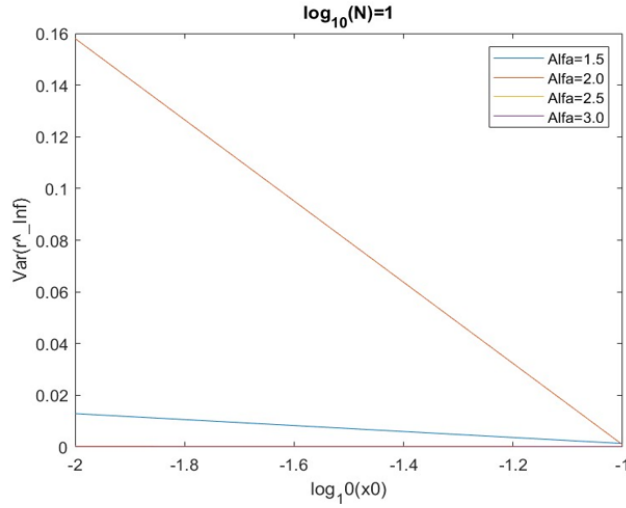


Figure 4: Decrecen.

Decremento general de las gráficas: Si la mayoría de las líneas en las gráficas disminuyen, esto podría indicar que a medida que aumenta x_0 o α , la variabilidad de los resultados del modelo disminuye en general. Esto sugiere que hay ciertos valores de x_0 o α que generan resultados más predecibles y estables en términos de variabilidad.

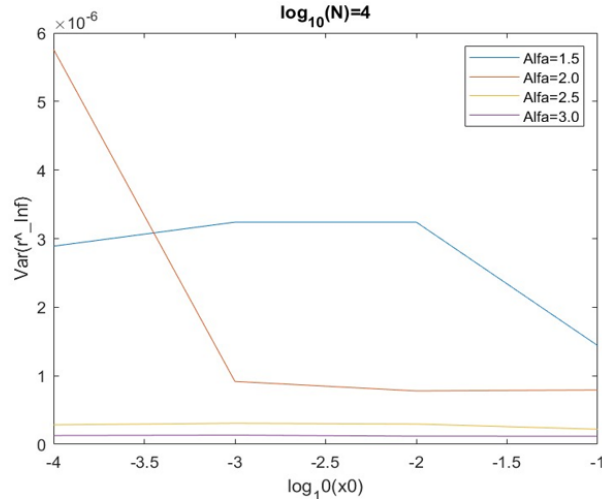


Figure 5: Crece y decrece.

Formas complejas de las gráficas: Si las gráficas muestran formas complejas, como puntos críticos, cambios en la pendiente o cambios en la dirección de las líneas, esto indica que la variabilidad de los resultados del modelo es sensible a las interacciones entre x_0 y α . Estas formas complejas pueden ser indicativas de comportamientos no lineales o de transiciones entre diferentes estados del sistema.

10 Conclusiones

Es un trabajo complejo donde se ve todo lo visto a lo largo de la carrera y se junta en un proyecto de simulación aunque notemos que tomamos uno de los modelos más sencillos de epidemiología, es todo un desafío en muchos aspectos pues se hizo lo siguiente, presentamos estos trabajos en equipos, y cada uno de los equipos un tema distinto, y cada uno de esos temas un grado de complejidad significativo, para luego exponerlo en clase y que los otros equipos entiendan el tema expuesto, necesitamos comprender la naturaleza del proceso que queremos modelar, sino no llegaremos a ninguna parte, para el caso del modelo SIR se puede decir que se logró capturar la escena de dicho modelo.