



Análisis de sensibilidad global en el modelo SIR aplicado en modelado de dispersión de VIH.

Dr. Prieto Moreno Kernel Enrique, Zamora Alemán Mario.
 Universidad Nacional Autónoma de México.
 Instituto de Matemáticas, Unidad Cuernavaca.

Resumen

Desde tiempos inmemorables las epidemias han sido participe de hecatombes en la historia de la humanidad, ahora con las recientes aplicaciones de los modelos matemáticos aplicados a la biología de sistemas, se sienta un nuevo paradigma para el estudio de las epidemias, esto desde un punto de vista de las ecuaciones diferenciales. A través del análisis de sensibilidad global podemos obtener información de utilidad para saber que tan robusto es el modelo que relaciona los valores biológicos que se ingresan comparados con los resultados, cuando hay variaciones en los parámetros del modelo. Dentro de los sistemas biológicos es común usar el modelo *SIR* (*Susceptibles, Infectious, Recovered*) para epidemias, es un modelo matemático que nos permite estudiar el comportamiento del VIH de allí que se puede implementar un análisis de sensibilidad global para los parámetros de este modelo.

Índice

Objetivo general y específicos:	2
Hipótesis de investigación:	2
Introducción:	2
Modelo SIR.	2
Método de Sobol para análisis de sensibilidad.	4
Método de Morris para análisis de sensibilidad.	4
Avances logrados a lo largo del verano:	5



Referencias

5

Objetivo general y específicos:

- Se plantea hacer un estudio bibliográfico sobre los modelos de sensibilidad local y global .
- Se aplicará un modelo de sensibilidad global para el modelo SIR aplicado al análisis del VIH.
- Estudiar los patrones de dispersión del VIH con el modelo SIR.
- Se identificarán los parámetros óptimos α y β para una distribución de probabilidad Beta.
- Se buscará aplicar un modelo de estadística inferencial bayesiano a través de la distribución Beta.

Hipótesis de investigación:

En el estudio de los patrones de dispersión del VIH el modelo SIR nos da una aproximación de como se comporta el virus, pero en el estudio empírico de los datos lo que se plantea es que la incertidumbre del modelo SIR para VIH se puede estudiar desde el punto de vista de la inferencia estadística con una distribución beta al hacer un estudio de análisis de sensibilidad global de los parámetros.

Introducción:

Entender como funciona una epidemia es de gran importancia desde la antigüedad, las enfermedades infecciosas que emergen en recientes años tales como el ébola y el SARS hacen que ahora sea de importancia el estudiar como funcionan los sistemas biológicos desde un punto de vista matemático, lo cual ha conllevado a que la bibliografía acerca del estudio de la dinámica de las enfermedades infecciosas sea basta, como lo menciona. En los últimos años se le ha dado principalmente la atención al modelo matemático concerniente a la infección por VIH para hacer la investigación acerca de cómo evoluciona el virus. (Capasso, 1993)

Conforme se estudian estas enfermedades surgen preguntas como lo son: ¿Cuál es el número máximo de personas que se pueden infectar? ¿Cuántos individuos han sobrevivido? ¿Cuántos individuos están infectados? entre otras, de allí que los parámetros en los modelos empíricos se deben determinar al hacer los experimentos, pero los datos experimentales contienen variabilidades o suelen estar incompletos para poder describir un fenómeno de allí que el hacer diagnósticos para encontrar la significancia y determinabilidad de los parámetros sea necesaria para asegurarnos que el modelo propuesto respalda a los datos experimentales obtenidos.(Jaqaman y Danuser, 2006)

Modelo SIR.

De acuerdo a (Capasso, 1993) el modelo SIR lo plantea Kermack y McKendrick en *Contributions to the mathematical theory of epidemics*, en este modelo una población de N individuos se divide en tres clases:

(S) la clase de los susceptibles, es decir aquellos individuos propensos a contraer una enfermedad y así a estar infectados.



(I) la clase de los infectados, es decir aquellos individuos capaces de transmitir la enfermedad a los susceptibles.

(R) la clase de los que se remueven, es decir aquellos individuos que habiendo contraído la enfermedad han muerto o si se han recuperado, son inmunes o han sido puestos en cuarentena y así ya no pueden transmitir la enfermedad.

De aquí para poder estudiarse desde el punto de vista de la dinámica del proceso de infección se usan las ecuaciones diferenciales para entender las tasas de transferencia de una clase de individuos a otra:

$$\begin{aligned}
 \frac{dS}{dt} &= f_1(I, S, R) \\
 \frac{dI}{dt} &= f_2(I, S, R) \\
 \frac{dR}{dt} &= f_3(I, S, R)
 \end{aligned} \tag{1}$$

Que para esta investigación se utiliza la siguiente variante, (Smith, 2013):

$$\begin{aligned}
 \frac{dS}{dt} &= \delta N - \delta S - \gamma \kappa IS, \quad S(0) = S_0, \\
 \frac{dI}{dt} &= \gamma \kappa IS - (r + \delta)I, \quad I(0) = I_0, \\
 \frac{dR}{dt} &= rI - \delta R, \quad R(0) = r_0,
 \end{aligned} \tag{2}$$

Donde $S(t)$, $I(t)$, y $R(t)$, son el número de susceptibles, infectados y removidos al tiempo t en la población de N individuos, $N = S_0 + I_0 + R_0$. Aquí los coeficientes γ , κ , y r denotan respectivamente el coeficiente de infección, el de interacción, que nos indica cuál es la probabilidad que un individuo entre en contacto con otros, y la tasa de recuperación.

Para este modelo tenemos que los parámetros son $q = [\gamma, \kappa, r, \delta]$, en el modelo SIR los parámetros γ y κ tienen una repercusión en la dinámica de la enfermedad ya que γ es una consecuencia de la enfermedad mientras que κ refleja el grado de interacción de los individuos. De lo cual tenemos que el coeficiente γ es de difícil control mientras que para κ se puede controlar a través de aislamiento o cuarentenas. La tasa de nacimientos y decesos (δ) para este modelo es la misma.

Tomaremos la distribución de parámetros de la siguiente forma:

$$\gamma \sim \mathcal{U}(0, 1), \quad \kappa \sim \text{Beta}(\alpha, \beta), \quad r \sim \mathcal{U}(0, 1), \quad \delta \sim \mathcal{U}(0, 1) \tag{3}$$

Es decir γ , r , δ se toman como distribuciones uniformes con soporte en el intervalo $(0, 1)$ y κ se toma como una distribución Beta con parámetros α y β . La distribución $\text{Beta}(\alpha, \beta)$ la definiremos como $f_X(x; \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$, $x \in [0, 1]$ con $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$, $\alpha > 0$. El que los coeficientes γ , r , δ sigan una distribución uniforme refleja el hecho de que se tiene un conocimiento limitado a priori de los parámetros.

Uno de los objetivos del análisis de sensibilidad global es escudriñar cuanta incertidumbre se tiene en la relación que hay en los datos con respecto a las predicciones del



modelo al momento de variar los parámetros. Este análisis se centra en el modelo en sí y no en los datos, la importancia de hacer el análisis de sensibilidad para los parámetros está en que se puede reducir el número de parámetros en el modelo matemático haciéndolo más simple si al momento de hacer el análisis nos da que el quitar ciertos parámetros no afecta en el estudio realizado. (Zi, 2011)

Método de Sobol para análisis de sensibilidad.

De acuerdo a (Zi, 2011) el método Sobol es del tipo basado en la varianza de los datos y no hace ninguna suposición entre los datos de entrada y salida del modelo. La base de este modelo es la descomposición de la varianza de la función $f(x)$ de resultados en sumandos de varianzas en combinaciones de los parámetros de entrada en dimensionalidad creciente

$$f(x) = \sum_{i=1}^k f_i(x_i) + \sum_{i=1}^k \sum_{j=i+1}^k f_{ij}(x_i, x_j) + \dots + f_{1\dots k}(x_1, x_2, \dots, x_k) \quad (4)$$

Así la varianza total D se define como:

$$\begin{aligned}
 D &= \int_{\Omega^k} f^2(x) dx - f(0)^2 \\
 &= \int_{\Omega^k} f^2(x) dx - \left(\int_{\Omega^k} f(x) dx \right)^2 \quad (5)
 \end{aligned}$$

Las varianzas parciales se calculan para cada elemento de la ecuación (4) y se definen como

$$D_{i_1 i_2 \dots i_s} = \int \dots \int f^2(x_{i_1}, x_{i_2}, \dots, x_{i_s}) dx_{i_1} dx_{i_2} \dots dx_{i_s}$$

Los índices de sensibilidad global de Sobol se calculan como

$$S_{i_1 i_2 \dots i_s} = \frac{D_{i_1 i_2 \dots i_s}}{D}$$

Estos terminos $S_{i_1 i_2 \dots i_s}$ nos dan la proporción de la varianza total que es asignada a un parámetro individual o a una combinación de ellos, por ejemplo a $S_i = \frac{D_i}{D}$ se le denomina índice de sensibilidad de primer orden

Método de Morris para análisis de sensibilidad.

El método Morris es un análisis de sensibilidad de detección que se basa en un efecto elemental (Zi, 2011). El efecto elemental se calcula cambiando un parámetro a la vez. El parámetro se muestrea p veces de manera discretas de una distribución predefinida. El análisis de sensibilidad de un modelo de salida para k parámetros requieren una matriz $k \times p$. Ahora considerando el análisis de sensibilidad de un modelo de salida y para los k parámetros del modelo, el parámetro i -ésimo del modelo se puede escalar en el intervalo $[0, 1]$ y puede tomar valores de $\{0, 1/(p-1), 2/(p-1), \dots, 1\}$. Los efectos elementales del i -ésimo parámetro x_i se definen como



$$d(x_i) = \frac{y(x_1, x_2, x_{i-1}, x_i + \Delta, x_{i+1}, \dots, x_k) - y(x)}{\Delta}$$

donde x es el conjunto de parámetros de referencia $\theta = [x_1, x_2, x_{k-1}, x_k]$, $x_i \leq 1 - \Delta$ y Δ es un múltiplo predefinido de $\frac{1}{p-1}$

El promedio de los efectos elementales cuantifica la importancia de los parámetros para la salida del modelo, mientras que la desviación estándar de los efectos elementales indica el efecto no lineal de los parámetros del modelo. Un valor grande en la media de efectos elementales implica que el parámetro modelo correspondiente tiene una importante influencia en la salida del modelo. Una desviación estándar alta de los efectos elementales indican que el parámetro es correlacionado con otros parámetros o el parámetro no tiene efectos lineales en la salida. El método Morris tiene un bajo costo computacional, que es apropiado para estudiar grandes modelos a escala.

Avances logrados a lo largo del verano:

Uno de los avances fue la búsqueda de la bibliografía para la elaboración de un trabajo posterior (tesis) para la licenciatura en matemáticas, además de elaborar los códigos en matlab y maple para los métodos numéricos asociados a encontrar los índices de Sobol y Morris.

En el transcurso del verano de investigación se se lograron recrear las tablas de análisis global para los índices de Sobol y Morris que se muestran a continuación para el valor $\kappa \sim Beta(2, 7)$

	γ	κ	r	δ
S_i	.0997	0.0312	0.7901	0.1750
S_{t_i}	-.0637	-0.0541	0.5634	0.2029
$\mu_i^*(\times 10^3)$	0.2532	0.2812	2.0184	1.2328
$\delta_i(\times 10^3)$	0.9539	1.6245	6.6748	3.9886

Agradecimientos: Este trabajo fue posible gracias al apoyo brindado por el Instituto de Matemáticas Unidad Cuernavaca UNAM.

Referencias

Capasso, V. (1993). *Mathematical structures of epidemic systems* (Vol. 88). Italia: Springer.

Jaqaman, K., y Danuser, G. (2006). Linking data to models: data regression. *Nature Reviews Molecular Cell Biology*, 7(11), 813.

Miao, H., Xia, X., Perelson, A. S., y Wu, H. (2011). On identifiability of nonlinear ode models and applications in viral dynamics. *SIAM review*, 53(1), 3-39.



- Shonkwiler, R. W., y Herod, J. (2009). *Mathematical biology: an introduction with maple and matlab*. EE.UU: Springer Science & Business Media.
- Smith, R. C. (2013). *Uncertainty quantification: theory, implementation, and applications* (Vol. 12). EE.UU: Siam.
- Zi, Z. (2011). Sensitivity analysis approaches applied to systems biology models. *IET systems biology*, 5(6), 336–346.