

Un intervalo de confianza robusto para el número de reproducción instantáneo, cómo lograrlo y por qué es necesario – Implementación mediante software: Apéndices

Contenido:

A. Suavización de series de incidencias.....	1
B. Distribución de la razón de dos variables aleatorias independientes distribuidas normalmente.....	8
C. Cálculo y aplicación de la función de densidad de probabilidad de R_t^*	9
D. Análisis de la fórmula de cálculo del R_t por el método de Cori et al. (2013) y su semejanza con R_t^m	12
E. Solución de un problema hipotético que sirve de modelo para calcular la desviación estándar, y el intervalo de confianza del R_t^m con incertidumbre en el intervalo serial.....	15
F. Implementación con MSExcel del método para el cálculo del R_t^m y aplicación	20
en una epidemia real: “La Pandemia COVID-19 de Panamá”	

A. Suavización de series de incidencias

Este apéndice es una ampliación del tema desarrollado en la sección de *Métodos* del artículo. En la Figura A.1 se vuelve a presentar el proceso hipotético de reconstrucción de la curva ideal de I_t^μ , que se vio en la Figura 3 del artículo. En una situación real solo se tendría una de las series de tiempo de incidencia I_t , por lo tanto, ya no será posible obtener una precisión progresiva que se acerque a I_t^μ , a medida que se adicionan series de incidencia I_t . Las series de tiempo de las epidemias solo tienen una observación por cada instante de conteo t . Si se tratara de conteos desvinculados, un solo valor de conteo en cada instante no serviría para desarrollar un análisis. Sin embargo, tratándose de una serie de tiempo, se puede aprovechar la proximidad de los demás conteos, hechos alrededor del instante t , para sustentar (o rechazar) la validez estadística de la relación entre los puntos próximos. Esta es la justificación de los métodos de suavización descritos en la teoría de *análisis de serie de tiempo*.

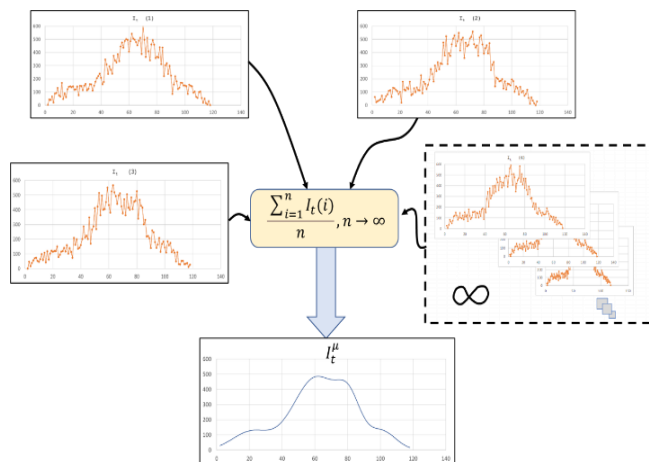


Figura A.1. Proceso de generación de series de incidencias aleatorias.

La hipótesis de suavidad permite suponer que los puntos de la curva se pueden organizar en sectores traslapados (Figura A.2), en donde los puntos de cada sector se pueden modelizar con funciones simples, tales como rectas, polinomios, exponenciales, logaritmos, etc. Trabajando con los puntos correspondientes a cada grupo, se pueden reasignar los valores centrales de cada grupo. Así, en la Figura A.2 el valor del punto en t se puede recalcular con base en los puntos del grupo A, y el valor del punto en $(t+1)$ se puede recalcular con base a los puntos del grupo B. Este mecanismo garantiza la suavización de la curva debido a que los grupos traslapados comparten todos los puntos salvo un punto. Así los cálculos consecutivos correspondientes no tendrán diferencias abruptas, se suavizan.

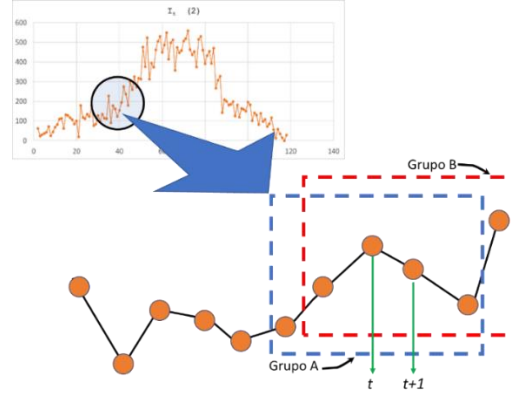


Figura A.2. Suavización de series mediante la reasignación del punto central de cada grupo de puntos traslapados.

Dentro de la amplia gama de funciones, hemos escogido a las rectas por su fácil *implementación*, por la simple *interpretación* y la suficiente *precisión de los resultados* relativos a otras funciones.

Método de media móvil, suavización con líneas rectas:

A continuación, se presenta la ecuación general que se usará para el cálculo de las medias Im_t de la serie de incidencias. Se trata de una media móvil *sin retraso*, con un ajuste especial al principio y al final de la serie para maximizar la cantidad de datos de incidencias empleados. N es el número total de puntos de la serie, n_k es el *tiempo de media móvil* convencional constante definido por el analista; y n_t es un tiempo de media móvil dinámico, que cambia al principio y al final de la serie, calculado para el instante t de observación. Así:

$$Si \quad 1 < t < \frac{n_k + 1}{2} :$$

$$Im_t = \frac{1}{n_t} \sum_{j=1}^{n_t} I_j, \quad n_t = 2(t-1) + 1$$

$$Si \quad \frac{n_k + 1}{2} \leq t \leq N - \frac{n_k + 1}{2} + 1 :$$

$$n_t = n_k$$

$$Im_t = \begin{cases} \frac{1}{n_t} \sum_{i=t-\frac{n_t-1}{2}}^{t+\frac{n_t-1}{2}} I_i & n_t \text{ impar} \\ \frac{\sum_{i=t-\frac{n_t}{2}}^{t+\frac{n_t}{2}-1} I_i + \sum_{i=t-\frac{n_t}{2}+1}^{t+\frac{n_t}{2}} I_i}{2n_t} & n_t \text{ par} \end{cases}$$

$$Si \quad N - \frac{n_k + 1}{2} + 1 < t < N :$$

$$Im_t = \frac{1}{n_t} \sum_{j=1}^{n_t} I_{N-j+1}, \quad n_t = 2(N-t) + 1 \quad (A.5)$$

El valor de n_k asignado no ha de ser ni muy pequeño que haga imprecisa a la estimación de I_t^μ , ni tan grande que no permita asegurar la aleatoriedad de **signo** de los errores.

¿Qué hacer cuando la magnitud de los errores crece conforme crece I_t^μ ?

La escogencia del valor de n_k afecta la precisión de la estimación de la media ideal I_t^μ con Im_t . Esta precisión también depende de la variabilidad real del error alrededor de I_t^μ , cuya desviación estándar se denota σ_t . En (A.9) descomponemos a la incidencia en sus partes de media y error:

$$I_t = I_t^\mu + e_t, \quad e_t \sim N(0, \sigma_t^2) \quad (\text{A.9})$$

En la Figura A.5 se muestra el diagrama de dispersión del error vs el valor del estimador Im_t de una de las series de tiempo presentadas en la Figura A.1. Se nota que el contorno superior de la serie crece conforme crece el valor de Im_t . Aunque esta es una serie de simulación, la figura ilustra una situación típica de la realidad. Evidentemente, el error no es homocedástico.

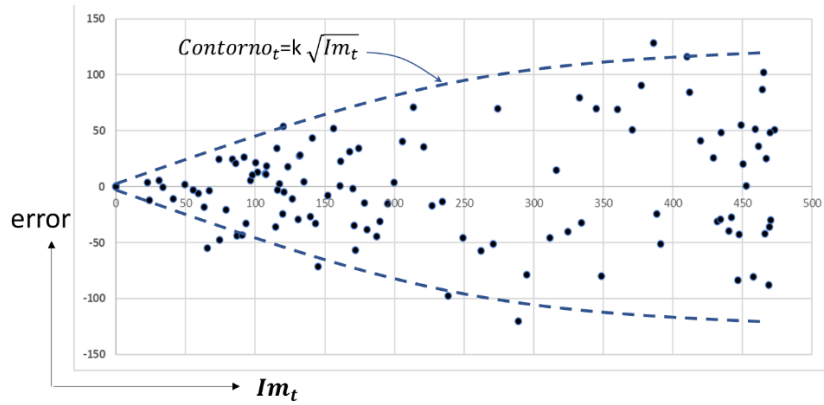


Figura A.5. Diagrama de dispersión del error vs Im_t .

Observar el contorno del diagrama de dispersión es un procedimiento estándar en el análisis de series de tiempo. Después de esto, simplemente se esboza una ecuación general que permita reproducir este contorno. Para el caso estudiado (Figura A.5), la ecuación general con la que se ensayará es:

$$\text{Contorno} = k \times \text{Im}_t^\gamma \quad (\text{A.10})$$

En (A.10) el factor de interés es Im_t^γ (función *potencia*), en el cual desconocemos cuál es el exponente γ más apropiado. Para definirlo, la práctica usual es dividir a los errores e_t entre este factor (A.11), y ensayar con distintos valores de γ , hasta que el diagrama de dispersión del *error transformado* e_t^* vs Im_t forme una banda horizontal (para hacerse homocedástico respecto a Im_t). Todos los valores e_t^* se asumen que se distribuyen normalmente en cada instante t , con una desviación estándar denotada por σ_t^* . Así:

$$e_t^* = \frac{e_t}{\text{Im}_t^\gamma} \sim N(0, \sigma_t^*) \quad (\text{A.11})$$

En la Figura A.6 se muestran ejemplos de los diagramas de dispersión para distintos valores de γ .

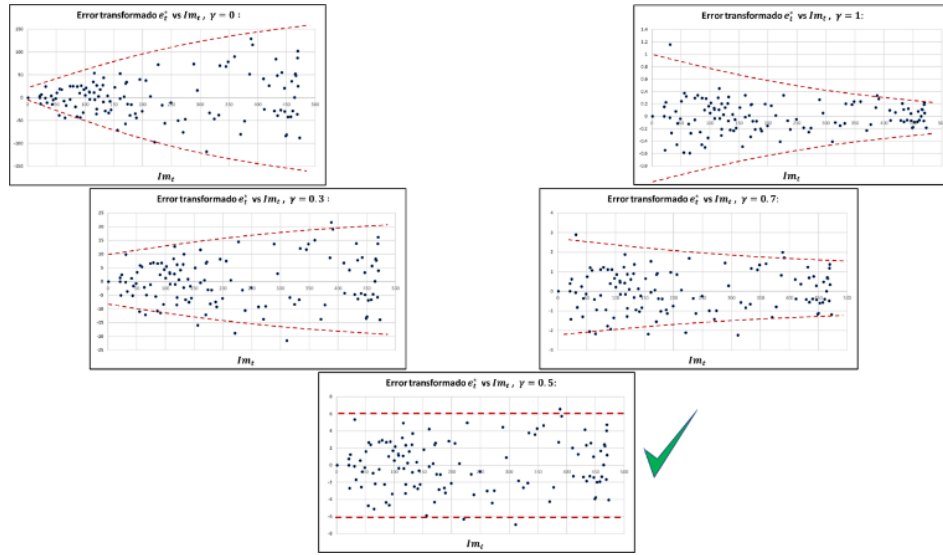


Figura A.6. Diagramas de dispersión del error trasformado vs Im_t para $\gamma = 0, 0.3, 0.5, 0.7, 1$.

Se descubre que para $\gamma=0.5$, el diagrama de dispersión adopta la forma deseada de banda horizontal. Con este valor de γ , el divisor de la función de transformación (A.11) no es más que la raíz cuadrada de Im_t . Lo cual era de esperar, porque este ejemplo procede de un caso simulado (ecuaciones (2) a (6) del artículo).

La forma del contorno del error, en los procesos de conteo como los de una epidemia, no siempre se podrá modelizar con una simple función *potencia*. Existen otras funciones con las cuales habrá que ensayar; tenemos, por ejemplo, las funciones logaritmo, seno (parte creciente positiva), tangente hiperbólica, logística, etc. Todas estas funciones comparten la cualidad de que son crecientes (primera derivada positiva) en el rango de su aplicación. En esta propuesta hemos seleccionado a las funciones *potencia*, *logística*, y una suma ponderada especial de estas, que hemos llamado *Logipot*. Además, tenemos a la función *Logist* (una variante de la función *logística* que estamos proponiendo), que es particularmente útil porque se puede modelizar para usar a discreción sus distintas secciones (ver Figura A.7).

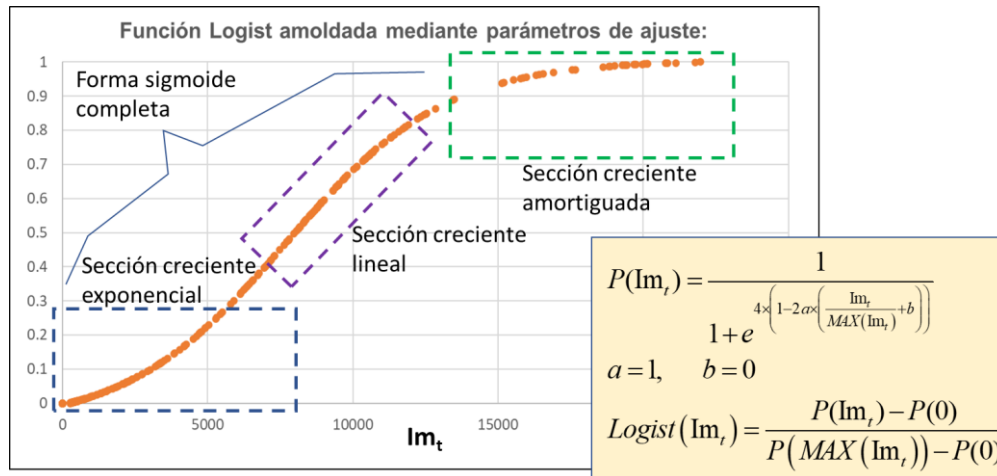


Figura A.7. Ejemplo de una función *Logist* amoldada mediante parámetros, y secciones de modelización.

La función de transformación *Logist* tiene 4 secciones crecientes claramente identificadas: exponencial, lineal, amortiguada, y la forma sigmoide completa. Si esta función se mezcla mediante una suma ponderada con la función *potencia*, se obtiene la función *Logipot* que mencionamos (Figura A.8).

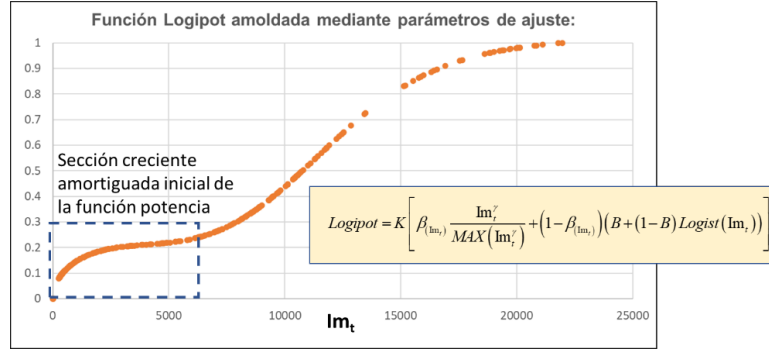


Figura A.8. Función Logipot ajustada mediante parámetros, y secciones de modelización.

En este trabajo se propone a la función *Logipot* debido a que la plasticidad de su diseño permite amoldar la forma del contorno del error en función de la media (Im_t), mediante la escogencia apropiada de parámetros. A continuación se hace la descripción de las funciones y los parámetros que se utilizan en la transformación del error.

$$P(Im_t) = \frac{1}{1 + e^{4 \times \left(1 - 2a \times \left(\frac{Im_t}{MAX(Im_t)} + b \right) \right)}} \quad (A.12)$$

Esta es la **función logística básica**, donde:

a (Factor de la pendiente): es un parámetro que afecta notoriamente a la pendiente en el punto de inflexión. Cuando su valor sea menor que 1, la pendiente disminuye; cuando sea mayor que 1, aumenta. Escoger un valor tal que $a > 0$.

b (Corrimiento horizontal): es un parámetro que permite ajustar el corrimiento a la izquierda ($b > 0$) o a la derecha ($b < 0$) de la curva de la función respecto al eje de Im_t . Este factor está definido en el rango $-1 < b < 1$.

$MAX(Im_t)$: es el valor máximo que adopta la curva de Im_t . El parámetro b se puede interpretar como una proporción de $MAX(Im_t)$.

$$Logist(Im_t) = \frac{P(Im_t) - P(0)}{P(MAX(Im_t)) - P(0)} \quad (A.13)$$

Esta función estandariza a la función logística básica. Su valor varía entre 0 y 1, cuando $0 \leq Im_t \leq MAX(Im_t)$.

$$Logipot = K \left[\beta_{(Im_t)} \frac{Im_t^\gamma}{MAX(Im_t)^\gamma} + (1 - \beta_{(Im_t)}) (B + (1 - B) Logist(Im_t)) \right] \quad (A.14)$$

Suma ponderada de las funciones *Logist* (A.13) y *potencia*. Su valor varía entre 0 y K cuando $0 \leq Im_t \leq MAX(Im_t)$.

K (Escala común): Es un factor de escala siempre positivo, que no afecta la forma, ni a los cálculos subsecuentes, pero que facilita el ajuste de los parámetros, mediante el apoyo visual en la computadora.

γ (Exponente): Exponente de la función *potencia*. Se recomiendan valores entre 0.5 y 1.

$\beta_{(Im_t)}$: Es un factor de ponderación dinámico, dependiente de Im_t , que determina las proporciones en que efectivamente se sumarán las funciones *potencia* y *Logist*. Se calcula según (A.15).

B (Base): Es un parámetro para el ajuste del nivel base sobre el que varía la función *Logist*. Solo puede tomar valores entre 0 y 1.

$$\beta_{(Im_t)} = \frac{\alpha}{\alpha + (1 - \alpha) \text{Logist}(Im_t)} \Big|_{a=1, b=BalSec} \quad (A.15)$$

Esta función actúa como proporción efectiva para la suma ponderada de las funciones *potencia* y *Logist*. El valor de α (Balance principal) representa la proporción ideal asignada por el analista a la función *potencia*, y $(1 - \alpha)$ es la proporción ideal asignada a la función *Logist*. Sin embargo, para fines de un mejor ajuste de la curva de transformación cuando Im_t toma valores pequeños, se está proponiendo sustituir al valor de α por $\beta_{(Im_t)}$. Así, cuando $Im_t \rightarrow 0$, $\beta_{(Im_t)} \rightarrow 1$; y cuando $Im_t \rightarrow MAX(Im_t)$, $\beta_{(Im_t)} \rightarrow \alpha$.

La escogencia del valor de α se puede hacer en el rango $0 < \alpha < 1$, y no debe estrañar que sea necesario escoger valores en el orden de 10^{-3} debido a la no linealidad de (A.15).

BalSec (Balance secundario): Es un ajuste fino al balance de la ponderación. Refuerza a la razón de crecimiento $\beta_{(Im_t)}$ cuando $Im_t \rightarrow 0$. BalSec > 0 , preferiblemente 0.

Estimación de la desviación estándar σ_t :

Ahora retornamos nuestra atención a (A.11), para el cálculo de la desviación estándar.

$$e_t^* = \frac{e_t}{Im_t^\gamma} \sim N(0, \sigma_t^*) \quad (A.16)$$

Ya que Im_t es prácticamente constante (muy poca variabilidad) respecto a la variabilidad de e_t (A.16) la desviación estándar esperada de e_t^* se puede aproximar:

$$\sigma_t^* = \frac{\sigma_t}{(I_t^\mu)^\gamma} \quad (A.17)$$

Resolviendo por σ_t :

$$\sigma_t = \sigma_t^* (I_t^\mu)^\gamma \quad (A.18)$$

Para lograr más generalidad, sustituiremos a la función *potencia* por la función de transformación *Logipot*:

$$\sigma_t = \sigma_t^* \text{Logipot}(I_t^\mu) \quad (A.19)$$

Se proponen dos métodos para estimar la desviación estándar σ_t^* :

1. Asumiendo que σ_t^* es constante por tramos de tiempo: Para procesos de conteo bajo control dentro de cada tramo.
2. Asumiendo que σ_t^* varía continuamente con el tiempo: Para procesos de conteo que mejoran (o empeoran) en forma continua, sin que se puedan especificar tramos de control.

Estimación de la desviación estándar σ_t^* asumiendo que es constante por tramos de tiempo:

Sean M tramos de la serie de e_t^* . Cada tramo de tamaño N_k (con $k=1,2,3,\dots,M$). Y sea t_k el tiempo inicial de cada tramo.

Media de rango móvil del tramo k :

$$\overline{RMe}_k^* = \frac{1}{N_k - 1} \sum_{t=t_k+1}^{t_k+N_k} |e_t^* - e_{t-1}^*|, \quad (A.20)$$

La desviación estándar estimada $\hat{\sigma}_k^*$ de cada tramo se calcula dividiendo a la media del rango móvil \overline{RMe}_k^* entre un factor de conversión $d_2 = 1.128$.

$$\hat{\sigma}_k^* = \left(\frac{\overline{RMe}_k^*}{d_2} \right) = \left(\frac{\overline{RMe}_k^*}{1.128} \right) \quad (A.21)$$

Dentro de cada tramo $\hat{\sigma}_t^*$ tendrá un valor constante de $\hat{\sigma}_k^*$ dado por (A.21).

Es claro que este método puede aplicarse al gran tramo de todo el tiempo de la epidemia.

El cálculo de la desviación estándar mediante rango móvil (A.21) es el método más preciso cuando solo se cuenta con un valor en cada instante de medición (Montgomery, 2001).

Estimación de la desviación estándar σ_t^* asumiendo que puede variar continuamente con el tiempo:

Sea la serie de e_t^* de tamaño N .

Se calculan todos los rangos móviles \overline{RMe}_k^* :

$$RMe_t^* = |e_t^* - e_{t-1}^*| \quad (A.22)$$

Tomando la media móvil \overline{RMe}_k^* de estos rangos, con tiempo de media móvil n , para cada instante t . La desviación estándar estimada $\hat{\sigma}_t^*$ se calcula dividiendo a la media móvil \overline{RMe}_k^* entre un factor de conversión $d_2 = 1.128$.

$$\hat{\sigma}_t^* = \left(\frac{\overline{RMe}_t^*}{d_2} \right) = \left(\frac{\overline{RMe}_t^*}{1.128} \right) \quad (A.23)$$

Finalmente, el trabajo para calcular el valor estimado de σ_t , se resume en calcular $\hat{\sigma}_t^*$ y luego aplicar (A.19) para cada instante t , con valores estimados:

$$\hat{\sigma}_t = \hat{\sigma}_t^* \text{Logipot}(\text{Im}_t) \quad (A.24)$$

Aclaración:

*Todo el ejercicio para encontrar una forma de modelizar el contorno del diagrama de dispersión del error vs Im_t (Figura A.5 y Figura 6) es importante para lograr la condición ideal de un error transformado normalizado y, consiguientemente, maximizar la precisión de los cálculos de σ_t , y otros cálculos derivados, como los intervalos de confianza. Sin embargo, gracias a las consecuencias de Teorema Central del Límite (TCL), la condición de **normalidad**, aunque deseable, no es un requisito “sine qua non” para los cálculos basados en medias, como los intervalos de confianza. La prueba de esto no la hemos aportado aquí, pero el lector interesado podrá constatarlo por su cuenta, simplemente ensayando con distintas curvas de modelización de contorno.*

B. Distribución de la razón de dos variables aleatorias independientes distribuidas normalmente

En este apéndice se presenta una función general de densidad de probabilidad que servirá como herramienta para el cálculo de la función de densidad de probabilidad de R_t^* , en el Apéndice C.

Sean las variables aleatorias independientes X y Y distribuidas según $N(\mu_x, \sigma_x^2)$ y $N(\mu_y, \sigma_y^2)$ respectivamente. Entonces, la razón $W = \frac{Y}{X}$ tiene una distribución dada por:

$$f_W(w) = \int_{-\infty}^{\infty} |x| f_X(x) f_Y(wx) dx \quad (B.1)$$

(Larsen & Marx, 2012)

Sabiendo que las funciones de densidad de probabilidad de X y Y son:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}, \quad f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2} \quad (B.2)$$

Entonces:

$$\begin{aligned} f_W(w) &= \int_{-\infty}^{\infty} |x| \left[\frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2} \right] \left[\frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{1}{2}\left(\frac{wx-\mu_y}{\sigma_y}\right)^2} \right] dx \\ &= \left(\frac{1}{2\pi\sigma_x\sigma_y} \right) \int_{-\infty}^{\infty} |x| e^{-\left[\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \frac{1}{2}\left(\frac{wx-\mu_y}{\sigma_y}\right)^2 \right]} dx \end{aligned} \quad (B.3)$$

El procedimiento de solución de esta integral lo omitiremos aquí debido a su extensión. La solución es:

$$f(w) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{\sigma_y^2\mu_x^2 + \sigma_x^2\mu_y^2}{2\sigma_x^2\sigma_y^2}} \left(\frac{1}{\frac{\sigma_y^2 + \sigma_x^2 w^2}{2\sigma_x^2\sigma_y^2}} + \frac{\left(\frac{\sigma_y^2\mu_x + \sigma_x^2 w\mu_y}{\sigma_x^2\sigma_y^2} \right) \sqrt{\pi} e^{\frac{\left(\frac{\sigma_y^2\mu_x + \sigma_x^2 w\mu_y}{\sigma_x^2\sigma_y^2} \right)^2}{4\left(\frac{\sigma_y^2 + \sigma_x^2 w^2}{2\sigma_x^2\sigma_y^2} \right)}}}{2\left(\frac{\sigma_y^2 + \sigma_x^2 w^2}{2\sigma_x^2\sigma_y^2} \right)^{\frac{3}{2}}} \operatorname{erf} \left(\frac{\frac{\sigma_y^2\mu_x + \sigma_x^2 w\mu_y}{\sigma_x^2\sigma_y^2}}{2\sqrt{\frac{\sigma_y^2 + \sigma_x^2 w^2}{2\sigma_x^2\sigma_y^2}}} \right) \right) \quad (B.4)$$

Con $w=y/x$

La función $\operatorname{erf}(z)$ es una función especial de matemática estadística llamada *función de error*:

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$$

En la Figura B.1 mostramos algunas de las gráficas de las funciones de densidad de probabilidad que se pueden obtener con distintos parámetros de medias y desviaciones estándar para las variables x y y .

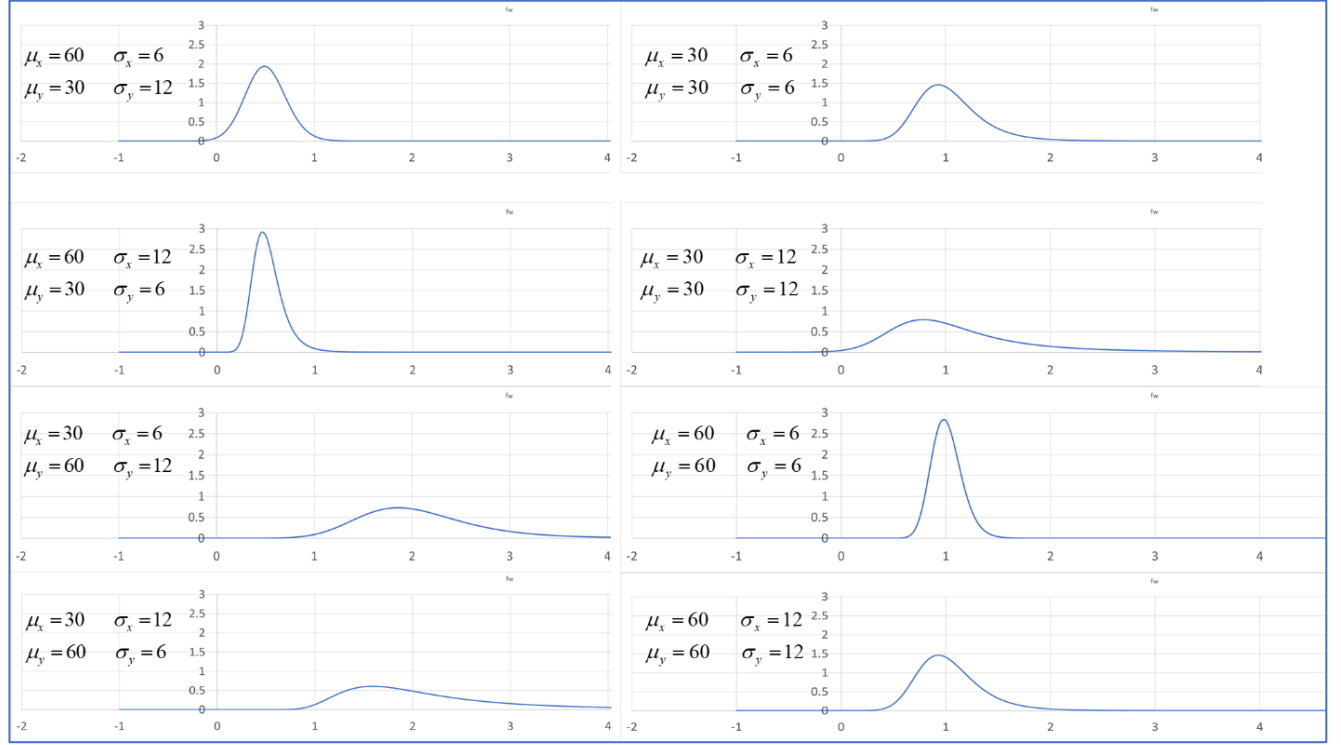


Figura B.1

C. Cálculo y aplicación de la función de densidad de probabilidad de R_t^*

En esta sección se encontrará la función de densidad de probabilidad de R_t^* necesaria como herramienta matemática de demostración para el cálculo de la media de R_t^* . En la ecuación (21) del artículo se definió a R_t^* :

$$R_t^* = \frac{I_t}{\sum_{s=1}^t w_{t,s} I_{t-s}} \quad (C.1)$$

Para empezar, se observa que el divisor $\sum_{s=1}^t w_{t,s} I_{t-s}$ es una variable aleatoria, de aquí en adelante llamada Λ_t , que se distribuye normalmente según:

$$\Lambda_t = \sum_{s=1}^t w_{t,s} I_{t-s} \sim N(\Lambda_t^\mu, \sigma_{\Lambda_t}^2) \quad (C.2)$$

Λ_t^μ se obtiene directamente:

$$\Lambda_t^\mu = \sum_{s=1}^t w_{t,s} I_{t-s}^\mu \quad (C.3)$$

Además, considerando que las incidencias I_{t-s} de (C.2) tienen sus medias y errores desacoplados, el cálculo de la varianza de Λ_t podrá hacerse sabiendo que las incidencias I_{t-s} son variables independientes. Así:

$$\sigma_{\Lambda_t}^2 = \sum_{s=1}^t w_{t,s}^2 \sigma_{t-s}^2 \quad (C.4)$$

Con lo anterior, y siendo R_t^* una razón (C.1), en la que el numerador I_t , y el denominador Λ_t son independientes, la función de densidad de probabilidad $f(R_t^*)$ se puede calcular. Veamos.

Sean las distribuciones normales de I_t y de Λ_t , $u(I_t/I_t^\mu, \sigma_t)$ y $v(\Lambda_t/\Lambda_t^\mu, \sigma_{\Lambda_t})$ respectivamente, entonces:

$$u(I_t/I_t^\mu, \sigma_t) = \frac{1}{\sqrt{2\pi}\sigma_t} e^{-\frac{1}{2}\left(\frac{I_t-I_t^\mu}{\sigma_t}\right)^2},$$

$$v(\Lambda_t/\Lambda_t^\mu, \sigma_{\Lambda_t}) = \frac{1}{\sqrt{2\pi}\sigma_{\Lambda_t}} e^{-\frac{1}{2}\left(\frac{\Lambda_t-\Lambda_t^\mu}{\sigma_{\Lambda_t}}\right)^2} \quad (C.5)$$

Así, la distribución de R_t^* se podrá calcular aplicando la solución general para el cálculo de la distribución de una razón (Y/X) que se explicó en el Apéndice B. En nuestro caso:

$$y = I_t, \quad x = \Lambda_t, \quad w = R_t^*, \quad f_X(x) = v(\Lambda_t), \quad f_Y(y) = u(I_t), \quad y \quad f_W(w) = f(R_t^*).$$

Reemplazando en (B.1):

$$f(R_t^*) = \int_{-\infty}^{\infty} |\Lambda_t| v(\Lambda_t) u(R_t^* \Lambda_t) d\Lambda_t \quad (C.6)$$

Para simplificar, se han suprimido los parámetros condicionales (.../...) de medias y desviaciones estándar. Aplicando la solución general presentada en el apéndice B:

$$f(R_t^*) = \frac{1}{2\pi\sigma_{\Lambda_t}\sigma_t} e^{-\frac{\sigma_t^2(\Lambda_t^\mu)^2 + \sigma_{\Lambda_t}^2(I_t^\mu)^2}{2\sigma_{\Lambda_t}^2\sigma_t^2}} \left[\frac{1}{\frac{\sigma_t^2 + \sigma_{\Lambda_t}^2(R_t^*)^2}{2\sigma_{\Lambda_t}^2\sigma_t^2}} + \frac{\left(\frac{\sigma_t^2\Lambda_t^\mu + \sigma_{\Lambda_t}^2R_t^*I_t^\mu}{\sigma_{\Lambda_t}^2\sigma_t^2}\right) \sqrt{\pi} e^{\frac{\left(\frac{\sigma_t^2\Lambda_t^\mu + \sigma_{\Lambda_t}^2R_t^*I_t^\mu}{\sigma_{\Lambda_t}^2\sigma_t^2}\right)^2}{4\left(\frac{\sigma_t^2 + \sigma_{\Lambda_t}^2(R_t^*)^2}{2\sigma_{\Lambda_t}^2\sigma_t^2}\right)}}}{2\left(\frac{\sigma_t^2 + \sigma_{\Lambda_t}^2(R_t^*)^2}{2\sigma_{\Lambda_t}^2\sigma_t^2}\right)^{\frac{3}{2}}} \operatorname{erf}\left(\frac{\frac{\sigma_t^2\Lambda_t^\mu + \sigma_{\Lambda_t}^2R_t^*I_t^\mu}{\sigma_{\Lambda_t}^2\sigma_t^2}}{2\sqrt{\frac{\sigma_t^2 + \sigma_{\Lambda_t}^2(R_t^*)^2}{2\sigma_{\Lambda_t}^2\sigma_t^2}}}\right) \right] \quad (C.7)$$

Con esta ecuación podrá calcularse el valor esperado y la *desviación estándar* de R_t^* , así como su *intervalo de predicción*. Aplicando esta función de densidad para calcular el valor esperado de R_t^* :

$$E(R_t^*) = \int_{-\infty}^{\infty} R_t^* f(R_t^*) dR_t^* \quad (C.8)$$

Es evidente que la solución de esta ecuación por medios analíticos es muy compleja, por lo que prescindiremos de este ejercicio para resolver numéricamente. Para cada instante t se aplicarán las medias I_t^μ y Λ_t^μ , y las desviaciones estándar σ_t y σ_{Λ_t} , resolviendo numéricamente la ecuación (C.8). Esto significa que, para conseguir la gráfica de la media en función del tiempo, habrá que aplicar este procedimiento para todos los N instantes t de la serie. Un trabajo que manualmente puede hacerse con apoyo computacional, pero que eventualmente ya no se querrá hacer más. Por esta razón procederemos a buscar una solución alternativa de aproximación, la que luego podremos comparar con la solución exacta (C.8).

La gráfica de $\sum_{s=1}^t w_{t,s} I_{t-s}$ es una curva suave, debido a que se trata de un promedio ponderado. A esta función la llamamos Λ_t en (C.2). Por tener poca variabilidad, ahora se asumirá que es una constante (no exhibe aleatoriedad en cada instante t), y para asegurar esta condición, se tomará su valor esperado, para aplicarlo en (C.1). Así:

$$R_t^* \approx \frac{I_t}{\sum_{s=1}^t w_{t,s} I_{t-s}^\mu} \quad (C.9)$$

De esta forma, abreviando cálculos, se encuentra que la función de densidad de probabilidad $f(R_t^*)$ ahora es mucho más simple:

$$f(R_t^*) = \frac{1}{\sqrt{2\pi}\sigma_{R_t^*}} e^{-\frac{1}{2}\left(\frac{R_t^* - R_t^\mu}{\sigma_{R_t^*}}\right)^2} \Rightarrow R_t^* \sim N(R_t^\mu, \sigma_{R_t^*}^2) \quad (C.10)$$

$$R_t^\mu = \frac{I_t^\mu}{\Lambda_t^\mu}, \quad \sigma_{R_t^*} = \frac{\sigma_t}{\Lambda_t^\mu}, \quad \Lambda_t^\mu = \sum_{s=1}^t w_{t,s} I_{t-s}^\mu \quad (C.11)$$

Los estimadores de las ecuaciones en (C.11) serán:

$$R_t^m = \frac{Im_t}{\Lambda_t^m}, \quad \hat{\sigma}_{R_t^*} = \frac{\hat{\sigma}_t}{\hat{\Lambda}_t^m}, \quad \hat{\Lambda}_t^m = \sum_{s=1}^t w_{t,s} Im_{t-s} \quad (C.12)$$

Sabiendo que Im_t y $\hat{\sigma}_t$ son valores que se pueden obtener (ver Apéndice A). Será labor del analista encontrar los parámetros de las distribuciones de R_t^* para cada instante t .

Ahora procederemos a hacer una prueba gráfica de la coincidencia de la aproximación normal (C.12) para el cálculo de R_t^m respecto al resultado exacto que se obtiene aplicando la (C.8).

En la Figura C.1 se muestra la superposición de las gráficas R_t^m con la aplicación del método exacto y el aproximado, usando los datos de la serie de la Figura 5 del artículo.

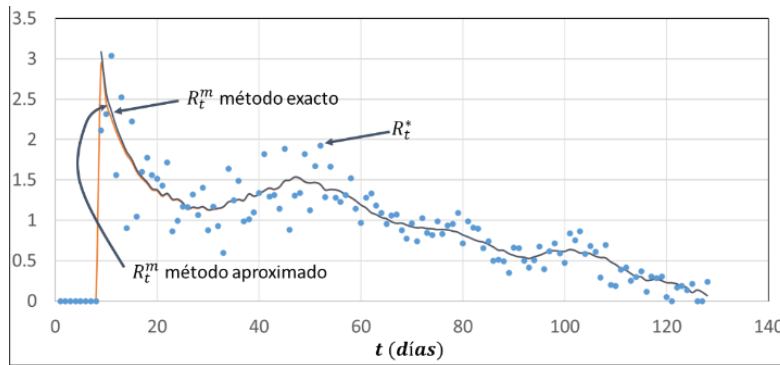


Figura C.1. Superposición de las curvas de calculadas por los métodos exacto y aproximado.

Como se observa (Figura C.1), la coincidencia es muy buena, a tal punto que no se puede distinguir con claridad cuál es cuál. Esto también prueba que la aproximación con la serie de Taylor (ecuación (29) del artículo), es correcta. Apenas se nota una diferencia muy leve cuando R_t^m tiene valores altos. No habiendo diferencias de importancia, es innegable que el método aproximado (C.12) es el mejor para calcular R_t^m , dada la sencillez de sus cálculos.

D. Análisis de la fórmula de cálculo del R_t por el método de Cori et al. (2013) y su semejanza con R_t^m

Sabiendo que la fórmula de cálculo del R_t por el método de Cori et al. (2013) es:

$$R_t = \frac{a + \sum_{s=t-\tau+1}^t I_s}{\frac{1}{b} + \sum_{s=t-\tau+1}^t \Lambda_s}, \quad \Lambda_s = \sum_h I_{s-h} w_h \quad (D.1)$$

Donde a y b son los parámetros a priori de la distribución gama.

Analizando la expresión de la sumatoria del denominador:

$$\sum_{s=t-\tau+1}^t \Lambda_s = \sum_{s=t-\tau+1}^t \left(\sum_h I_{s-h} w_h \right) \quad (D.2)$$

$$\begin{aligned} &= \sum_{s=t-\tau+1}^t (I_{s-1} w_1 + I_{s-2} w_2 + I_{s-3} w_3 + \dots) = \sum_{s=t-\tau+1}^t I_{s-1} w_1 + \sum_{s=t-\tau+1}^t I_{s-2} w_2 + \sum_{s=t-\tau+1}^t I_{s-3} w_3 + \dots \\ &= w_1 \sum_{s=t-\tau+1}^t I_{s-1} + w_2 \sum_{s=t-\tau+1}^t I_{s-2} + w_3 \sum_{s=t-\tau+1}^t I_{s-3} + \dots \end{aligned} \quad (D.3)$$

Esta ecuación se puede resumir como una sumatoria con índice k :

$$\sum_{s=t-\tau+1}^t \Lambda_s = \sum_{k=1}^{t-1} \left(w_k \sum_{s=t-\tau+1}^t I_{s-k} \right) \quad (D.4)$$

Sumando para todo valor de k dentro de los límites de la sumatoria en donde se cumpla que $s-k > 0$.

Para conocer otra forma de expresar esta ecuación, suponga que especificamos $t = 5$ y $\tau = 3$. Así, los términos de la sumatoria interna de (D.4) $(w_k \sum_{s=t-\tau+1}^t I_{s-k})$, para cada valor de k vienen a ser:

$$\begin{aligned} k=1 &\Rightarrow w_1 \sum_{s=3}^5 I_{s-1} = w_1 (I_2 + I_3 + I_4) \\ k=2 &\Rightarrow w_2 \sum_{s=3}^5 I_{s-2} = w_2 (I_1 + I_2 + I_3) \\ k=3 &\Rightarrow w_3 \sum_{s=3}^5 I_{s-3} = w_3 (X_0 + I_1 + I_2) \\ k=4 &\Rightarrow w_4 \sum_{s=3}^5 I_{s-4} = w_4 (X_{-1} + X_{-0} + I_1) \end{aligned} \quad (D.5)$$

Por inducción matemática, es fácil notar que las ecuaciones de (D.5) se pueden organizar considerando a las sumatorias que tienen τ variables I_s dentro del paréntesis ($k = 1, 2$), y considerando a las sumatorias que tienen menos de τ variables dentro del paréntesis ($k = 3, 4$). En resumen, los términos de la (D4) se pueden reagrupar como sigue:

$$\sum_{s=t-\tau+1}^t \Lambda_s = \sum_{k=1}^{t-\tau} \left(w_k \sum_{s=t-\tau+1}^t I_{s-k} \right) + \sum_{k=t-\tau+1}^{t-1} \left(w_k \sum_{s=1}^{t-k} I_s \right) \quad (D.6)$$

La primera sumatoria de miembro derecho de la (D.6) está relacionada a medias móviles, como veremos a continuación.

Sabiendo que las medias móviles, sin ajuste temporal hacia el centro, calculadas desde el instante $(t - \tau + 1)$ a t , están dadas por:

$$\text{Im}_{t-k}^{\text{sac}} = \frac{\sum_{s=t-\tau+1}^t I_{s-k}}{\tau}, \quad t > \tau, \quad \text{siempre que } s-k > 0 \quad (\text{D.7})$$

Las restricciones para t y $(s-k)$ son necesarias para garantizar las τ secuencias completas de incidencias para el cálculo de la media móvil. El superíndice “sac” indica que no se ha hecho el ajuste temporal hacia el centro del intervalo de la media móvil.

Rearreglando la ec.D.7:

$$\sum_{s=t-\tau+1}^t I_{s-k} = \tau \text{Im}_{t-k}^{\text{sac}}, \quad t > \tau, \quad \text{siempre que } s-k > 0 \quad (\text{D.8})$$

Reemplazando en (D.6):

$$\sum_{s=t-\tau+1}^t \Lambda_s = \tau \sum_{k=1}^{t-\tau} w_k \text{Im}_{t-k}^{\text{sac}} + \sum_{k=t-\tau+1}^{t-1} \left(w_k \sum_{s=1}^{t-k} I_s \right) \quad (\text{D.9})$$

Considerando ahora a la sumatoria del numerador de (D.1):

$$\sum_{s=t-\tau+1}^t I_s \quad (\text{D.10})$$

Tomando en cuenta nuevamente la expresión con las medias móviles (D.8), si hacemos $k=0$, encontramos:

$$\sum_{s=t-\tau+1}^t I_s = \tau \text{Im}_t^{\text{sac}} \quad (\text{D.11})$$

Reemplazando las ecuaciones (D.9) y (D.11), en (D.1):

$$R_t = \frac{a + \tau \text{Im}_t^{\text{sac}}}{\frac{1}{b} + \tau \sum_{k=1}^{t-\tau} w_k \text{Im}_{t-k}^{\text{sac}} + \underbrace{\sum_{k=t-\tau+1}^{t-1} \left(w_k \sum_{s=1}^{t-k} I_s \right)}} \quad (\text{D.12})$$

En el denominador de la ecuación anterior, la expresión marcada con una llave es insignificante para valores relativamente grandes de t . Esto se debe a que los pesos w_k adoptan valores pequeños para valores grandes de k . En la Figura D.1 presentamos nuevamente la gráfica del intervalo serial de la Figura 4 del artículo.

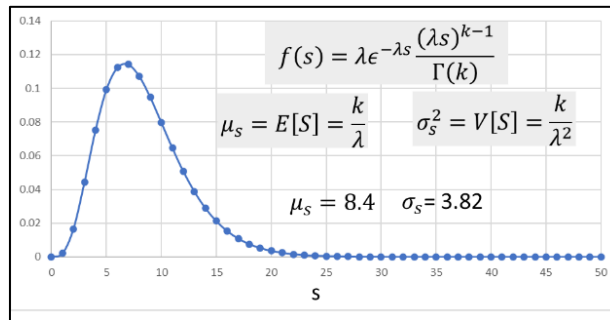


Figura D.1. Ejemplo típico de una gráfica del intervalo serial.

Ya que w_k adopta los valores de esta gráfica, para:

$$k > \mu_s + 2\sigma_s = 8.4 + 2(3.82) = 16, \quad w_k \approx 0$$

Así, (D.12) se puede expresar como:

$$R_t \approx \frac{a + \tau \text{Im}_t^{\text{sac}}}{\frac{1}{b} + \tau \sum_{k=1}^{t-\tau} w_k \text{Im}_{t-k}^{\text{sac}}}, \quad k > \mu_s + 2\sigma_s, \quad \text{ó} \quad t > \tau + \mu_s + 2\sigma_s \quad (\text{D.13})$$

Además, si consideramos los valores que típicamente se dan a los parámetros a y b del modelo de Cori et al. (2013), por ejemplo $a = 1$, y $b=5$, con un “length of time step” (para nuestro caso, el tiempo de media móvil) $\tau = 15$ días, (D.13) se puede simplificar aún más:

$$R_t \approx \frac{\tau \text{Im}_t^{\text{sac}}}{\tau \sum_{k=1}^{t-\tau} w_k \text{Im}_{t-k}^{\text{sac}}} = \frac{\text{Im}_t^{\text{sac}}}{\sum_{k=1}^{t-\tau} w_k \text{Im}_{t-k}^{\text{sac}}} \quad (\text{D.14})$$

Recordando la ecuación (29) del artículo para el cálculo del *número de reproducción instantáneo*:

$$R_t^m = \frac{\text{Im}_t}{\sum_{s=1}^t w_{t,s} \text{Im}_{t-s}} \quad (\text{D.15})$$

Es evidente la similitud entre (D.14) y (D.15), aunque se aprecian las siguientes diferencias:

1. Las medias móviles para el cálculo de R_t (D.14) se presentan sin ajuste temporal hacia el centro del intervalo de la media móvil (sac). Esto produce el efecto de desfase entre las curvas correspondientes de R_t y R_t^m . Ver Figura D.2.
2. El límite superior de la sumatoria del denominador de R_t (D.14) llega hasta $(t-\tau)$, mientras el límite superior de la sumatoria de R_t^m (D.15) llega hasta t . No obstante, esta diferencia no es importante para valores de $t > \mu_s + 2\sigma_s$, debido a la pequeñez de los pesos w_k .
3. Los pesos w_k de R_t (D.14) son independientes del tiempo; mientras que los pesos de $w_{t,s}$ de R_t^m (D.15) dependen del tiempo. Sin embargo, para valores de t grandes, $w_{t,s} \approx w_s$ ($k=s$), según se nota en la ecuación (20) del artículo.

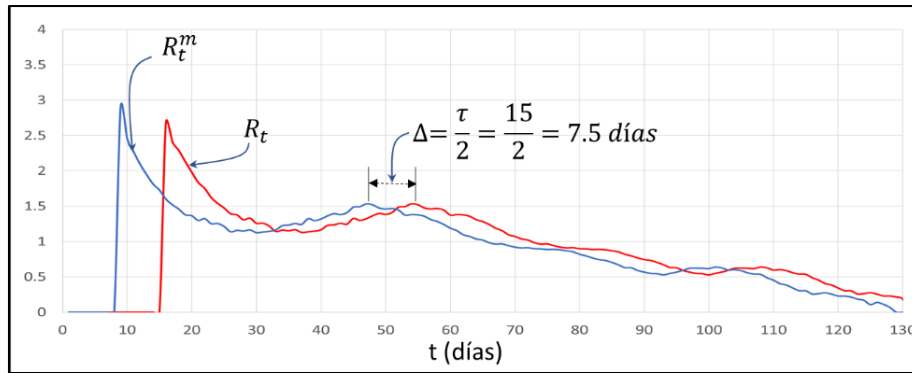


Figura D.2. Superposición de las curvas de R_t y R_t^m .

Las diferencias 2 y 3 no son suficientemente importantes, ya que aún tenemos muy buena coincidencia entre la curva adelantada de $R_t \rightarrow R_{t+\frac{\tau}{2}}$ y la curva de R_t^m como se aprecia en la Figura D.3.

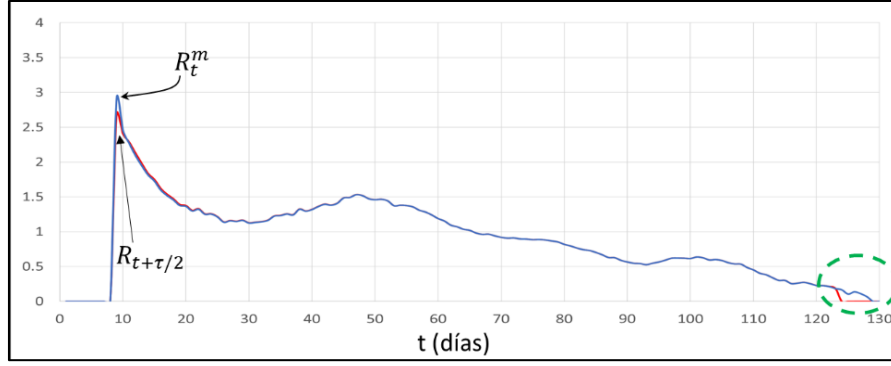


Figura D.3 Superposición de las curvas de $R_{t+\frac{\tau}{2}}$ y R_t^m .

E. Solución de un problema hipotético que sirve de modelo para calcular la desviación estándar, y el intervalo de confianza de R_t^m con incertidumbre en el intervalo serial

Supóngase que cuenta con M valores $x_{j,k}$, ($k = 1, 2, 3, \dots, M$) de una variable aleatoria x_j (la necesidad de subíndice j la explicaremos más adelante), generados por un proceso aleatorio discreto de M iteraciones, de tal forma que para un número M suficientemente grande se cumpla con:

$$\frac{\sum_{k=1}^M x_{j,k}}{M} = \mu_j, \text{ y } \frac{\sum_{k=1}^M (x_{j,k} - \mu_j)^2}{M} = \sigma_j^2 \quad (\text{E.1})$$

Condicionado a:

$$x_j \sim N(\mu_j, \sigma_j^2) \quad (\text{E.2})$$

Obviamente, (E.1) debe cumplirse si se acepta (E.2).

Supóngase ahora que se pasa a un escenario más complejo. Ahora el subíndice j es un apuntador de N variables x_j ($j = 1, 2, 3, \dots, N$), cada una de las cuales con medias y varianzas organizadas en duplas (μ_j, σ_j^2) . Además, cada x_j se relaciona con M valores aleatorios $x_{j,k}$, que cumplen con (E.1) y (E.2). Así, los $x_{j,k}$ valores son, evidentemente, los elementos de una matriz:

$$\begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,M} \\ x_{2,1} & x_{2,2} & \dots & x_{2,M} \\ \vdots & \vdots & \vdots & \vdots \\ x_{N,1} & x_{N,2} & \dots & x_{N,M} \end{bmatrix} \Rightarrow \begin{matrix} x_1 \sim N(\mu_1, \sigma_1^2) \\ x_2 \sim N(\mu_2, \sigma_2^2) \\ \vdots \\ x_N \sim N(\mu_N, \sigma_N^2) \end{matrix} \quad (\text{E.3})$$

Desde esta perspectiva matricial el planteamiento específico de nuestro problema es el siguiente:

Condiciones:

1. Los valores de las medias μ_j y las varianzas σ_j^2 son **conocidos**, y provienen de un proceso no especificado con el que se generan N duplas (μ_j, σ_j^2) de estos parámetros.
2. Los valores del renglón j de esta matriz son versiones aleatorias de la variable $x_j \sim N(\mu_j, \sigma_j^2)$.
3. Existe una variable global que llamaremos \mathcal{X} , cuyos valores aleatorios se obtienen del banco de datos de la matriz $[x_{j,k}]$, y que tiene media y varianza poblacional μ y σ^2 , respectivamente. No se conoce la función de densidad de probabilidad $g(x)$ correspondiente.

Preguntas:

Dada las condiciones, nos interesa calcular la media y la varianza de la variable \mathcal{X} :

1. Con base en las medias y varianzas prefijadas en la condición 1, y con el conocimiento de la normalidad de las variables x_j declarado en la condición 2.
2. Con base en todos los valores “crudos” aleatorios de la matriz.
3. ¿Se puede decir que los resultados en las preguntas 1 y 2 son razonablemente aproximados?

Solución 1 - Cálculo de la media y la varianza de \mathcal{X} con base a las duplas (μ_j, σ_j^2)

Ya que estamos asumiendo que conocemos las N versiones aleatorias de las medias y varianzas correspondientes a las variables x_j , el cálculo de la media de \mathcal{X} es inmediato:

$$\mu = \frac{\sum_{j=1}^N \mu_j}{N} \quad (E.4)$$

Esto en sí, ya es una de las respuestas que se buscaban, pero se aprovechará este resultado para mirar el problema desde una perspectiva diferente. Ya que los parámetros de medias μ y μ_j son los valores esperados de las variables aleatorias x y x_j , respectivamente, la ecuación (E.4) se puede presentar con más detalle:

$$E(x) = \frac{1}{N} \sum_{j=1}^N E(x_j) = \frac{1}{N} \sum_{j=1}^N \int_{x_j \rightarrow -\infty}^{x_j \rightarrow +\infty} x_j \frac{e^{-\frac{1}{2} \left(\frac{x_j - \mu_j}{\sigma_j} \right)^2}}{\sqrt{2\pi} \sigma_j} dx_j \quad (E.5)$$

Puesto que todas las variables x_j varían dentro del mismo rango $(-\infty, +\infty)$, el resultado de las N integraciones no se verá afectado si hacemos que las x_j varíen en conjunto, lo que implica que todas ellas son iguales a una misma variable, que provisionalmente llamaremos y . Así:

$$E(x) = \frac{1}{N} \sum_{j=1}^N \int_{-\infty}^{+\infty} y \frac{e^{-\frac{1}{2} \left(\frac{y - \mu_j}{\sigma_j} \right)^2}}{\sqrt{2\pi} \sigma_j} dy \quad (E.6)$$

Reordenando factores:

$$E(x) = \int_{-\infty}^{+\infty} y \frac{1}{N} \sum_{j=1}^N \frac{e^{-\frac{1}{2} \left(\frac{y - \mu_j}{\sigma_j} \right)^2}}{\sqrt{2\pi} \sigma_j} dy \quad (E.7)$$

Se nota que, de esta forma, el valor esperado de x no se afecta:

$$E(x) = \int_{-\infty}^{+\infty} y \frac{1}{N} \sum_{j=1}^N \frac{e^{-\frac{1}{2}\left(\frac{y-\mu_j}{\sigma_j}\right)^2}}{\sqrt{2\pi}\sigma_j} dy = \frac{\sum_{j=1}^N \mu_j}{N} = \mu \quad (\text{E.8})$$

Ya que la parte subrayada de la ecuación anterior cumple con todos los requisitos para ser una función de densidad de probabilidad (función siempre positiva, y el área total bajo la curva es igual a 1), y que el valor esperado calculado es igual al valor esperado de la variable x , podemos considerar la posibilidad de que $x=y$. Así, la función de probabilidad de x podría ser:

$$g(x) = \frac{1}{N} \sum_{j=1}^N \frac{e^{-\frac{1}{2}\left(\frac{x-\mu_j}{\sigma_j}\right)^2}}{\sqrt{2\pi}\sigma_j} = \frac{1}{N\sqrt{2\pi}} \sum_{j=1}^N \frac{e^{-\frac{1}{2}\left(\frac{x-\mu_j}{\sigma_j}\right)^2}}{\sigma_j} \quad (\text{E.9})$$

El cálculo de la *posible* varianza es inmediato:

$$V(x) = \frac{1}{N\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-\mu)^2 \sum_{j=1}^N \frac{e^{-\frac{1}{2}\left(\frac{x-\mu_j}{\sigma_j}\right)^2}}{\sigma_j} dx \quad (\text{E.10})$$

Definiendo:

$$V_j(x) = \frac{1}{N\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-\mu)^2 \frac{e^{-\frac{1}{2}\left(\frac{x-\mu_j}{\sigma_j}\right)^2}}{\sigma_j} dx \quad (\text{E.11})$$

Resolviendo $V_j(x)$:

$$V_j(x) = \frac{1}{N\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-\mu)^2 \frac{e^{-\frac{1}{2}\left(\frac{x-\mu_j}{\sigma_j}\right)^2}}{\sigma_j} dx = \frac{\sigma_j^2 + (\mu_j - \mu)^2}{N} \quad (\text{abreviando pasos}) \quad (\text{E.12})$$

Así:

$$V(x) = \sum_{j=1}^N \left[\frac{\sigma_j^2 + (\mu_j - \mu)^2}{N} \right] = \frac{1}{N} \sum_{j=1}^N \left[\sigma_j^2 + (\mu_j - \mu)^2 \right] \quad (\text{E.13})$$

Aclaración:

El resultado de (E.8) es una condición necesaria para considerar la posibilidad de que (E.9) sea la función de densidad de probabilidad de x , pero esto no implica que sea una condición suficiente. Ya que no estamos proveyendo una prueba analítica de esta suficiencia, en su defecto haremos una prueba experimental (simulación Montecarlo) en un ejemplo numérico de cálculo de varianzas, cumpliendo con las condiciones del problema. Los resultados experimentales se compararán con los resultados teóricos calculados con la (E.13).

Solución 2 - Cálculo de la media y la varianza de \mathcal{X} con base en los datos de la matriz $[x_{j,k}]$:

Para este fin construimos una matriz $[x_{j,k}]$ para generar los datos y hacer los cálculos requeridos. En la Figura E.1 mostramos parte de esta matriz y los resultados que obtuvimos para 50 iteraciones de una simulación Montecarlo.

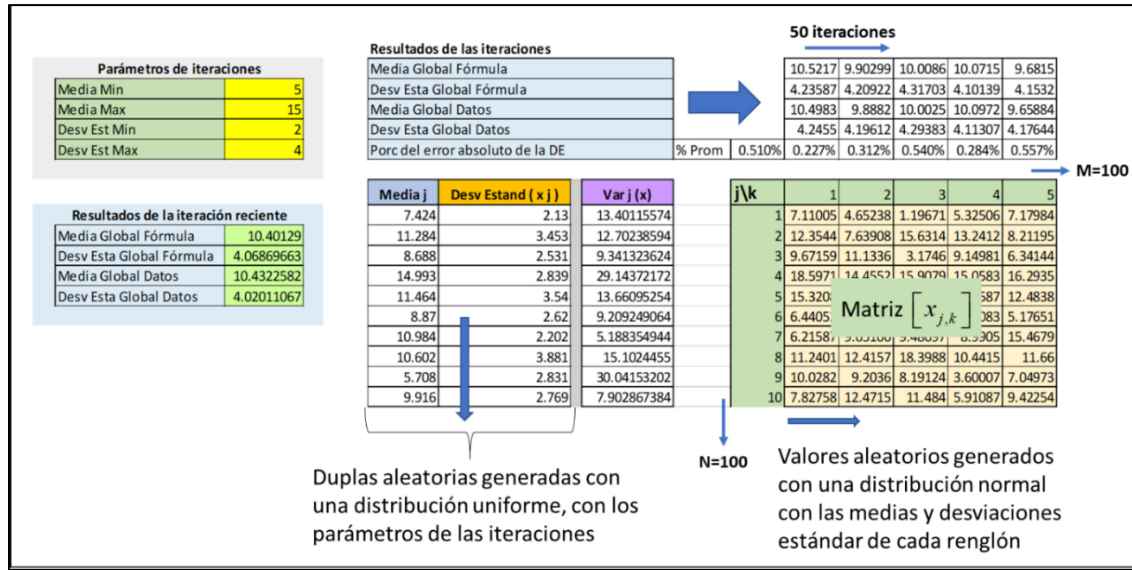


Figura E.1. Simulación Montecarlo.

Solución 3 – Comparación de los resultados calculados con la función de densidad de probabilidad propuesta (E.9) vs los resultados calculados directamente en la matriz $[x_{j,k}]$:

En la Figura E.1 se muestran los resultados obtenidos en 50 iteraciones (los resultados de cada iteración se producen en el recuadro “Resultados de la iteración reciente”, y se van copiando en los primeros cuatro renglones titulados “50 iteraciones”), y se comparan los resultados obtenidos por ambos métodos. Se encuentra que el promedio del porcentaje de error absoluto para el cálculo de la desviación estándar es de apenas 0.51 %. De hecho, con el mecanismo de prueba descrito, repetido una y otra vez, siempre obtuvimos porcentajes menores que 1%. La certeza absoluta de que los resultados son iguales en el infinito solo se lograría con una prueba analítica. Así, de aquí en adelante asumiremos que (E.9) es, al menos, una muy buena aproximación a la verdadera distribución de \mathcal{X} .

Aplicación de los resultados de este problema hipotético en el cálculo del número de reproducción instantáneo R_t^m cuando se considera que hay incertidumbre en el intervalo serial:

En el artículo principal se demuestra que, cuando se asume que el intervalo serial es constante, el R_t^m se distribuye normalmente con base en los siguientes parámetros estimados:

$$R_t^m = \frac{\text{Im}_t}{\Lambda_t^m}, \quad \hat{\sigma}_{R_t^m} = \frac{\hat{\sigma}_t}{\Lambda_t^m \sqrt{n_t}}, \quad \Lambda_t^m = \sum_{s=1}^t w_{t,s} \text{Im}_{t-s} \quad (\text{E.14})$$

Cuando se considera incertidumbre en el intervalo serial, los parámetros calculados con la (E.14) resultan también inciertos. La incertidumbre en el intervalo serial que se considera aquí es causada por una falta de información precisa, no es que se diga que el intervalo serial es un patrón de pesos intrínsecamente aleatorio. Esta incertidumbre se presenta comúnmente en los inicios de una epidemia, por la escases de datos para medir al intervalo serial. Para abordar este problema, se ha propuesto que los parámetros de la media μ_S y la desviación estándar σ_S del *intervalo serial* se traten como si fueran variables aleatorias que siguen alguna distribución, usualmente normal truncada (Cori et al., 2013). Como el cálculo por métodos analíticos se vuelve extremadamente complejo, se ha optado por aplicar simulaciones de Montecarlo (Cori et

al., 2013). El procedimiento consiste en hacer variar aleatoriamente a μ_S y a σ_S , asignándoles medias y desviaciones estándar (una media y desv. est. para μ_S ; y una media y desv. est. para σ_S). De esta forma se pueden generar versiones aleatorias de estos parámetros en parejas (duplas), las cuales producirán versiones distintas del intervalo serial. Para garantizar que el intervalo serial se anule en $t=0$, debe asegurarse que $\sigma_S < \mu_S$ en cada dupla (Cori et al., 2013). Cada intervalo serial conduce, a su vez, a versiones distintas de los parámetros de la media y desviaciones estándar de R_t^m (E.14). Es aquí donde entramos precisamente con los resultados de problema abstracto que resolvimos en este apéndice. Así, la conversión a las nuevas variables es la siguiente:

$$\mu_j \rightarrow R_{t,j}^m = \frac{\text{Im}_t}{\Lambda_{t,j}^m}, \quad \sigma_j \rightarrow \hat{\sigma}_{R_t^m}^{(j)} = \frac{\hat{\sigma}_t}{\Lambda_{t,j}^m \sqrt{n_t}}, \quad \Lambda_{t,j}^m = \sum_{s=1}^t w_{t,s}^{(j)} \text{Im}_{t-s} \quad (\text{E.15})$$

Los cálculos de arriba se harán, para cada instante t , y cada iteración j del intervalo serial.

De aquí se calcula la media de las N versiones $R_{t,j}^m$, que llamaremos R_t^m :

$$\mu = \frac{\sum_{j=1}^N \mu_j}{N} \rightarrow R_t^m = \frac{\sum_{j=1}^N R_{t,j}^m}{N} \quad (\text{E.16})$$

También, la desviación estándar de las N versiones de $R_{t,j}^m$, que llamaremos $\hat{\sigma}_{R_t^m}$:

$$\sqrt{V(x)} = \sqrt{\frac{1}{N} \sum_{j=1}^N \left[\sigma_j^2 + (\mu_j - \mu)^2 \right]} \rightarrow \hat{\sigma}_{R_t^m} = \sqrt{\frac{1}{N} \sum_{j=1}^N \left[\left(\sigma_{R_t^m}^{(j)} \right)^2 + \left(R_{t,j}^m - R_t^m \right)^2 \right]} \quad (\text{E.17})$$

Y, finalmente, la función de densidad de probabilidad de R_t^m :

$$g(x) = \frac{1}{N\sqrt{2\pi}} \sum_{j=1}^N \frac{e^{-\frac{1}{2} \left(\frac{x - \mu_j}{\sigma_j} \right)^2}}{\sigma_j} \rightarrow f(R_t^m) = \frac{1}{N\sqrt{2\pi}} \sum_{j=1}^N \frac{e^{-\frac{1}{2} \left(\frac{R_t^m - R_{t,j}^m}{\sigma_{R_t^m}^{(j)}} \right)^2}}{\sigma_{R_t^m}^{(j)}} \quad (\text{E.18})$$

La ecuación (E.18), por sí sola es suficiente para describir estadísticamente a R_t^m , pero desafortunadamente, resulta en procesos computacionales que tardan mucho para completarse. Por ejemplo, para calcular intervalos de confianza con (E.18), habría que resolver numéricamente las siguientes ecuaciones, para cada instante t :

$$\alpha/2 = \frac{1}{N\sqrt{2\pi}} \sum_{j=1}^N \int_{-\infty}^{\text{Lin}f} \frac{e^{-\frac{1}{2} \left(\frac{R_t^m - R_{t,j}^m}{\sigma_{R_t^m}^{(j)}} \right)^2}}{\sigma_{R_t^m}^{(j)}} dR_t^m, \quad (1 - \alpha/2) = \frac{1}{N\sqrt{2\pi}} \sum_{j=1}^N \int_{-\infty}^{\text{Lsup}} \frac{e^{-\frac{1}{2} \left(\frac{R_t^m - R_{t,j}^m}{\sigma_{R_t^m}^{(j)}} \right)^2}}{\sigma_{R_t^m}^{(j)}} dR_t^m \quad (\text{E.19})$$

Donde $\text{Lin}f$ y Lsup , son los límites de confianza inferior y superior de la media poblacional de R_t^m , dado un nivel de significancia α . Esta ecuación está compuesta por N distribuciones normales que se pueden programar con la función NORM.DIST (*acumulado*) de MSeExcel, cuyos resultados sumados y promediados, deberán igualarse a $\alpha/2$ ó a $(1-\alpha/2)$, luego de un proceso de aproximaciones cambiando el argumento ($\text{Lin}f$ ó el Lsup). Aunque este es un proceso semi-manual que requiere paciencia, los resultados obtenidos son los más exactos que se pueden lograr.

Si no se cuenta con el tiempo para aplicar (E.19), se puede recurrir a distribuciones subrogadas tales como la distribución t-Student, o la distribución gama, aplicando los parámetros de la media y la desviación estándar de las ecuaciones (E.16), y (E.17). En la Figura E.2 comparamos los límites del intervalo de confianza obtenidos aplicando el método exacto (E.19) y las

distribuciones subrogadas t-Student y gama. Los datos de la serie de incidencias son los mismos que se usan en la Figura 5 del artículo.

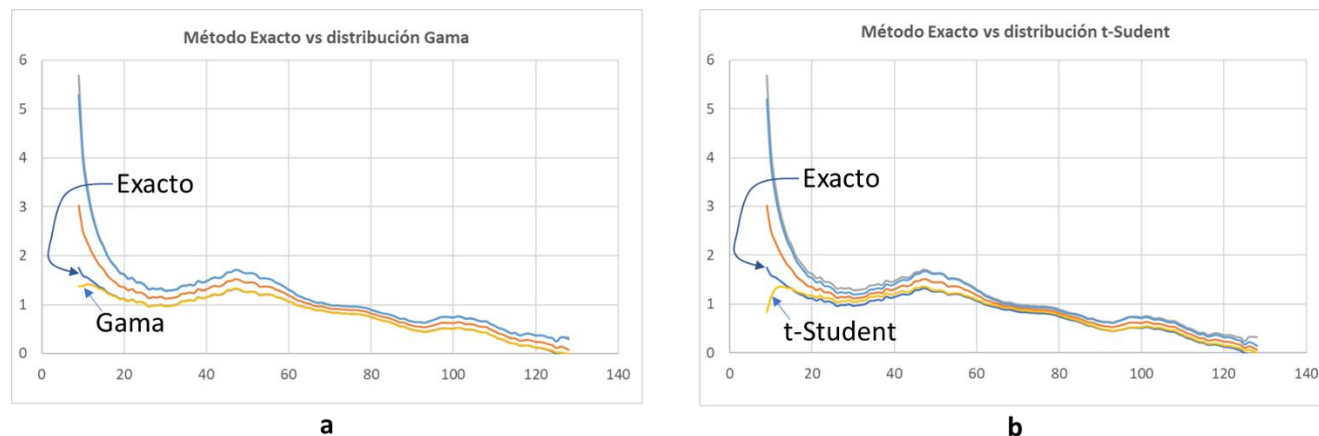


Figura E.2. Comparación del desempeño de las distribuciones Gama y t-Student para servir como funciones subrogadas de la distribución exacta (E.18).

Es evidente que entre las distribuciones Gama y t-Student, la distribución Gama es la que ofrece un mejor desempeño para aproximar a la distribución exacta.

F. Implementación con MSEXcel del método para el cálculo del R_t^m y aplicación en una epidemia real: “La Pandemia COVID-19 de Panamá”

La teoría desarrollada en el artículo principal, y los presentes apéndices se implementó en el software SeguiEpi, disponible en <https://github.com/mitzicubillamontilla/SeguiEpi.xlsm>

En la Figura F1 se muestra parte de la serie de tiempo de las incidencias (casos diarios) de la pandemia COVID-19 de Panamá, y la curva de suavización (medias). Esta es una carpeta del software llamada “Panel de Control 1”. Aquí se ingresa el tiempo de media móvil ($n_k = T_{mm}$), que ha de ser el máximo tiempo posible, tal que la serie del error pueda considerarse aleatoria. También se ingresa el nivel de significancia α para las pruebas de aleatoriedad, y de normalidad; y para el cálculo de los intervalos de confianza.

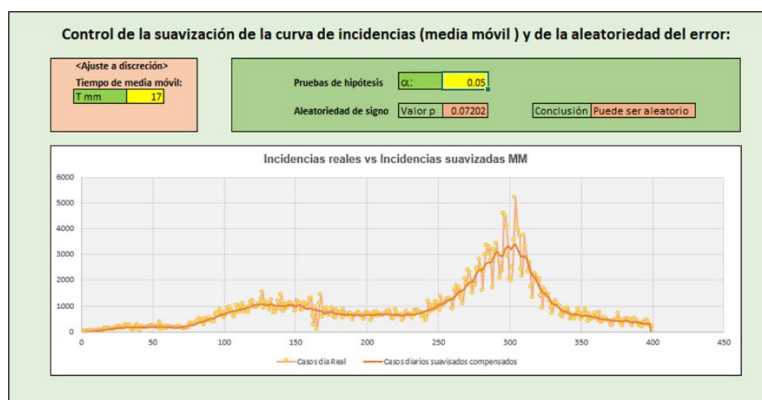


Figura F1. Carpeta “Panel de Control 1”: Gráfica de incidencias de la Pandemia COVID-19, Panamá 2020, y la curva de suavización.

Además del tiempo de media móvil, para calcular el error transformado e_t^* , es necesario ingresar otros parámetros con los que se “dibujará” la función para transformar el error (ver Figura F2). El *error plegado* es la superposición de los valores positivos del error con los valores absolutos de la parte negativa, de esta forma se busca dibujar un contorno (la función de transformación) que represente a ambas partes del error. En el software se tienen 3 formas para calcular la función de transformación, usando una función *Potencia*, una función *Logit*, y una función *Logipot*, según se propuso en el apéndice A. En la Figura F2 se observan los parámetros de entrada para modelizar el contorno.

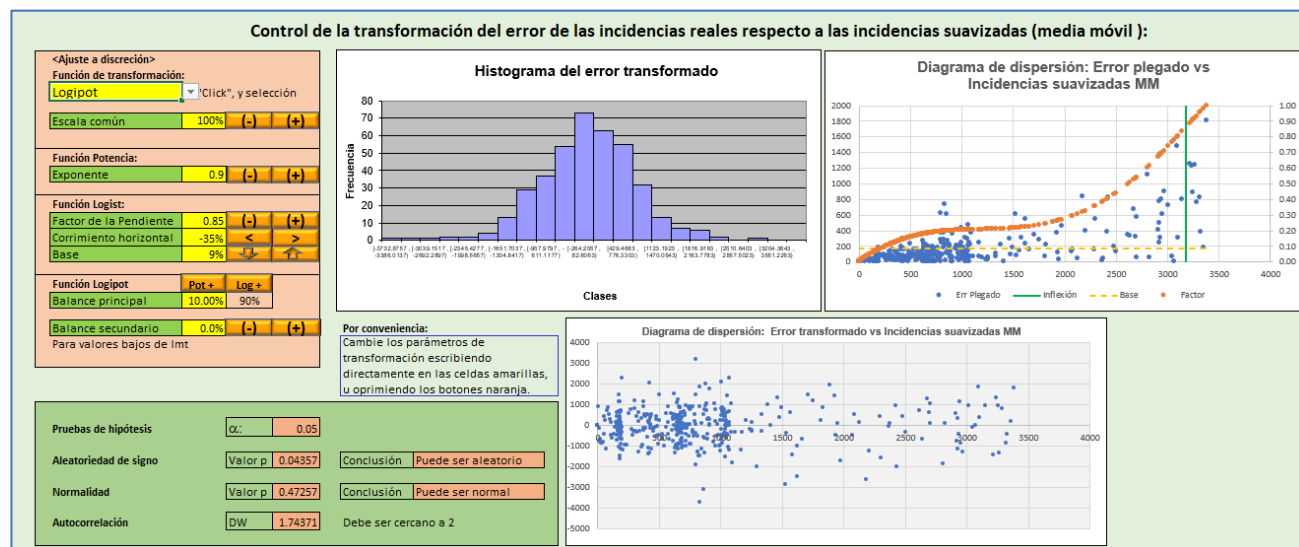


Figura F2. Carpeta “Panel de Control 2”: Entrada de parámetros para la transformación del error.

La definición de los parámetros se hace con el apoyo de las pruebas de hipótesis de normalidad y aleatoriedad, complementándose con una prueba de autocorrelación de Durbin Watson. También contamos con apoyos gráficos: histograma, diagrama de dispersión del error transformado vs la media de las Incidencias, y el diagrama de dispersión del error plegado vs la media de las incidencias. Se aprecia que el error transformado e_t^* se ubica en una franja horizontal en el diagrama de dispersión de este error con la media de las incidencias (casos suavizados MM), tal como se desea.

La transformación del error y la definición del tiempo de media móvil se requieren para calcular la desviación estándar del error transformado $\hat{\sigma}_t^*$. En el software se dan tres modalidades: Constante total, Constante por tramos, y Continuo (ver Figura F3). El modo continuo requiere ingresar un tiempo de media móvil específico para calcular la media móvil del rango móvil del error transformado.

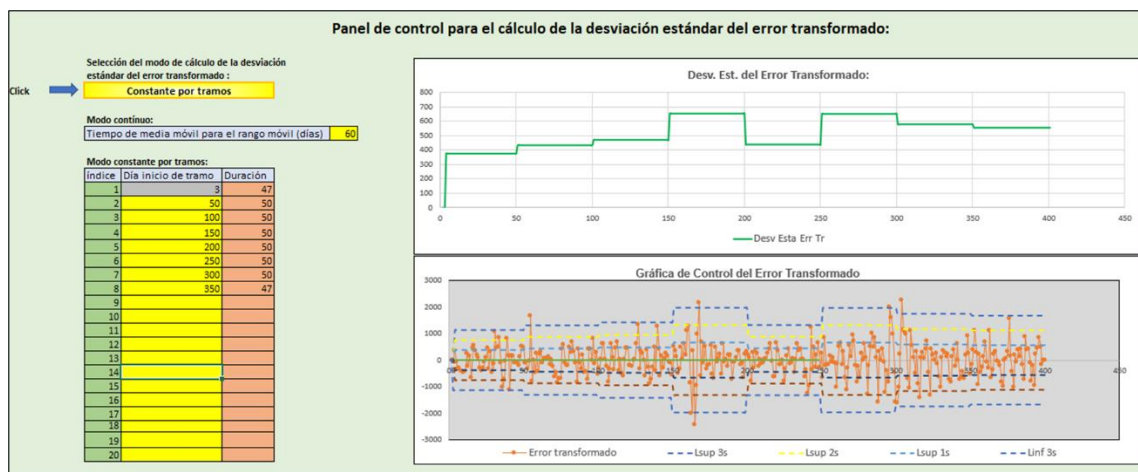


Figura F3. Carpeta “Panel de Control 3”: Cálculo de la desviación estándar del error transformado.

En la Figura F4 se muestra la superposición las incidencias, su media estimada, y los límites de confianza para la media poblacional.

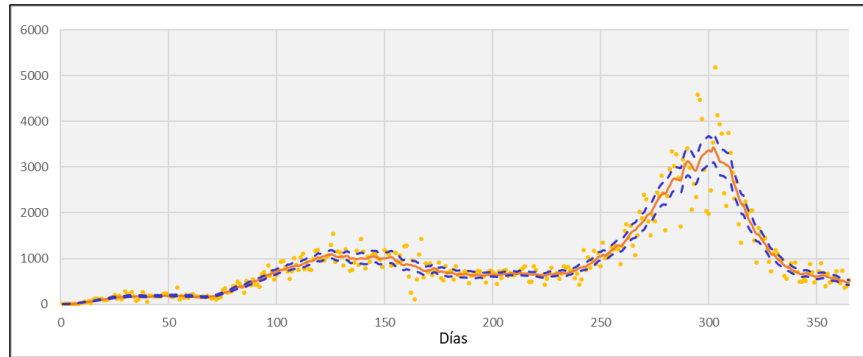


Figura F4. Gráfica de incidencias, media e intervalo de confianza de la Pandemia COVID-19, Panamá 2020.

Para los cálculos del *número de reproducción instantáneo* (R_t^m) se requieren definir los pesos $w_{t,s}$ a partir de las constantes w_s del *intervalo serial*. Esto se hace con el apoyo del software (Figura F5), usando como entrada los parámetros de la media y la desviación estándar. En el programa se dan 3 alternativas de distribuciones: gamma, lognormal, y una distribución libre definida por el usuario. Los cálculos del intervalo serial se ofrecen sin incertidumbre, y con incertidumbre.

Definición del intervalo Serial Sin Incertidumbre

El intervalo Serial (IS) es una distribución de probabilidad discreta que define la importancia (wi) que tienen los casos contados (incidentes) en el pasado para influir el conteo de los casos en el tiempo presente. Es la distribución del tiempo que tarda un infectado primario desde que manifiesta síntomas, hasta que se manifiestan los síntomas en un infectado secundario, infectado por el primario. Es un patrón, que se asume constante, resultante de la interacción biológica del ser humano con el agente infeccioso. Las entradas que usted haga en esta carpeta definirán los resultados en el cálculo del R_t .

Cuando no hay incertidumbre acerca del Intervalo Serial (IS), en este software el IS puede construirse de 3 maneras:

- (1) Mediante una distribución gamma, dando como parámetros la media y la desviación estándar.
- (2) Mediante una distribución lognormal, dando como parámetros la media y la desviación estándar.
- (3) Mediante una distribución discreta de probabilidad asignada por el usuario con base en alguna teoría.

¿Cómo hará la definición del IS? (1 ó 2 ó 3):

Si escogió (1 ó 2):

Media:
 Desv Est:

Si escogió (3):

Llene las casillas de abajo con el patrón de pesos (wi) correspondiente, cumpliendo con que la suma de los wi sea 1, ó ligeramente menor (ej., 0.99+sum(wi=1)).

i	wi
1	0.00067
2	0.01117
3	0.04267
4	0.07785
5	0.10409
6	0.14811
7	0.11843
8	0.10736
9	0.09075
10	0.07918
11	0.06292
12	0.04944

Intervalo Serial (días) para presentar el R_t :

Actualización de resultados SI

Intervalo Serial

i	wi
1	0.0019548
2	0.0261192
3	0.0436638
4	0.0746535
5	0.0991123
6	0.1127967
7	0.1147933
8	0.1077542
9	0.0953523
10	0.0801289
11	0.0648975
12	0.0508436
13	0.0387897
14	0.0289177
15	0.0211188
16	0.01518
17	0.0107389
18	0.0074905
19	0.0051892
20	0.0037386
21	0.0025827
22	0.0019397
23	0.0013662
24	0.0008991
25	0.0006379
26	0.0004681

Intervalo Serial

a.

Definición del intervalo Serial Con Incertidumbre

El significado del Intervalo Serial que se da en la carpeta "Intervalo Serial" también aplica en esta carpeta. La diferencia en esta carpeta está en el grado de certeza que asumimos acerca del Intervalo Serial. En la carpeta "Intervalo Serial SI", se asume total certeza. Sin embargo, como al principio de una epidemia realmente no se cuenta con datos suficientes para aventurarse a asumir certeza total, se propone modelar la incertidumbre correspondiente mediante medias y desviaciones estándar de los parámetros del Intervalo Serial (Cori et al, 2013). Se sigue asumiendo que el Intervalo Serial es constante, pero se acepta que los cálculos que derivan de éste (ej., R_t) están afectados por la incertidumbre inherente a los datos con los que se calcula el Intervalo Serial.

Cuando hay incertidumbre acerca del Intervalo Serial (IS), en este software el IS puede construirse de 2 maneras:

- (1) Mediante una distribución gamma, dando como parámetros la media y la desviación estándar.
- (2) Mediante una distribución lognormal, dando como parámetros la media y la desviación estándar.

¿Cómo hará la definición del IS? (1 ó 2):

Simulación de la incertidumbre del Intervalo Serial:

μS		σS	
Media	8.4	Media	3.8
Desv. Estándar	1	Desv. Estándar	0.5
Min	4.4	Min	1.8
Max	12.4	Max	5.8

Actualización de resultados CI

Intervalo Serial Medio

i	wi
1	0.001954781
2	0.026119176
3	0.043663789
4	0.074653664
5	0.099112190
6	0.1127967
7	0.114793294
8	0.107754205
9	0.095351479
10	0.080129885
11	0.064887547
12	0.050845087
13	0.038789748
14	0.028917096
15	0.021113576
16	0.01518
17	0.010738019
18	0.007495471
19	0.005189148
20	0.003735262
21	0.002582718
22	0.001939644

Intervalo Serial Medio

b.

Figura F5. Carpetas para la definición del intervalo serial: a) Intervalo Serial sin Incertidumbre. b) Intervalo Serial con Incertidumbre.

La aplicación de este intervalo serial con incertidumbre condujo a la curva de R_t^m , y el intervalo de confianza que se muestran en la Figura F6.

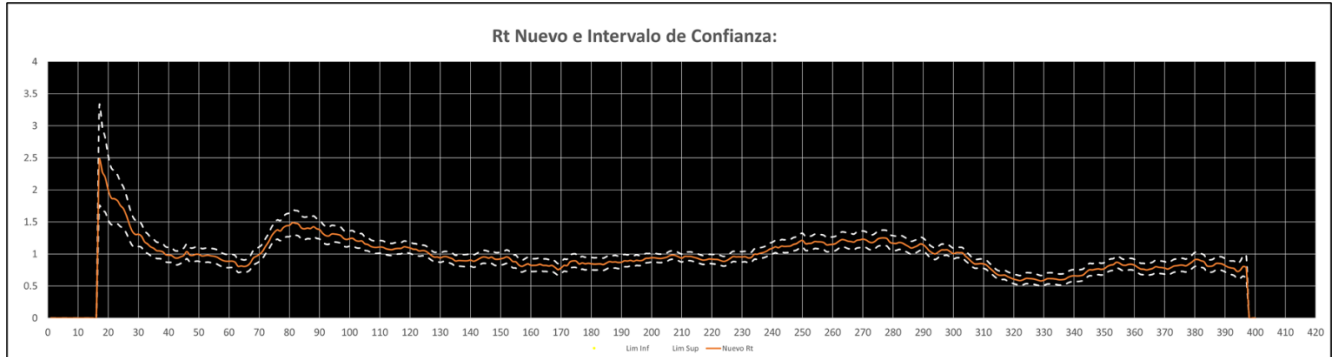


Figura F6. Gráfica del número de reproducción instantáneo R_t^m , y su intervalo de confianza, de la Pandemia COVID-19, Panamá 2020.

Referencias

- Cori, A., Ferguson, N. M., Fraser, C., & Cauchemez, S. (2013). Practice of Epidemiology / A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. *American Journal of Epidemiology*, 178(9), 1505–1512. <https://doi.org/10.1093/aje/kwt133>
- Larsen, R., & Marx, M. (2012). *An introduction to mathematical statistics* (5th ed.). Prentice Hall / Pearson.
- Makridakis, S., Wheelwright, S., & Hyndman, R. (1998). *Forecasting: methods and applications* (3th ed.). John Wiley & Sons, Inc.
- Montgomery, D. (2001). *Introduction to statistical quality control* (4th ed.). John Wiley & Sons, Inc.