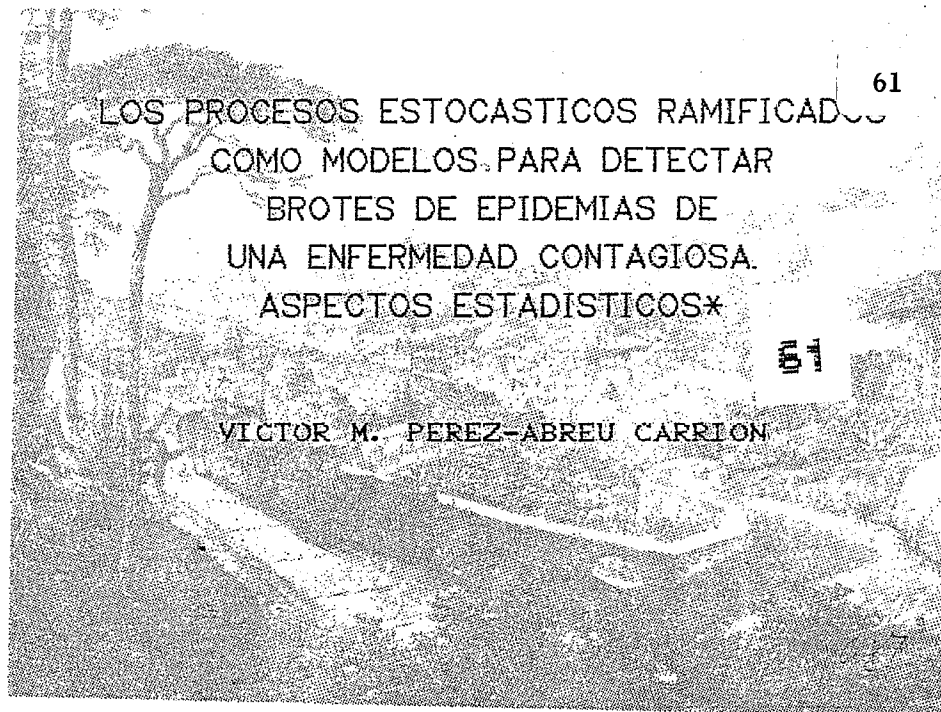


COMUNICACIONES DEL CIMAT



- * Parte de este trabajo fue presentado en el II Foro de Estadística Aplicada, celebrado en el Auditorio del Jardín Botánico Exterior de Cd. Universitaria, D. F., del 28 al 30 de octubre de 1987, así como en el Seminario de los Miércoles del CIMAT, el 9 de septiembre del mismo año.

CENTRO DE INVESTIGACION EN MATEMATICAS

Apartado Postal 402

Guanajuato, Gto.

México

Tels. (473) 2-25-50

2-02-58

INTRODUCCION

El objetivo de este trabajo es presentar de una manera elemental algunos aspectos estadísticos de los procesos estocásticos ramificados y su aplicación como modelos para describir brotes de epidemias de una enfermedad contagiosa. Asimismo se quiere ilustrar cómo un estadístico podría ser útil en la solución de algunos problemas asociados al control de enfermedades contagiosas.

La detección temprana del brote de una epidemia es importante ya que puede ayudar a las autoridades del Sector Salud a prevenir que la enfermedad se extienda, tomando medidas pertinentes como puede ser el determinar la fracción de la comunidad susceptible a la enfermedad que debe ser vacunada, en caso de que la vacuna exista. Los modelos estadísticos que aquí se presentan también ayudan a detectar el brote de una epidemia después que alguna de estas medidas ha sido tomada, pero aun existen en la población algunos casos infectados no identificados o existe el riesgo de que lleguen a la comunidad portadores de la enfermedad.

La organización del trabajo es la siguiente. En la Sección 1 damos una introducción elemental a los procesos estocásticos ramificados simples, también conocidos como procesos de Galton-Watson. Se hace énfasis en su aplicación dentro de un lenguaje y contexto epidemiológico. En la Sección 2 describimos algunos métodos estadísticos de estos modelos y

sus principales propiedades y desventajas. En la Sección 3 presentamos una simulación del modelo con el objeto de ilustrar el uso de los métodos estadísticos, así como discutimos la imposibilidad de aplicarlos a unos datos reales de casos de Hepatitis Vírica registrados en el estado de Guanajuato.

En la realización de este trabajo me he beneficiado de los comentarios de los alumnos de mi curso de Procesos Estocásticos de Maestría: Paul Blanquet y Antonio González, así como de los de la M. en C. Rebeca Ponce de León. El trabajo de cómputo fue hecho por Emilio Marmolejo, alumno de la Escuela de Matemáticas de la Universidad de Guanajuato. La mecanografía estuvo a cargo de Laura Rincón Gallardo. A todos ellos mi agradecimiento.

1. PROCESOS DE GALTON-WATSON

Un bosquejo histórico amplio de la teoría de los procesos estocásticos ramificados aparece en el Capítulo 1 del libro de P. Jagers (1975), al cual el lector interesado es remitido. Sólo mencionaremos aquí que el origen de los procesos estocásticos ramificados está asociado a los nombres, entre otros de Thomas Malthus, Sir Francis Galton y el clérigo y matemático H. W. Watson. En los trabajos de los dos primeros autores se estudia la extinción del apellido de una familia. Así por ejemplo, T. Malthus observó que de 487 familias burguesas en la ciudad de Berna, 379 se extinguieron en un período de dos siglos (1583-1783), cosa que T. Malthus reconoció como un problema en el que no está presente la "pura

coincidencia". Por otro lado, en sus "Ensayos sobre los Principios de Poblaciones", Malthus expresa la idea fundamental de que una población, cuando no se controla, tiende a crecer exponencialmente. Las características probabilísticas de una población que determinan la posible extinción o explosión de ella son uno de los problemas de la teoría de los procesos estocásticos ramificados resuelto por H. W. Watson.

Los procesos estocásticos ramificados como modelos para describir la magnitud del inicio de una epidemia menor han sido usados, entre otros, por Neyman y Scott (1964) y Becker (1975), trabajos en los cuales se pueden encontrar más referencias sobre el tema. Estos procesos son adecuados como modelos de epidemias siempre y cuando la población de susceptibles sea grande, lo cual es cierto durante los principios del brote, cuando no se ha tomado ninguna medida pertinente.

La descripción de los procesos estocásticos ramificados simples, también conocidos como procesos de Galton-Watson, es la siguiente: supongamos que a un determinado tiempo inicial se tiene un ($X_0 = 1$) enfermo de una enfermedad contagiosa, el cual contagia a un número aleatorio X_1 en un determinado tiempo al cual llamaremos generación. A su vez, durante la segunda generación del brote, estos X_1 enfermos contagian, cada uno de manera independiente y con la misma probabilidad, a un número aleatorio de enfermos, cuyo total denotaremos por X_2 . Así X_2 es el número de enfermos contagiados durante la segunda generación

de la enfermedad por los enfermos X_1 , y por lo tanto X_2 depende de X_1 .

El proceso continúa de esta manera y denotamos por X_0, \dots, X_n las primeras n generaciones del modelo. La probabilidad de que un enfermo contagie k individuos será denotada por P_k y estas probabilidades se conocen como probabilidades de contagio o distribución de contagio. En el modelo se supone que la distribución de contagio no varía de generación en generación, por lo que se dice que el modelo es homogéneo.

Las distribuciones de contagio paramétricas más usadas en la práctica son las siguientes: la de Poisson con media $\lambda > 0$

$$P_k = \lambda^k \frac{e^{-\lambda}}{k!} \quad k = 0, 1, 2, \dots, \quad (1.1)$$

la geométrica con parámetro P

$$P_k = P(1-P)^k \quad k = 0, 1, 2, \dots, \quad 0 < P < 1 \quad (1.2)$$

y la binomial con parámetros ℓ y P

$$P_k = \frac{\ell!}{k!(\ell-k)!} P^k (1-P)^{\ell-k} \quad k = 0, 1, \dots, \ell, \quad 0 < P < 1 \quad (1.3)$$

La media de contagio m del modelo se define como

$$m = \sum_{k=0}^{\infty} k P_k \quad (1.4)$$

y la varianza σ^2 de la siguiente manera:

$$\sigma^2 = \sum_{k=0}^{\infty} (k-m)^2 P_k \quad (1.5)$$

Las medias de contagio para las distribuciones de contagio arriba mencionadas se presentan en la Tabla 1.1.

Distribución de contagio	Poisson λ	Geométrica P	Binomial ℓ, P
Media de contagio	$m = \lambda$	$m = \frac{1-P}{P}$	$m = \ell P$

Tabla 1.1.
Media para algunas distribuciones de contagio

Una herramienta matemática importante en el estudio de los procesos ramificados es la función generadora de probabilidades f que se define de la siguiente manera

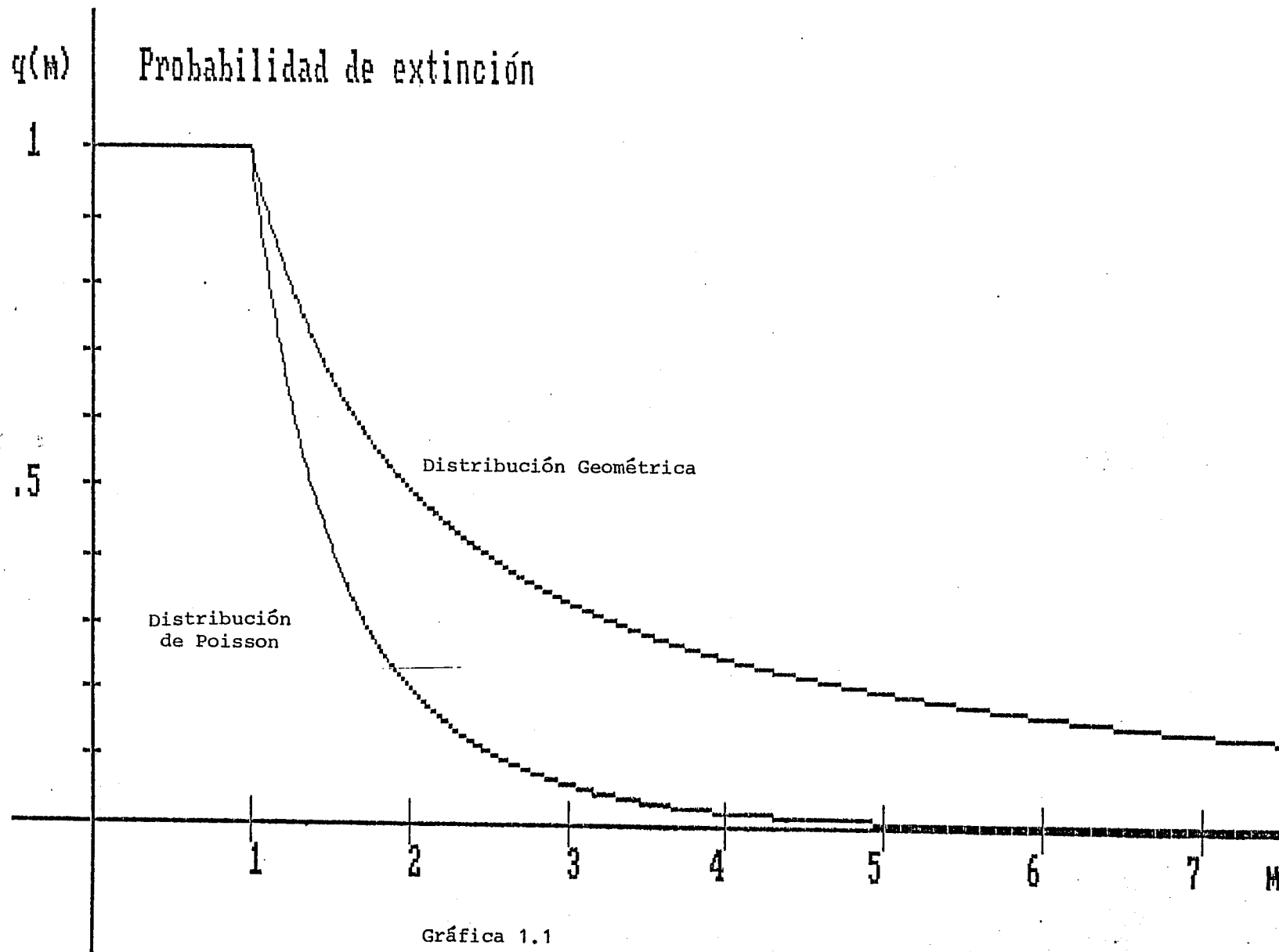
$$f(t) = \sum_{k=0}^{\infty} t^k P_k \quad |t| < 1 \quad (1.6)$$

Finalmente, denotaremos por q a la probabilidad de extinción de la epidemia, o sea q es la probabilidad de que X_n sea cero para alguna generación n . La importancia de la función generadora de probabilidades queda de manifiesto en el siguiente teorema probado por H. W. Watson y J. Steffensen.

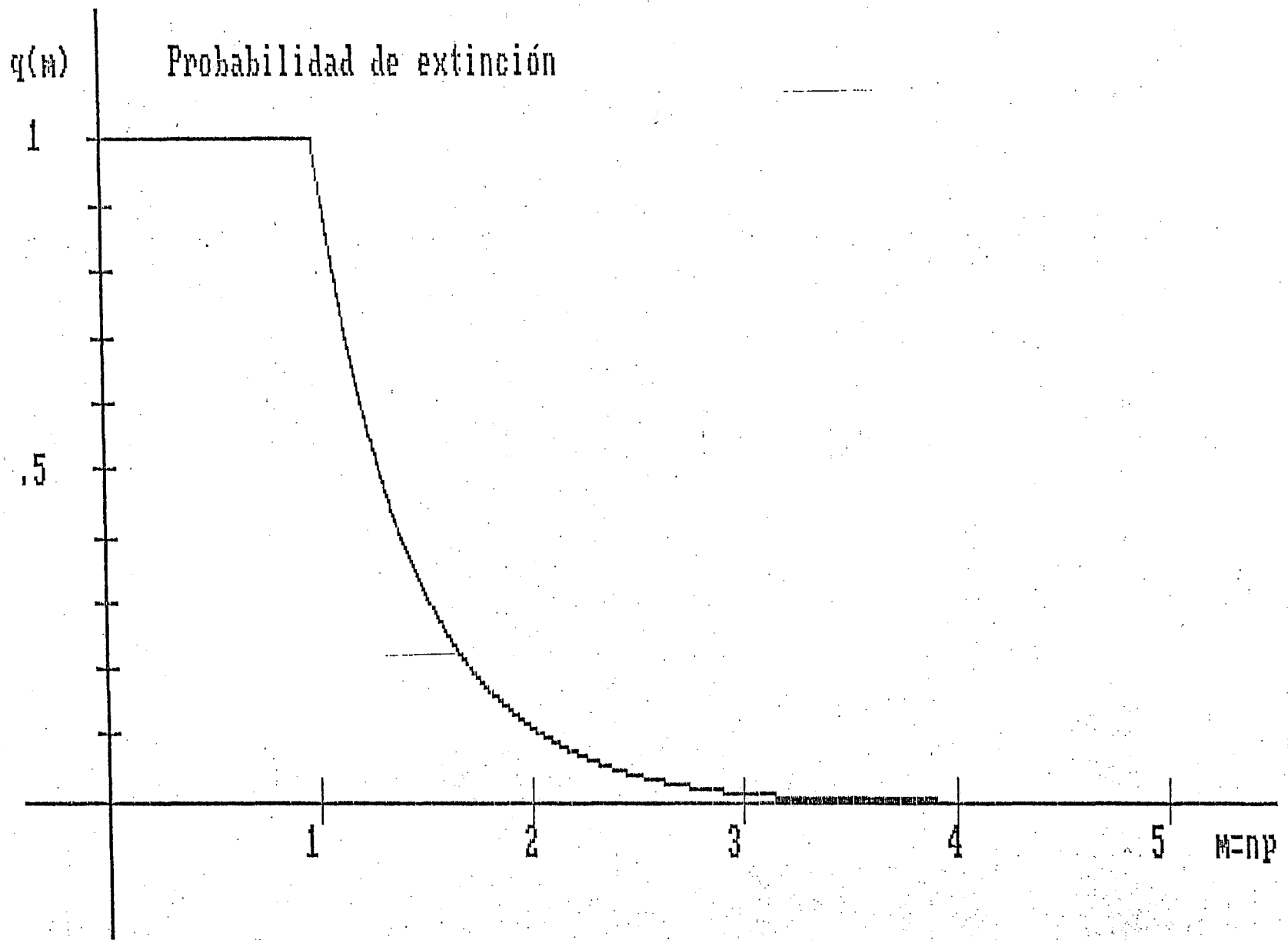
Teorema 1.

- a) Si $m \leq 1$ entonces $q = 1$.
- b). Si $m > 1$ entonces q es la única raíz mayor o igual a cero y menor a uno de la ecuación $s = f(s)$.

El teorema anterior nos expresa que si la media de contagio es menor o igual a uno, entonces la probabilidad de extinción de la epidemia es uno; de otra forma, si la media de contagio es mayor que uno la probabilidad de extinción es la raíz diferente de uno en el intervalo $[0,1]$ de la ecuación $s = f(s)$. En las gráficas 1.1 y 1.2 se muestra el



Probabilidad de extinción para las distribuciones de contagio Geométrica y Poisson.



Gráfica 1.2

Probabilidad de extinción para la distribución
de contagio binomial $n=5$

comportamiento de la probabilidad de extinción con respecto a la media de contagio para las distribuciones de Poisson, geométrica y binomial con parámetros λ y P . En todas ellas se observa que el decrecimiento de la probabilidad de extinción es bastante rápido cuando m se aleja de uno. Las tablas 1.2 y 1.3 muestran algunos valores de q para distintos valores de m para las distribuciones Poisson y geométrica respectivamente.

λ	1	1.05	1.1	1.2	1.4	1.6	1.8	2.0
q	1	0.91	0.82	0.69	0.49	0.36	0.27	0.20

Tabla 1.2
Probabilidad de extinción para algunos valores de la media λ de una distribución de Poisson

m	1	1.05	1.1	1.2	1.4	1.6	1.8	2.0
q	1	0.95	0.91	0.83	0.71	0.63	0.55	0.50

Tabla 1.3
Probabilidad de extinción para algunos valores de la media $m = (1-P)/P$ de una distribución geométrica

Del teorema 1 y de las gráficas y tablas mencionadas anteriormente observamos que cualquier control de una enfermedad contagiosa debe tender a mantener la media de contagio de una comunidad menor o igual a uno o muy poco superior a uno. Si m se conoce y se sabe que es mayor a uno, debemos determinar la proporción θ de individuos susceptibles a

la enfermedad que deben ser vacunados para evitar una epidemia. Si seleccionamos estos individuos al azar dentro de la comunidad es razonable esperar que $m^* = (1-\theta)m$ será la nueva media de contagio de la comunidad si estos individuos son vacunados. Puesto que queremos una extinción de la epidemia debemos tomar θ de tal forma que $m^* \leq 1$, o sea que θ debe ser al menos $1-1/m$. Así por ejemplo, si $m = 1.5$ la proporción de individuos vacunados debe ser al menos del 30%. Este resultado no depende de la distribución de contagio sino únicamente de la media. Sin embargo, una vez que es probable la ocurrencia de una epidemia, la probabilidad con que ocurra sí está determinada por la distribución de contagio. Al respecto comentaremos más en la sección 2.

2. METODOS ESTADISTICOS

En la sección anterior hemos visto la importancia de conocer el valor de la media de contagio m . En esta sección presentamos algunos métodos estadísticos para estimarla. Primeramente haremos la inferencia bajo el supuesto que observamos N muestras independientes de un proceso de Galton-Watson de tamaño n ; es decir suponemos que es posible observar $X_0^{(i)}, \dots, X_n^{(i)}$ para $i = 1, \dots, N$. Si la distribución de contagio es Poisson, geométrica o binomial, el estimador de máxima verosimilitud \hat{m} de m es

$$\hat{m} = \frac{\sum_{i=1}^N \sum_{k=0}^{n-1} X_{k+1}^{(i)}}{\sum_{i=1}^N \sum_{k=0}^{n-1} X_k^{(i)}} \quad (2.1)$$

el cual representa el "número total de contagiados entre el número total de contagiadores".

Si N es grande y n se mantiene fija

$$Z = \frac{1}{\hat{\sigma}} \left\{ N \frac{\binom{\hat{m}}{n-1}}{\binom{\hat{m}-1}{n-1}} \right\}^{1/2} (\hat{m}-m) \quad (2.2)$$

tiene una distribución aproximadamente normal con media cero y varianza uno. La expresión para $\hat{\sigma}^2$, el estimador de máxima verosimilitud de σ^2 , depende de la distribución de contagio de la siguiente manera: para la distribución de Poisson $\hat{\sigma}^2 = \hat{m}$, para la geométrica $\hat{\sigma}^2 = \hat{m}(1+\hat{m})$ y para la distribución binomial con parámetros ℓ y P $\hat{\sigma}^2 = \hat{m}(1-\hat{m}/\ell)$. Por lo tanto un intervalo aproximado de $(1-\alpha)\%$ de confianza para la media será

$$\left[\hat{m} - Z_{\alpha/2} \hat{\sigma} \left\{ \frac{1}{N} \frac{\hat{m}-1}{\binom{\hat{m}}{n-1}} \right\}^{1/2}, \hat{m} + Z_{\alpha/2} \hat{\sigma} \left\{ \frac{1}{N} \frac{\hat{m}-1}{\binom{\hat{m}}{n-1}} \right\}^{1/2} \right] \quad (2.3)$$

en donde $Z_{\alpha/2}$ es el $1-\alpha/2$ percentil de una distribución normal estándar.

En el caso en que sólo es posible observar una realización del proceso de Galton-Watson, es decir X_0, X_1, \dots, X_n , el estimador de máxima verosimilitud \hat{m} de m cuando las distribuciones de contagio son Poisson, geométrica o binomial es

$$\hat{m} = \frac{X_1 + \dots + X_n}{X_0 + \dots + X_{n-1}} \quad (2.4)$$

Si n es grande y ninguno de los X_i es cero

$$Z = \frac{1}{\hat{\sigma}} \left\{ X_0 + \dots + X_{n-1} \right\}^{1/2} (\hat{m} - m) \quad (2.5)$$

tiene una distribución aproximadamente normal con media cero y varianza uno, en donde $\hat{\sigma}^2$ depende de la distribución de contagio de la misma manera que en la primera parte de esta sección. Así, un intervalo aproximado de $(1-\alpha)\%$ de confianza para la media será, si n es grande y ninguno de los X_i es cero,

$$\left[\hat{m} - \frac{Z_{\alpha/2} \hat{\sigma}}{(X_0 + \dots + X_n)^{1/2}}, \hat{m} + \frac{Z_{\alpha/2} \hat{\sigma}}{(X_0 + \dots + X_n)^{1/2}} \right] \quad (2.6)$$

Es importante observar que aunque la expresión para \hat{m} , en ambos casos presentados, no depende de la distribución de contagio, la expresión para los intervalos de confianza sí depende de ella.

Lo anterior refleja la importancia de tener métodos de estimación para las probabilidades de contagio P_k . En el presente éstas se pueden estimar únicamente si es posible observar cuántos contagios produce cada enfermo, lo cual no siempre es factible. En la práctica tiene que suponerse alguna distribución de contagio.

Otros métodos de estimación de m , así como inferencia estadística para la varianza σ^2 y la probabilidad de extinción q se encuentran en el Capítulo 3 de la excelente monografía de Nanthi (1983) o en el Capítulo 2 del libro de Basawa y Prakasa-Rao (1980), trabajos en los cuales se pueden encontrar más referencias al respecto.

3. EJEMPLOS

En la tabla 3.1 se presentan casos de Hepatitis Vírica por mes registrados en la Secretaría de Salud del estado de Guanajuato en los años 1977 a 1987. Con el objeto de ver las limitaciones y alcances de los modelos y métodos de las secciones anteriores, explicaremos por qué éstos no son aplicables a los datos. Esto tiene el doble objetivo de sugerir otra forma de recopilación de los datos así como las condiciones adecuadas para la aplicación correcta del modelo. Empezaremos por mencionar que la recopilación mensual no es la más adecuada, ya que ésta debiera hacerse más de acuerdo al período de incubación-contagio de la enfermedad, con el objeto de tener definidas las generaciones del proceso lo mejor posible. En segundo lugar observamos que los datos sólo incluyen los casos reportados en la Secretaría de Salud y no todos los casos en el estado de Guanajuato. Esto viola una suposición del modelo de la Sección 1, en la que se supone que todos los casos de una generación fueron contagiados por los de la generación anterior. Esto no sucede necesariamente, pues no hay seguridad de que los 10 casos reportados en junio de 1977 hayan sido contagiados sólo por los seis casos reportados en mayo del mismo año. Otra violación del modelo podría ser la homogeneidad del contagio pues no es necesariamente cierto que la distribución de contagio es la misma en una zona rural que en una urbana.

SECRETARIA DE SALUD Y SEGURIDAD SOCIAL DEL ESTADO
 SUBDIRECCION DE PLANEACION
 DEPARTAMENTO DE ESTADISTICA Y EVALUACION

CASOS DE HEPATITIS VIRICA POR MES
 REGISTRADOS EN LA SECRETARIA DE SALUD
 1977 - 1987

AÑOS	N U M E R O D E C A S O S												TOTAL
	E	F	M	A	M	J	J	A	S	O	N	D	
1977	3	1	4	5	6	10	6	7	8	11	2	4	67
1978	2	8	5	3	0	7	1	6	5	6	5	3	51
1979	2	0	2	2	11	4	8	8	16	18	8	5	84
1980	1	1	0	3	3	0	4	2	6	5	4	1	37
1981	5	2	3	5	1	5	6	4	8	2	5	4	50
1982	3	11	5	0	5	7	5	19	24	16	13	0	117
1983	7	2	5	3	5	3	4	10	14	22	19	17	111
1984	8	15	12	11	9	13	15	17	14	25	19	4	162
1985	3	3	8	5	5	5	8	6	5	10	8	2	68
1986	1	12	4	5	6	10	4	19	21	22	18	8	131
1987	8	3	2	1	6	1	2	1*					25

FUENTE: BOLETIN EPIDEMIOLOGICO SSSSG
 *HASTA SEMANA 35 (AGOSTO 2- SEPT 5-)

Tabla 3.1

13

Lo anterior sugiere que la recopilación de los datos se haga con un período mejor definido y por comunidades más pequeñas y homogéneas (escuelas, por ejemplo). De esta manera, con los datos de las comunidades similares se estima la media mediante la ecuación (2.1) y suponiendo una distribución de contagio adecuada se construye un intervalo de confianza como en (2.3). De esta forma se podría tener un control de la enfermedad pues cada vez que se reportaran en la comunidad casos de ella, mediante (2.4) se estima la media y se observa si cae o no dentro de un intervalo de confianza peligroso, lo que serviría para detectar el surgimiento de una epidemia.

Con el objeto de ilustrar lo anterior, realizamos una simulación en la computadora de $N=15$ procesos de Galton-Watson de $n=5$ generaciones con distribución de contagio geométrica con media 1.3. Los resultados se muestran en la tabla 3.2. Utilizando el estimador (2.1) obtenemos

$$m = \frac{267}{200} = 1.335 \quad (3.1)$$

lo cual corresponde a una probabilidad de extinción alrededor de 0.75. El estimador de máxima verosimilitud de σ^2 es

$$\hat{\sigma}^2 = \hat{m}(1+\hat{m}) = 3.117 \quad (3.2)$$

Por lo tanto de (2.3) obtenemos el intervalo de 90% de confianza ($Z_{\alpha/2} = 1.645$)

$$(1.094, 1.576) \quad (3.3)$$

Observamos que este intervalo no contiene al uno. Asimismo, y de la gráfica 1.1, observamos que un intervalo aproximado de

90% de confianza para la probabilidad de extinción es (0.635, 0.914).

Ejemplificamos ahora el control de la enfermedad simulando un proceso de Galton-Watson de 5 generaciones con distribución de contagio geométrica con media 0.7. Los resultados fueron los siguientes

X_0	X_1	X_2	X_3	X_4	X_5
1	1	2	1	2	1

De (2.4) obtenemos que $\hat{m} = 1$ que no se encuentra en el intervalo (3.3). Por lo tanto no hay peligro de una epidemia.

Observamos finalmente que si se supone una distribución de contagio de Poisson el valor de $\hat{m} = 1.335$ correspondería a una probabilidad de extinción aproximadamente menor a 0.5, comparada con 0.75 para la distribución geométrica. La importancia de precisar la distribución de contagio queda expresada en la gráfica 1.1 en donde para la misma media, la probabilidad de extinción de la distribución geométrica siempre está por arriba de la correspondiente para la distribución de Poisson. Esperamos investigar métodos estadísticos que nos permitan conocer mejor las distribuciones de contagio en un futuro.

N	X_0	X_1	X_2	X_3	X_4	X_5
1	1	1	5	6	5	8
2	1	1	2	2	2	2
3	1	0	0	0	0	0
4	1	1	1	3	2	5
5	1	2	0	0	0	0
6	1	2	4	2	3	2
7	1	0	0	0	0	10
8	1	3	4	6	8	10
9	1	2	3	5	14	16
10	1	5	4	3	7	17
11	1	8	11	20	15	12
12	1	3	1	2	2	2
13	1	1	0	0	0	0
14	1	0	0	0	0	0
15	1	1	2	4	7	8

Tabla 3.2
Simulación de 15 procesos de Galton-Watson con
distribución de contagio geométrica con media 1.3

BIBLIOGRAFIA

1. Basawa I. y B. L. S. Prakasa-Rao (1980). Statistical Inference for Stochastic Processes. Academic Press, New York.
2. Becker N. (1975). The use of mathematical models in determining vaccination policies. Bulletin of the International Statistical Institute, 46,I, 478-490.
3. Jagers P. (1975). Branching Processes with Biological Applications. John Wiley & Sons, New York.
4. Nanthi K. (1983). Statistical Estimation for Stochastic Processes. Queen's papers in Pure and Applied Mathematics No. 62, Queen's University, Ontario.
5. Neyman J. y E. L. Scott (1964). In Stochastic Models in Medicine and Medicine and Biology. J. Gurland editor, 45-55, University of Wisconsin Press, Madison.