

Nº 39  
2EV.

Universidad Nacional Autónoma de México

Facultad de Ciencias

EL PROCESO DE GALTON-WATSON,  
ESTIMACIÓN DE LA PROBABILIDAD DE EXTINCIÓN  
Y LA MEDIA DE REPRODUCCIÓN

T E S I S

que para recibir el título de:

A C T U A R I O

Presenta

DINO MON VÁSQUEZ

México, D.F. 1992

FALLA DE ORIGEN



## **UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso**

### **DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

# Contenido

## INTRODUCCIÓN

<b>1</b>	<b>PRELIMINARES</b>	<b>1</b>
1.1	CADENAS DE MARKOV . . . . .	1
<b>2</b>	<b>PROCESO DE GALTON-WATSON</b>	<b>9</b>
2.1	DESCRIPCIÓN MATEMÁTICA . . . . .	11
2.2	FUNCIÓN GENERADORA DE PROBABILIDAD Y PROPIEDADES ELEMENTALES . . . . .	13
2.3	PROBABILIDAD DE EXTINCIÓN . . . . .	25
2.4	EJEMPLOS IMPORTANTES . . . . .	29
2.5	TEOREMAS LÍMITES . . . . .	35
2.5.1	CASO SUBCRÍTICO . . . . .	36
2.5.2	CASO CRÍTICO . . . . .	40
2.5.3	CASO SUPERCRÍTICO . . . . .	44
2.6	LA PROGENIE TOTAL . . . . .	47
<b>3</b>	<b>ESTIMACIÓN DE LA MEDIA DE REPRODUCCIÓN Y LA PRO- BABILIDAD DE EXTINCIÓN</b>	<b>57</b>
3.1	ESTIMACIÓN DE LA MEDIA DE REPRODUCCIÓN . . . . .	65
3.2	ESTIMACIÓN DE LA PROBABILIDAD DE EXTINCIÓN . . . . .	88

## REFERENCIAS

# INTRODUCCIÓN

La decadencia de las familias de hombres notables en tiempos pasados ha sido objeto de frecuentes observaciones y estudios, los cuales fueron el preámbulo de lo que hoy día se conoce como la Teoría de Procesos de Ramificación. Las primeras observaciones sobre la extinción de familias de hombres importantes, fueron hechas por Thomas Malthus, quien relata en uno de sus ensayos ([4]) que en la ciudad de Berna de 487 familias burguesas 379 se extinguieron en espacio de dos siglos, de 1583 a 1783.

Sir Francis Galton (1822-1911) hace un estudio de la decadencia de la nobleza inglesa y otras familias de hombres insignes; desde un punto de vista eugénico (estudio de todo aquello que afecta la calidad de la progenie humana)[9]. Galton dio explicaciones sociales y biológicas de las frecuentes extinciones de familias que observó. Luego un renombrado botánico suizo Alfonso de Candolle, en su publicación de 1873, volteó la atención de Galton a la posibilidad de una interpretación probabilística del fenómeno ([10], pág. 389) y ese mismo año Galton proporciona una formulación precisa al problema, apareciendo como el " Problema 4001 " en el Educational Time[2]: 'Una nación grande en la cual sólo nos preocuparemos de los varones adultos, de número  $N$ , y de los cuales cada uno lleva diferente apellido, colonizan un distrito. Su ley de población es tal que en cada generación,  $a_0\%$  de los hombres adultos no tienen niños varones que alcancen la edad adulta,  $a_1$  tienen sólo un niño varón,  $a_2$  tienen dos; y así hasta  $a_s$  que tiene cinco. Las preguntas que se formulan son las siguientes:

- 1.- Encontrar qué proporción de los apellidos se va a extinguir después de  $r$  generaciones.
- 2.- Encontrar cuántos casos habrá en el que el mismo apellido sea llevado por  $n$  personas'.

Galton, recibió sólo una respuesta al problema, de una persona que nunca percibió su complejidad, y se la mostró a un amigo, el clérigo y matemático Reverendo H. W. Watson. Éste transformó el problema en uno de iteración de funciones generadoras de probabilidad[3]; de hecho, así es tratado hoy día. Si  $f(x) = \sum p_j x^j$ ,  $p_j = a_j/100$ , es decir la probabilidad de que un padre tenga  $j$  hijos varones que alcancen la edad adulta, entonces Watson define la sucesión recursiva  $f_r = f_{r-1}(f)$ . Concluye que la respuesta a la primera pregunta es el término independiente de  $x$  en  $f_r(x)$  y da el número de apellidos con  $s$  representantes en la  $r$ -ésima generación como el coeficiente de  $x^s$  en

$f_r(x)$  multiplicado por  $N$ . Además Watson descubre, para la función generadora de probabilidad binomial, que la probabilidad  $q$ , de extinción final, debe satisfacer la ecuación  $q = f(q)$ , a partir de aquí agrega erróneamente que siempre  $q = 1$ . Él estaba consciente que esto podría parecer una contradicción al principio Malthusiano de incremento exponencial del total de la población y argumentó con sutileza en forma extensa, que el total de la población puede bien crecer a pesar de que cada apellido específico eventualmente se perderá.

Fue un actuario danés, J. F. Steffensen, quien resolvió totalmente el problema formulado por Galton y publicó el primer análisis completo de la probabilidad de extinción[14]. Este problema se ha denominado en la literatura el *Proceso de Galton-Watson*. También las formulaciones de Steffensen están todavía en uso: Suponga que un individuo y sus descendientes procrean  $n$  hijos con probabilidad  $p_n$ . Entonces la probabilidad  $q$  de que su familia se extinga es la más pequeña raíz no negativa de la ecuación  $s = f(s)$ , de aquí que  $q < 1$  si y sólo si  $m = f'(1) > 1$ .

En el presente, existen en la literatura procesos de ramificación más generales que el proceso de Galton-Watson, como: Procesos de ramificación de tiempo continuo de nacimiento y muerte, procesos de ramificación donde se permite la dependencia entre vida y reproducción, procesos con individuos de tipos diferentes, procesos de ramificación con inmigración, etc.

Parámetros muy importantes en un proceso de Galton-Watson, son la media del número de descendientes por individuo y la probabilidad de extinción del proceso. Por lo cual, dos problemas que merecen especial atención, es la estimación de la media de reproducción basada sobre una historia completa del tamaño de la población, y la estimación de la probabilidad de extinción, basada sobre el tamaño actual de la población y la distribución de descendientes.

El objetivo principal de este trabajo, es presentar, algunos resultados generales de los estudios que se han realizado acerca de la estimación de esos dos parámetros y extender las conclusiones del tratamiento Bayesiano. Un compendio de todo lo que se ha desarrollado en cuanto a la estimación de parámetros en los procesos estocásticos, en particular los de ramificación se puede encontrar en[57].

Es natural presentar como preludeo al problema de estimación, los principales resultados probabilísticos que sobre el proceso de Galton-Watson existen, esto es, un estudio general de la función generadora de probabilidad y su relación con los diferentes

parámetros que el proceso involucra, así como una revisión de Teoremas Límite sobre el comportamiento asintótico del mismo. Todo esto permite abordar la problemática de la estimación mencionada anteriormente.

Históricamente el problema de estimar la media de reproducción y la probabilidad de extinción, ha sido resuelto utilizando Inferencia Estadística Clásica. Casos particulares como el de una distribución de descendientes parametrizada por una ley Poisson o el de descendencia binaria, han sido resueltos utilizando técnicas de estimación Bayesiana. El presente trabajo considera la aplicación de las técnicas de estimación Bayesianas en el caso general, esto es, sin parametrizar la distribución de descendientes, analiza las dificultades del procedimiento y obtiene la solución para una clase de problemas restringidos. El lector podrá también observar, la dificultad en la aplicación de los métodos Bayesianos de inferencia para estimar la probabilidad de extinción.

Para la comprensión del material que aquí se presenta se suponen conocimientos básicos sobre Procesos Estocásticos, en particular Cadenas de Markov, también conocimientos de Estadística Clásica y Bayesiana.

El trabajo está organizado de la siguiente manera: El capítulo 1, presenta resultados generales sobre Cadenas de Markov que se supondrán conocidos a lo largo del trabajo; El capítulo 2, describe el Modelo de Galton-Watson, da un estudio general de la función generadora de probabilidad y de Teoremas Límite sobre el comportamiento asintótico del proceso, así como un breve estudio sobre la progenie total del proceso; En el capítulo 3 se exponen métodos de estimación y resultados asintóticos para la media de reproducción y la probabilidad de extinción.

# Capítulo 1

## PRELIMINARES

### 1.1 CADENAS DE MARKOV

En esta sección y sus apartados se presentan resultados básicos acerca de Cadenas de Markov que se suponen conocidos para el lector. Es por esta razón que las definiciones y resultados, se presentan de una manera general y sin entrar en detalles. Para un mayor análisis sobre este tema, el lector puede revisar ([49, 40]).

La propiedad de Markov se define precisamente con la siguiente ecuación:

$$P(X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_n = x_n) = P(X_{n+1} = x_{n+1} | X_n = x_n) \quad (1.1.1)$$

para cualquier selección de enteros no-negativos  $n$  y los números  $x_0, \dots, x_{n+1}$ , cada uno de los cuales es elemento de  $\mathcal{F}$ . Las probabilidades condicionales  $P(X_{n+1} = y | X_n = x)$  son llamadas las probabilidades de transición de la cadena. A los sistemas que poseen esta propiedad se les llama Cadenas de Markov. En este trabajo se estudiarán Cadenas de Markov que tienen probabilidades de transición estacionaria, es decir, aquéllas tales que  $P(X_{n+1} = y | X_n = x)$  no depende de  $n$ , éstas también son llamadas cadenas de Markov Homogéneas.

### I. FUNCIÓN DE TRANSICIÓN Y DISTRIBUCIÓN INICIAL

Sea  $X_n$ ,  $n \geq 0$ , una cadena de Markov que tiene espacio de estados  $\mathcal{F}$ . La función  $P(x, y)$ ,  $x, y \in \mathcal{F}$ , definida por:

$$P(x, y) = P(X_1 = y | X_0 = x), \quad x, y \in \mathcal{F}, \quad (1.1.2)$$

es llamada la *Función de Transición* de la cadena. Ésta es tal que:

$$P(x, y) \geq 0, \quad x, y \in \mathcal{F}, \quad (1.1.3)$$

y

$$\sum_y P(x, y) = 1, \quad x \in \mathcal{F}. \quad (1.1.4)$$

Dado que la cadena de Markov tiene probabilidades estacionarias, nosotros vemos que:

$$P(X_{n+1} = y | X_n = x) = P(x, y), \quad n \geq 1. \quad (1.1.5)$$

y de la propiedad de Markov se sigue que:

$$P(X_{n+1} = y | X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x) = P(x, y). \quad (1.1.6)$$

En otras palabras, si la cadena de Markov está en el estado  $x$  al tiempo  $n$ , entonces no importa como llegamos a  $x$ , esta tiene probabilidad  $P(x, y)$  de estar en el estado  $y$  en el siguiente paso. Por esta razón los números  $P(x, y)$  son llamadas las probabilidades de transición en *un paso* de la cadena de Markov.

La función  $\pi_0(x)$ ,  $x \in \mathcal{F}$ , definida por

$$\pi_0(x) = P(X_0 = x), \quad x \in \mathcal{F}, \quad (1.1.7)$$

es llamada la distribución inicial de la cadena. Ésta es tal que

$$\pi_0(x) \geq 0, \quad x \in \mathcal{F}. \quad (1.1.8)$$

y

$$\sum_x \pi_0(x) = 1. \quad (1.1.9)$$

La distribución conjunta de  $X_0, \dots, X_n$  puede ser fácilmente expresada en términos de la función de transición y la distribución inicial. Por ejemplo:

$$\begin{aligned} P(X_0 = x_0, X_1 = x_1) &= P(X_0 = x_0)P(X_1 = x_1 | X_0 = x_0) \\ &= \pi_0(x_0)P(x_0, x_1). \end{aligned}$$

También,

$$\begin{aligned} P(X_0 = x_0, X_1 = x_1, X_2 = x_2) &= P(X_0 = x_0, X_1 = x_1)P(X_2 = x_2 | X_0 = x_0, X_1 = x_1) \\ &= \pi_0(x_0)P(x_0, x_1)P(x_1, x_2). \end{aligned}$$

Por inducción se ve fácilmente que

$$P(X_0 = x_0, \dots, X_n = x_n) = \pi_0(x_0)P(x_0, x_1) \cdots P(x_{n-1}, x_n). \quad (1.1.10)$$

## II. CÁLCULOS CON FUNCIONES DE TRANSICIÓN

Sea  $X_n, n \geq 0$ , una cadena de Markov sobre el espacio  $\mathcal{F}$ , que tiene función de transición  $P$ . En este apartado Mostraremos como varias probabilidades condicionales pueden ser expresadas en términos de  $P$ . También podremos definir la función de transición en  $n$ -pasos de la cadena de Markov.

Comenzaremos con la ecuación

$$\begin{aligned} P(X_{n+1} = x_{n+1}, \dots, X_{n+m} = x_{n+m} | X_0 = x_0, \dots, X_n = x_n) \\ = P(x_n, x_{n+1}) \cdots P(x_{n+m-1}, x_{n+m}). \end{aligned} \quad (1.1.11)$$

Para probar (1.1.11) escribimos el lado izquierdo de la ecuación como sigue

$$\frac{P(X_0 = x_0, \dots, X_{n+m} = x_{n+m})}{P(X_0 = x_0, \dots, X_n = x_n)}.$$

Por (1.1.10) el cociente es igual

$$\frac{\pi_0(x_0)P(x_0, x_1) \cdots P(x_{n+m-1}, x_{n+m})}{\pi_0(x_0)P(x_0, x_1) \cdots P(x_{n-1}, x_n)},$$

lo cual se reduce al lado derecho de (1.1.11).

Aquí es conveniente enunciar el siguiente Lema:

**LEMA 1.1.1** *Sea  $(\Omega, \mathcal{A}, P)$  un espacio de probabilidad y suponga que los conjuntos que se mencionan abajo están todos en  $\mathcal{A}$ .*

- (i) *Si  $D_i$  son ajenos y  $P(C|D_i) = p$  independientemente de  $i$ , entonces  $P(C|\cup_i D_i) = p$ .*
- (ii) *Si  $C_i$  son ajenos, entonces  $P(\cup_i C_i|D) = \sum_i P(C_i|D)$ .*

La demostración del Lema se basa en el uso de probabilidad condicional.

Podemos reescribir (1.1.11) como sigue

$$\begin{aligned} P(X_{n+1} = y_1, \dots, X_{n+m} = y_m | X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x) \\ = P(x, y_1)P(y_1, y_2) \cdots P(y_{m-1}, y_m). \end{aligned} \quad (1.1.12)$$

Sean  $A_0, \dots, A_{n-1}$  subconjuntos de  $\mathcal{F}$ . Se sigue de (1.1.12) y del Lema 1.1.1 (i) que

$$\begin{aligned} P(X_{n+1} = y_1, \dots, X_{n+m} = y_m | X_0 \in A_0, \dots, X_{n-1} \in A_{n-1}, X_n = x) \\ = P(x, y_1)P(y_1, y_2) \cdots P(y_{m-1}, y_m). \end{aligned} \quad (1.1.13)$$

Sean  $B_1, \dots, B_m$  subconjuntos de  $\mathcal{F}$ . Se sigue de (1.1.13) y del Lema 1.1.1 (ii) que

$$\begin{aligned} P(X_{n+1} \in B_1, \dots, X_{n+m} \in B_m | X_0 \in A_0, \dots, X_{n-1} \in A_{n-1}, X_n = x) \\ = \sum_{y_1 \in B_1} \cdots \sum_{y_m \in B_m} P(x, y_1) P(y_1, y_2) \cdots P(y_{m-1}, y_m). \end{aligned} \quad (1.1.14)$$

La función de transición en  $m$ -pasos  $P_m(x, y)$ , se define por

$$P_m(x, y) = \sum_{y_1} \cdots \sum_{y_{m-1}} P(x, y_1) P(y_1, y_2) \cdots P(y_{m-2}, y_{m-1}) P(y_{m-1}, y), \quad (1.1.15)$$

para  $m \geq 2$ , diremos  $P_1(x, y) = P(x, y)$ , y

$$P_0(x, y) = \begin{cases} 1, & x = y \\ 0, & \text{en otro caso.} \end{cases}$$

Vemos que si tomamos  $B_1 = \cdots = B_{m-1} = \mathcal{F}$  y  $B_m = \{y\}$  en (1.1.14) tenemos

$$P_m(x, y) = P(X_{n+m} = y | X_0 \in A_0, \dots, X_{n-1} \in A_{n-1}, X_n = x). \quad (1.1.16)$$

En particular, tomando  $A_0 = \cdots = A_{n-1} = \mathcal{F}$ , vemos que

$$P_m(x, y) = P(X_{n+m} = y | X_n = x). \quad (1.1.17)$$

También se sigue de (1.1.15) que

$$P_m(z, y) = P(X_{n+m} = y | X_0 = x, X_n = z). \quad (1.1.18)$$

Y por análisis del  $n$ -ésimo paso y probabilidad condicional tenemos

$$\begin{aligned} P_{n+m}(x, y) &= P(X_{n+m} = y | X_0 = x) \\ &= \sum_z P(X_n = z | X_0 = x) P(X_{n+m} = y | X_0 = x, X_n = z) \\ &= \sum_z P_n(x, z) P_m(z, y), \end{aligned}$$

Y concluimos por (1.1.17) que

$$P_{n+m}(x, y) = \sum_z P_n(x, z) P_m(z, y). \quad (1.1.19)$$

Esto es lo que en la literatura se conoce como la ecuación de Chapman-Kolmogorov.

Sea  $\pi_0$  la distribución inicial de la cadena de Markov. Dado que

$$\begin{aligned} P(X_n = y) &= \sum_x P(X_0 = x, X_n = y) \\ &= \sum_x P(X_0 = x)P(X_n = y|X_0 = x), \end{aligned}$$

vemos que

$$P(X_n = y) = \sum_x \pi_0(x)P_n(x, y). \quad (1.1.20)$$

Esta ecuación da el cálculo de la distribución de  $X_n$  en términos de la distribución inicial  $\pi_0$  y la función de transición en  $n$ -pasos  $P_n$ .

Es conveniente usar la notación  $P_x(\cdot)$  para denotar las probabilidades de varios eventos definidos en términos de una cadena de Markov que comienza en  $x$ . Así, (1.1.14) puede reescribirse como

$$\begin{aligned} P(X_{n+1} \in B_1, \dots, X_{n+m} \in B_m | X_0 \in A_0, \dots, X_{n-1} \in A_{n-1}, X_n = x) \\ = P_x(X_1 \in B_1, \dots, X_m \in B_m). \end{aligned} \quad (1.1.21)$$

### III. TIEMPOS DE ESPERA

Sea  $A$  un subconjunto de  $\mathcal{F}$ . El tiempo de espera  $T_A$  de  $A$  se define por

$$T_A = \min(n > 0 : X_n \in A)$$

si  $X_n \in A$  para alguna  $n > 0$ , y por  $T_A = \infty$  si  $X_n \notin A$  para toda  $n > 0$ . En otras palabras,  $T_A$  es el primer tiempo positivo para el cual la cadena de Markov está (espera) en  $A$ . En este trabajo se estudiarán sólo tiempos de espera de conjuntos que consisten de un sólo punto. Denotamos el tiempo de espera de un punto  $a \in \mathcal{F}$  por  $T_a$ .

Una importante ecuación que involucra tiempos de espera es dada por

$$P_n(x, y) = \sum_{m=1}^n P_x(T_y = m)P_{n-m}(y, y), \quad n \geq 1. \quad (1.1.22)$$

Para verificar (1.1.23) debemos observar que los eventos  $\{T_y = m, X_n = y\}$ ,  $1 \leq m \leq n$ , son ajenos y que

$$\{X_n = y\} = \bigcup_{m=1}^n \{T_y = m, X_n = y\}.$$

Hemos descompuesto el evento  $\{X_n = y\}$  de acuerdo al tiempo de espera de  $y$ . De esta descomposición tenemos

$$\begin{aligned} P_n(x, y) &= P_x(X_n = y) \\ &= \sum_{m=1}^n P_x(T_y = m, X_n = y) \\ &= \sum_{m=1}^n P_x(T_y = m)P(X_n = y | X_0 = x, T_y = m) \\ &= \sum_{m=1}^n P_x(T_y = m)P(X_n = y | X_0 = x, X_1 \neq y, \dots, X_{m-1} \neq y, X_m = y) \\ &= \sum_{m=1}^n P_x(T_y = m)P_{n-m}(y, y), \end{aligned}$$

y de aquí que (1.1.22) se cumple.

#### IV. ESTADOS TRANSITORIOS Y RECURRENTES

Sea  $X_n, n \geq 0$ , una cadena de Markov con espacio de estados  $\mathcal{F}$  y función de transición  $P$ . Sea

$$\rho_{xy} = P_x(T_y < \infty).$$

Entonces  $\rho_{xy}$  denota la probabilidad de que la cadena de Markov comenzando en  $x$  este en el estado  $y$  en algún tiempo positivo. En particular,  $\rho_{yy}$  denota la probabilidad de que la cadena de Markov comenzando en  $y$  regrese a  $y$ . Un estado  $y$  es llamado recurrente si  $\rho_{yy} = 1$  y transitorio si  $\rho_{yy} < 1$ . Si  $y$  es un estado recurrente, una cadena de Markov que comienza en  $y$  regresa a  $y$  con probabilidad uno. Si  $y$  es un estado transitorio, una cadena de Markov que empieza en  $y$  tiene una probabilidad positiva  $1 - \rho_{yy}$  de nunca regresar a  $y$ . Si  $y$  es un estado absorbente, entonces  $P_y(T_y = 1) = P(y, y) = 1$  y de aquí que  $\rho_{yy} = 1$ ; así un estado absorbente es necesariamente recurrente.

Denotamos por  $1_y(z)$ ,  $z \in \mathcal{F}$ , la función indicadora del conjunto  $\{y\}$  definido por

$$1_y(z) = \begin{cases} 1, & z = y \\ 0, & z \neq y. \end{cases}$$

Denotamos por  $N(y)$  el número de tiempos  $n \geq 1$  para los cuales la cadena está en el estado  $y$ . Dado que  $1_y(X_n) = 1$  si la cadena está en el estado  $y$  al tiempo  $n$  y  $1_y(X_n) = 0$  en otro caso, vemos que

$$N(y) = \sum_{n=1}^{\infty} 1_y(X_n). \quad (1.1.23)$$

Usaremos la notación  $E_x(\cdot)$  para denotar la esperanza de variables aleatorias definidas en términos de una cadena de Markov que empieza en  $x$ . Por ejemplo

$$E_x(1_y(X_n)) = P_x(X_n = y) = P_n(x, y). \quad (1.1.24)$$

Se sigue de (1.1.22) y (1.1.23) que

$$\begin{aligned} E_x(N(y)) &= E_x\left(\sum_{n=1}^{\infty} 1_y(X_n)\right) \\ &= \sum_{n=1}^{\infty} E_x(1_y(X_n)) \\ &= \sum_{n=1}^{\infty} P_n(x, y). \end{aligned}$$

Sea

$$G(x, y) = E_x(N(y)) = \sum_{n=1}^{\infty} P_n(x, y).$$

Entonces  $G(x, y)$  denota el valor esperado del número de visitas a  $y$  para la cadena de Markov que da comienzo en  $x$ .

**TEOREMA 1.1.1** (i) *Sea  $y$  un estado transitorio. Entonces*

$$P_x(N(y) < \infty) = 1$$

y

$$G(x, y) = \frac{\rho_{xy}}{1 - \rho_{yy}}, \quad x \in \mathcal{F},$$

la cual es finita para toda  $x \in \mathcal{F}$ .

(ii) *Sea  $y$  un estado recurrente. Entonces  $P_y(N(y) = \infty) = 1$  y  $G(y, y) = \infty$ . También*

$$P_x(N(y) = \infty) = P_x(T_y < \infty) = \rho_{yy}, \quad x \in \mathcal{F}.$$

Si  $\rho_{x,y} = 0$ , entonces,  $G(x, y) = 0$ , mientras si  $\rho_{x,y} > 0$ , entonces  $G(x, y) = \infty$ .

Para la demostración de este Teorema ver ([40], pág.20)

**OBSERVACIÓN 1.1.1** *Si  $y$  es un estado transitorio, el valor esperado  $G(x, y)$  es finito, por lo cual tiene que ocurrir que cada una de las*

$$P_n(x, y) \rightarrow 0,$$

de donde obtenemos que

$$\lim_{n \rightarrow \infty} P[X_n = y] = 0.$$

## Capítulo 2

# PROCESO DE GALTON-WATSON

Imaginemos objetos que pueden producir o generar un número aleatorio de descendientes de la misma clase (estos quizás pueden ser hombres, bacterias conocidos o neutrones producidos por una reacción en cadena), durante su vida, con una función de probabilidad (ley de reproducción ) digamos:

$$P[\text{número de desc.} = k] = p_k \text{ para } k = 0, 1, 2, \dots$$

donde  $p_k \geq 0$  y  $\sum_{k=0}^{\infty} p_k = 1$ . Si suponemos que todos los descendientes se reproducen independientemente uno de otro e individualmente tienen progenie de acuerdo con la distribución de probabilidad anterior, el proceso  $\{Z_n\}$ , donde  $Z_n$  es el tamaño de la  $n$ -ésima generación, es llamado un proceso de ramificación de GALTON-WATSON.

### *EJEMPLOS DE PROCESOS DE RAMIFICACIÓN DE GALTON- WATSON*

Hay numerosos ejemplos de procesos de Galton-Watson que surgen naturalmente en varias disciplinas científicas. Listaremos algunos de los mas prominentes.

**a- Multiplicador de Electrones:** Un Multiplicador de Electrones, es un aparato que amplifica una corriente débil de electrones. Una serie de láminas son colocadas en el camino de los electrones emitidos por una fuente. Cada electrón, al golpear la primera lámina, genera un número aleatorio de nuevos electrones, los cuales al golpear la siguiente lámina producen otros y así se continúa. Sea  $Z_0$  el número de electrones inicialmente emitidos,  $Z_1$  el número de electrones producidos sobre el impacto en la primera lámina de los  $Z_0$  iniciales; en general sea  $Z_n$

el número de electrones emitidos por impacto en la  $n$ -ésima lámina de los electrones emitidos en la lámina  $(n - 1)$ . La sucesión  $Z_0, Z_1, \dots, Z_n, \dots$  constituye un proceso de ramificación de Galton-Watson.

**b- Reacción en Cadena de Neutrones:** Un núcleo es dividido por una colisión casual con un neutrón. El resultado de la fisión da un número aleatorio de nuevos neutrones. Cada uno de esos neutrones secundarios puede chocar con algún otro núcleo produciendo un número aleatorio de neutrones adicionales y así sucesivamente. En este caso el número inicial de neutrones es  $Z_0 = 1$ . La primera generación de neutrones comprende todos aquellos producidos de la fisión causada por el neutrón inicial. El tamaño de la primera generación es una variable aleatoria  $Z_1$ . En general la población  $Z_n$  en la  $n$ -ésima generación es producida por los choques al azar de  $Z_{n-1}$  neutrones individuales de la  $(n - 1)$ -ésima generación.

**c- Supervivencia de los Apellidos de Familias:** Aquí (como a menudo sucede en la vida) sólo cuentan los descendientes masculinos, los cuales desempeñan el papel de las partículas, y  $p_k$  es la probabilidad de que un varón recién nacido se convierta en progenitor de exactamente  $k$  niños varones. Este esquema supone dos simplificaciones artificiales. La fertilidad está sujeta a tendencias seculares y, por lo tanto, la distribución  $\{p_k\}$  cambia en la realidad de generación en generación. Por otra parte, es seguro que la herencia y el medio ambiente comunes produzcan semejanzas entre hermanos, con lo que se contradice la suposición de independencia estocástica. Este modelo se puede refinar de modo que se tomen en cuenta estas objeciones, pero las características esenciales permanecen inalterables.

En general tenemos un árbol con múltiples ramas que surgen a raíz de los nacimientos de partículas.

## 2.1 DESCRIPCIÓN MATEMÁTICA

Sea  $\{X_{nj}\}_{n \geq 0, j \geq 1}$  variables aleatorias independientes e idénticamente distribuidas (v.a.i.i.d.) enteras con densidad  $\{p_k\}_{k \geq 0}$  (ley de reproducción). Si consideramos que el proceso da inicio con un individuo, ( $Z_0 = 1$ ), entonces podemos construir el número de descendientes de cada generación como sigue:

$$\begin{aligned}
 Z_0 &= 1 \\
 Z_1 &= X_{0,1} (\# \text{ de descendientes de la } 1^{\text{ra}} \text{ partícula}) \\
 Z_2 &= \sum_{j=1}^{Z_1} X_{1,j} \\
 &\vdots \\
 Z_{n+1} &= \sum_{j=1}^{Z_n} X_{n,j}, \quad n \in \mathbb{N}
 \end{aligned} \tag{2.1.1}$$

es claro entonces que el número de partículas en la  $n$ -ésima generación,  $Z_n$ , es una variable aleatoria y evidentemente si  $Z_n = 0$  entonces  $Z_{n+k} = 0$  para toda  $k \geq 0$ . Decimos entonces que 0 es un estado absorbente y si  $Z_n$  toma ese valor, la población se extingue.

Obsérvese que:

$$\begin{aligned}
 P[Z_n = k | Z_0 = 1, Z_1 = k_1, Z_2 = k_2, \dots, Z_{n-1} = k_{n-1}] &= \\
 &= P\left[\sum_{j=1}^{Z_{n-1}} X_{n-1,j} = k \mid Z_0 = 1, \dots, Z_{n-1} = k_{n-1}\right] \\
 &= P\left[\sum_{j=1}^{k_{n-1}} X_{n-1,j} = k \mid Z_0 = 1, \dots, Z_{n-1} = k_{n-1}\right]
 \end{aligned}$$

como las  $X_{k,j}$  son independientes, los dos argumentos son independientes

$$= P\left[\sum_{j=1}^{k_{n-1}} X_{n-1,j} = k\right] \text{ que es la } k_{n-1} \text{ convolución de } \{p_k\} \tag{2.1.2}$$

ya que la  $\sum_{j=1}^{k_{n-1}} X_{n-1,j}$  es independiente de  $Z_k$ , para  $k = 1, \dots, n-1$ .

. Por otro lado

$$\begin{aligned}
 P\left[\sum_{j=1}^{Z_{n-1}} X_{n-1,j} = k \mid Z_{n-1} = k_{n-1}\right] &= P\left[\sum_{j=1}^{k_{n-1}} X_{n-1,j} = k \mid Z_{n-1} = k_{n-1}\right] \\
 &= P\left[\sum_{j=1}^{k_{n-1}} X_{n-1,j} = k\right]
 \end{aligned}$$

y por lo tanto,

$$P[Z_n = k \mid Z_0 = 1, \dots, Z_{n-1} = k_{n-1}] = P[Z_n = k \mid Z_{n-1} = k_{n-1}],$$

la cual no depende de  $n$ . Esto es, el modelo de Galton-Watson es una cadena de Markov homogénea.

Ahora, si comenzamos con una partícula  $Z_0 = 1$  tenemos:

$$P(1, k) = P[Z_1 = k \mid Z_0 = 1] = p_k \quad (2.1.3)$$

y nos interesa estudiar, cómo crece la población a partir de este individuo (el esquema se puede generalizar a tener una población inicial de  $k$  individuos). En particular es de interés determinar si la población se extingue y con qué probabilidad lo hace, esto es calcular  $P[Z_i = 0 \text{ para alguna } i \geq 1]$ . Este problema fue abordado por H.W. Watson, quien convirtió el problema en uno de iteración de funciones generadoras de probabilidad, y fue tan útil su método, que hoy día se sigue aplicando al estudiar los procesos de ramificación. Por lo cual la siguiente sección tratará brevemente la propiedades de la función generadora de probabilidad (f.g.p), y su relación con los procesos de ramificación.

## 2.2 FUNCIÓN GENERADORA DE PROBABILIDAD Y PROPIEDADES ELEMENTALES

La función generadora de probabilidad (f.g.p) de una variable aleatoria  $X$  entera, con función de probabilidad  $\{p_k\}_{k \geq 0}$ , es por definición

$$f(s) = E[s^X] = \sum_{k=0}^{\infty} p_k s^k \quad (2.2.1)$$

para  $-1 \leq s \leq 1$ . Hay dos relaciones importantes que conciernen a la f.g.p, la primera la relaciona la densidad  $\{p_k\}$  como sigue:

$$p_k = \frac{1}{k!} \left. \frac{d^k f(s)}{ds^k} \right|_{s=0},$$

la segunda con la función generadora de momentos  $M(t) = E[e^{tX}]$ . Si  $f(t)$  y  $M(t)$  existen para una variable aleatoria  $X$ , para  $t \in \mathbb{R}$ , entonces

$$M(t) = f(e^t).$$

Sea  $P(i, j)$ ,  $i \geq 0$ ,  $j \geq 0$  definida por

$$P(i, j) = P\{Z_{m+1} = j | Z_m = i\} = P\left[\sum_{k=1}^i X_{m,k} = j\right]$$

Puesto que para cada  $i \geq 0$   $P(i, j)$  es una probabilidad, la función generadora de probabilidad  $g(s)$  de  $P(i, j)$  está dada por

$$g(s) = \sum_{j=0}^{\infty} P(i, j) s^j = [f(s)]^i, \quad (2.2.2)$$

esto es, para cada  $i \geq 1$  fija, es la  $i$ -ésima potencia de  $f(s)$  ya que  $P(i, j)$  es la densidad de una suma de  $i$  v.a.i.i.d.

Nos interesa ahora obtener la función generadora de probabilidad (f.g.p), cuando tenemos que  $n$  generaciones han pasado y comenzamos con una partícula, i.e si

$$P_n(i, j) = P\{Z_n = j | Z_0 = 1\} \implies f.g.p = f_n(s) = \sum_{j=0}^{\infty} P_n(1, j) s^j$$

utilizando la ecuación de Chapman-Kolmogorov (1.1.19) tenemos que

$$\begin{aligned}
 f_{n+1}(s) &= \sum_{j=0}^{\infty} P_{n+1}(1, j) s^j \\
 &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} P_n(1, k) P(k, j) s^j \\
 &= \sum_{k=0}^{\infty} P_n(1, k) \sum_{j=0}^{\infty} P(k, j) s^j \\
 &= \sum_{k=0}^{\infty} P_n(1, k) [f(s)]^k \\
 &= f_n(f(s))
 \end{aligned} \tag{2.2.3}$$

Análogamente se puede obtener que:

$$f_{n+1}(s) = f(f_n(s)) \tag{2.2.4}$$

Resultados importantes los cuales se enunciarán más adelante, en cuanto al comportamiento del proceso de Galton-Watson, son obtenidos de los momentos de la v.a.  $Z_n$ .

Los momentos de variables aleatorias enteras cuando existen, pueden ser expresados en términos de las derivadas de la f.g.p en  $s = 1$ . Para la *media*,  $m$ , tenemos:

$$m = E[Z_1] = \sum_j P(1, j) j = f'(1)$$

y tomando regla de cadena obtenemos:

$$E[Z_n] = \sum_j P_n(1, j) j = f'_n(1) = f'_{n-1}(1) f'(1) = \dots = [f'(1)]^n = m^n \tag{2.2.5}$$

$$\text{ya que } f_n(1) = f(f_{n-1}(1)) \Rightarrow f'_n(1) = f'(f_{n-1}(1)) f'_{n-1}(1) = f'(1) f'_{n-1}(1)$$

$$\text{y } f_{n-1}(1) = \sum_j P(1, j) = 1$$

similarmente usando el hecho de que

$$f''_{n+1}(1) = f''(1) [f'_n(1)]^2 + f'(1) f''_n(1)$$

se prueba que

$$f_n''(1) = f''(1)[m^{2n-2} + m^{2n-3} + \dots + m^{n-1}]$$

y de aquí que, tomando  $\sigma^2 = \text{varianza}(Z_1)$ , se concluye que

$$\text{VAR}(Z_n) = \begin{cases} \frac{\sigma^2 \times m^{n-1} \times (m^{n-1})}{n \times \sigma^2} & \text{Si } m \neq 1 \\ \sigma^2 & \text{Si } m = 1 \end{cases} \quad (2.2.6)$$

otra forma de ver esto, es aplicando los resultados para la media y varianza de sumas aleatorias de variables aleatorias:

Sean  $\xi_k$  y  $N$  v.a con momentos finitos:

$$E[\xi_k] = \mu \quad \text{y} \quad \text{VAR}[\xi_k] = \sigma^2$$

$$E[N] = \nu \quad \text{y} \quad \text{VAR}[N] = \gamma^2$$

lo que se quiere es determinar la media y varianza para  $X = \xi_1 + \dots + \xi_N$ . El resultado se obtiene fácilmente utilizando la esperanza condicional, y a lo que se llega es que

$$E[X] = \mu \cdot \nu \quad (2.2.7)$$

$$\text{VAR}[X] = \nu \cdot \sigma^2 + \mu^2 \cdot \gamma^2 \quad (2.2.8)$$

ahora bien para nuestro caso la ecuación  $Z_{n+1} = \sum_{j=1}^{Z_n} X_{n,j}$  caracteriza la evolución del proceso de ramificación como sucesivas sumas aleatorias de variables aleatorias. Sean

$$m = E[X_{n,j}]_{n \geq 1, j \geq 1} \quad \text{y} \quad \sigma^2 = \text{VAR}[X_{n,j}]_{n \geq 1, j \geq 1},$$

la media y la varianza de la distribución de descendientes  $P[X_{n,j} = k] = p_k$  (ley de reproducción).

Sea  $M(n)$  y  $V(n)$  la media y la varianza de  $Z_n$  bajo la condición inicial de que  $Z_0 = 1$ . Entonces por las ecuaciones (2.2.7) y (2.2.8) tenemos:

$$M(n+1) = mM(n)$$

$$V(n+1) = \sigma^2 M(n) + m^2 V(n)$$

la condición inicial  $Z_0 = 1$  genera una recursividad en las ecuaciones anteriores. Como  $M(0) = 1$  y  $V(0) = 0$ , tenemos que  $M(1) = m \times 1 = m$ ,  $M(2) = m \times M(1) = m^2$  y en general  $M(n) = m^n$  para  $n = 0, 1, 2, \dots$ . De aquí que la media aumenta

geometricamente si  $m > 1$ , decrece geometricamente cuando  $m < 1$  y permanece constante si  $m = 1$ .

Ahora sustituyendo  $M(n)$  dentro de  $V(n+1)$  tenemos:

$$V(n+1) = \sigma^2 m^n + m^2 V(n)$$

con  $V(0) = 0$  tenemos la siguiente recursividad:

$$V(1) = \sigma^2$$

$$V(2) = \sigma^2 \times m + m^2 \times V(1) = \sigma^2 \times m + \sigma^2 \times m^2$$

$$V(3) = \sigma^2 \times m^2 + m^2 \times V(2) = \sigma^2 \times m^2 + \sigma^2 \times m^3 + \sigma^2 \times m^4$$

y en general se tiene:

$$\begin{aligned} V(n) &= \sigma^2 [m^{n-1} + m^n + \dots + m^{2n-2}] \\ &= \sigma^2 m^{n-1} [1 + m + \dots + m^{n-1}] \end{aligned}$$

$$V(n) = \sigma^2 m^{n-1} \begin{cases} n & \text{si } m = 1 \\ \frac{1-m^n}{1-m} & \text{si } m \neq 1 \end{cases}$$

la varianza del tamaño de la población crece geometricamente si  $m > 1$ , crece linealmente si  $m = 1$  y decrece geometricamente si  $m < 1$ .

Mencionamos anteriormente que la extinción de la población ocurre cuando el tamaño de ésta se reduce a cero. Ahora definimos

$$T_0 = \min(n > 0 : Z_n \in \{0\}),$$

como el primer tiempo  $n$  para el cual  $Z_n = 0$ , y es obvio entonces, que  $Z_k = 0 \forall k \geq n$ .

Haciendo uso del análisis del primer paso podemos calcular la probabilidad de que la población muera a lo más en  $n$  generaciones, esta probabilidad se denotará como sigue:

$$q_n = P[T \leq n] = P[Z_n = 0.] \quad (2.2.9)$$

Para calcularla, supongamos que una sola partícula digamos  $Z_0 = 1$  tiene  $X_{01} = k$  descendientes, y cada uno de ellos generará otra descendencia, y si la población original muere a lo más en  $n$  generaciones, entonces cada una de esas  $k$ -líneas de descendientes debe morir a lo más en  $n-1$  generaciones. Ahora las  $k$  sub-poblaciones generadas por

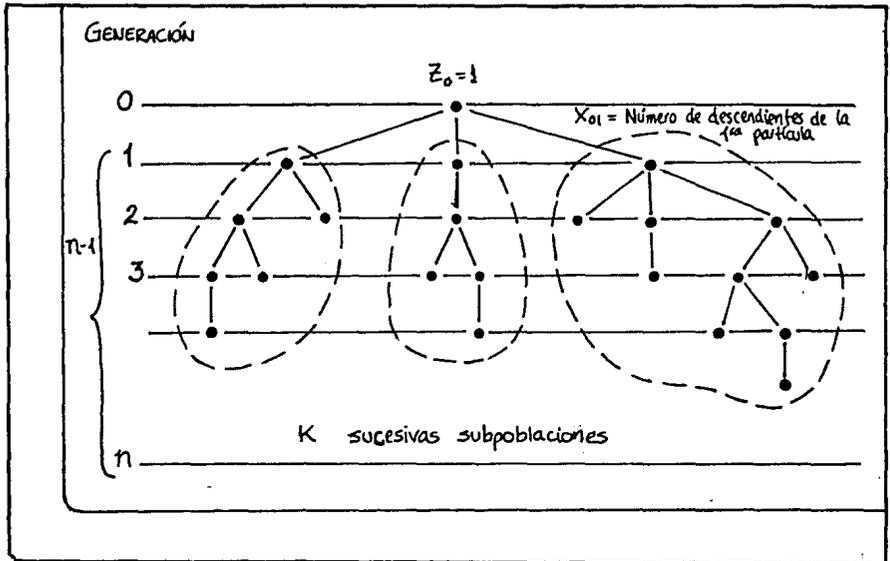


Figura 2.1: El diagrama ilustra que si la población original muere a lo más en  $n$  generaciones, entonces las sub-poblaciones generadas por distintos descendientes iniciales debe toda morir en a lo más  $n - 1$  generaciones

los distintos descendientes de la partícula original son independientes, y ellos tienen las mismas propiedades estadísticas de la partícula original. Por tanto la probabilidad de que una partícula cualquiera muera a lo más en  $n - 1$  generaciones es  $q_{n-1}$ , por definición, y la probabilidad de que las  $k$  sub-poblaciones mueran a lo más en  $n - 1$  generaciones  $[q_{n-1}]^k$ , ya que ellas son independientes, Así:

$$\begin{aligned} q_n &= P\{Z_n = 0\} = \sum_{k=0}^{\infty} P\{Z_n = 0; Z_1 = k\} = \sum_{k=0}^{\infty} P\{Z_n = 0 | Z_1 = k\} P\{Z_1 = k\} \\ &= \sum_{k=0}^{\infty} [q_{n-1}]^k p_k = f(q_{n-1}), \end{aligned} \quad (2.2.10)$$

es claro que  $q_0 = 0$  y  $q_1 = p_0$  (probabilidad de que la partícula original no tenga descendientes).

**EJEMPLO 1:** Supongamos que una partícula no tiene descendientes con probabilidad  $p_0 = \frac{1}{4}$  y tiene dos descendientes con probabilidad  $p_2 = \frac{3}{4}$

entonces por la recursión anterior tenemos :

$$\begin{aligned} q_n &= \sum_{k=0}^2 p_k (q_{n-1})^k = p_0 (q_{n-1})^0 + p_2 (q_{n-1})^2 \\ &= \frac{1}{4} + \frac{3}{4} (q_{n-1})^2 = \frac{1 + 3(q_{n-1})^2}{4} \end{aligned}$$

como  $q_0 = 0$  introduciéndolo en la recursión anterior tenemos :

$$\begin{aligned} q_1 &= .2500 & q_6 &= .3313 \\ q_2 &= .2969 & q_7 &= .3323 \\ q_3 &= .3161 & q_8 &= .3328 \\ q_4 &= .3249 & q_9 &= .3331 \\ q_5 &= .3292 & q_{10} &= .3332 \end{aligned}$$

La función generadora de probabilidad está dada por:

$$f(s) = \frac{1}{4} + \frac{3}{4}s^2.$$



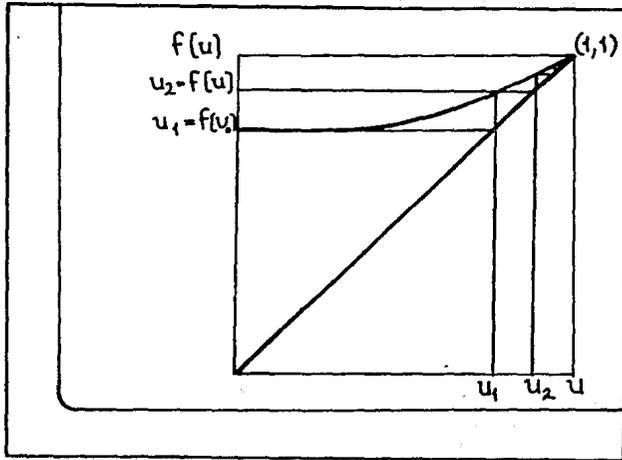


Figura 2.3: función generadora correspondiente a la distribución  $p_0 = 3/4$  y  $p_2 = 1/4$ .

Resolviendo  $q = f(q) = \frac{3}{4} + \frac{1}{4}q^2$  obtenemos.

$$q = \frac{4 \pm \sqrt{16 - 12}}{2} = \begin{cases} 1 \\ 3 \end{cases}$$

la más pequeña solución es  $q = 1$ . Es lógico que esto pasara pues la probabilidad de no tener descendientes es más alta que la probabilidad de tener dos, pero que pasa en el caso extremo donde las dos probabilidades son iguales. El siguiente ejemplo mostrará lo que llega a pasar.

**EJEMPLO 3:** La distribución de descendientes es  $p_0 = \frac{1}{2}$  y  $p_2 = \frac{1}{2}$ .

Resolviendo  $q = f(q) = \frac{1}{2} + \frac{1}{2}q^2$  obtenemos

$$q = \frac{2 \pm \sqrt{4 - 4}}{2} = 1$$

la más pequeña solución es  $q = 1$ . Vemos pues que aún en el caso extremo la población se extingue seguramente. En general la clave es, como lo muestran las figuras (2.2) y (2.3), si la función generadora cruza o no la identidad, esto es, si  $f(s) = s$ . Dentro

de un momento se probará que la función  $f(s)$  es estrictamente convexa en  $[0, 1]$ , de aquí que la gráfica de  $f(s)$  puede intersectar a la recta de  $45^\circ$  en a lo más dos puntos. Sabemos que  $f(1) = 1$ , es decir, en el punto  $(1, 1)$  hay una intersección y tenemos una de las dos raíces, como se puede observar en las figuras (2.2) y (2.3). Si  $m = f'(1) > 1$ , entonces la pendiente de la tangente a la curva  $f(s)$  en  $s = 1$  excede a uno y el caso representado en la figura (2.2) es similar, esto nos dice que hay una solución  $0 < q < 1$ . Si  $m = f'(1) \leq 1$ , entonces la pendiente de la recta tangente en  $s = 1$  es menor o igual a uno y esta situación queda representada por la figura (2.3), en donde necesariamente  $q = 1$ . En otras palabras la extinción es segura si la media del número de descendientes por individuo no excede a uno. Aquí sólo hemos dado un resultado intuitivo en base a las figuras (2.2) y (2.3), por lo cual más adelante formalizaremos este resultado con un Teorema.

Ahora bien, como todas las propiedades de las funciones de transición  $P_n(i, j)$  están contenidas en la función generadora de probabilidad  $f_n(s)$ , y el estudio de teoremas límites que enunciaremos más adelante, también depende de esta función, es necesario precisar un poco más acerca de las propiedades más simples de la función generadora de probabilidad. Hay dos casos que se deben observar sobre la f.g.p: el primero es que la  $f(s)$  puede ser una función lineal si  $p_0 + p_1 = 1$ , en cuyo caso, como se muestra en la sección de ejemplos importantes, iteradas de esta función y la probabilidad de extinción del proceso son fácilmente calculadas. El segundo caso y el más común, es cuando  $f(s)$  es estrictamente convexa, esto se da si  $p_0 + p_1 < 1$  como se verá en el siguiente Lema y aquí, tanto iteradas como la probabilidad de extinción no son triviales de calcular.

**LEMA 2.2.1** *Si  $f(s)$  es la función generadora de probabilidad de un proceso de Gallon-Watson, y suponemos que  $p_0 + p_1 < 1$  y  $p_j < 1$  para toda  $j \geq 0$ , entonces:*

- (i)  $f(\cdot)$  es estrictamente convexa y creciente en  $[0, 1]$ ;
- (ii)  $f(0) = p_0$ ;  $f(1) = 1$ ;
- (iii) Si  $m \leq 1$  entonces  $f(t) > t$  para  $t \in [0, 1]$ ;
- (iv) Si  $m > 1$  entonces  $f(t) = t$  tiene una única raíz en  $[0, 1]$ ;
- (v) Sea  $q$  la más pequeña raíz de  $f(t) = t$  para  $t \in [0, 1]$ . Si  $m > 1$ , y  $q \in [0, 1]$  es tal que  $f(q) = q$ , entonces  $f(t) > t$  para toda  $t \in [0, q]$  y  $f(t) < t$ , para toda  $t \in (q, 1)$ .

**Demostración:** ([48], pág. 4)

(i) p.d.  $f$  creciente en  $[0,1]$ :

$$f(s) = \sum_{j=0}^{\infty} p_j s^j \text{ con } s \in [0,1]$$

si  $s_1 < s_2$  entonces  $s_1^j < s_2^j$ ,  $p_i s_1^i < p_i s_2^i$  y esto implica:

$$\sum_{i=0}^{\infty} p_i s_1^i < \sum_{i=0}^{\infty} p_i s_2^i$$

$f(s_1) < f(s_2)$  por lo tanto es creciente en  $[0,1]$ .

p.d.  $f$  Convexa

Sabemos que si  $f$  es derivable y  $f'$  es creciente  $\Rightarrow f$  es convexa

$$f'(s) = \sum_{j=0}^{\infty} j p_j s^{j-1}$$

como  $f$  es derivable ya que la suma converge absolutamente, basta probar que  $f'$  es creciente, es decir, si  $s_1 < s_2$  entonces  $f'(s_1) \leq f'(s_2)$ . Así, por el mismo argumento anterior, término a término  $f'(s_1) < f'(s_2)$  con desigualdad estricta, ya que  $p_0 + p_1 < 1$ . De igual manera, esto es, utilizando el mismo argumento se demuestra que la derivada de la f.g.p  $f'(s)$ , es creciente y convexa. (ii) p.d.  $f(0) = p_0$ ;  $f(1) = 1$

$$f(s) = \sum_{j=0}^{\infty} p_j s^j = p_0 + p_1 s + p_2 s^2 + \dots = p_0 + \sum_{j=1}^{\infty} p_j s^j$$

de aquí que

$$f(0) = p_0 + \sum_{j=1}^{\infty} p_j (0)^j = p_0, \quad f(1) = \sum_{j=0}^{\infty} p_j = 1.$$

(iii) p.d Si  $m \leq 1$  entonces  $f(t) > t$  para  $t \in [0,1]$

Obsérvese que si  $m = f'(1) \leq 1$  entonces:

$$f'(s) = \sum_{j=0}^{\infty} j p_j s^{j-1} < \sum_{j=0}^{\infty} j p_j \leq 1, \quad s \in [0,1]$$

Por lo tanto, aplicando el Teorema del Valor Medio tenemos que para toda  $t \in [0,1]$ , existe  $s \in (t,1)$  tal que

$$\frac{f(1) - f(t)}{1 - t} = f'(s) < 1,$$

entonces  $f(1) - f(t) < 1 - t$  y por lo tanto  $t < f(t)$ .

(iv) p.d Si  $m > 1$  entonces  $f(t) = t$  tiene una única raíz en  $[0, 1)$

Si  $m = f'(1) > 1$  entonces existe  $\delta > 0$  tal que  $f'(s) > 1$  para  $s \in (1 - \delta, 1)$  ya que la derivada es continua. y si seguimos el procedimiento aplicado en el inciso anterior tenemos que  $f(t) < t$ , para toda  $t \in (1 - \delta, 1)$ .

Por otro lado tenemos que  $f(0) = p_0 \geq 0$ , de aquí que  $t = f(t)$  tiene al menos una solución en el intervalo abierto  $[0, 1)$ .

Ahora bien para probar la unicidad, supongamos que tenemos dos soluciones a la ecuación  $t = f(t)$ , digamos  $s_0$  y  $t_0$  con  $0 \leq s_0 < t_0 < 1$ ; por el Teorema del Valor Medio tenemos que existe  $\xi \in (s_0, t_0)$  tal que

$$f'(\xi) = \frac{f(t_0) - f(s_0)}{t_0 - s_0} = \frac{t_0 - s_0}{t_0 - s_0} = 1,$$

y también existe  $\eta \in (t_0, 1)$  tal que

$$f'(\eta) = \frac{f(1) - f(t_0)}{1 - t_0} = \frac{1 - t_0}{1 - t_0} = 1,$$

por lo tanto  $f'(\xi) = f'(\eta) = 1$  lo cual es una contradicción, ya que  $f$  es estrictamente convexa por (i) y si  $\xi < \eta$  entonces  $f'(\xi) < f'(\eta)$ .

(v) p.d Si  $m > 1$  y  $q \in [0, 1)$  es tal que  $f(q) = q$ , entonces  $f(t) > t$  para toda  $t \in [0, q)$  y  $f(t) < t$ , para toda  $t \in (q, 1)$ .

Por el Lema(2.2.1) inciso (i) vemos que la función generadora de probabilidad es estrictamente convexa y creciente por lo cual si  $t \in [0, q)$  entonces  $f(t) > t$ . Ahora si  $t \in (q, 1)$ , por el mismo argumento usado en (iv), tomando  $q = 1 - \delta$  tenemos que  $f(t) < t$ .

□

**LEMA 2.2.2** Sea  $q$  la más pequeña raíz de  $f(t) = t$  para  $t \in [0, 1)$ . Si  $t \in [0, q)$  entonces  $f_n(t) \uparrow q$  cuando  $n \rightarrow \infty$ . Si  $t \in (q, 1)$  entonces  $f_n(t) \downarrow q$  cuando  $n \rightarrow \infty$ . Si  $t = q$  ó 1 entonces  $f_n(t) = t \forall n$ .

**Demostración:** ([48], pág. 5)

Si  $0 \leq t < q$ , iterando ésta desigualdad por (i), (iii) y (v) del Lema (2.2.1) y recordando

que  $f_n(s) = f(f_{n-1}(s))$ , tenemos:

$$t < f(t) < f(q)$$

$$t < f(t) < f_2(t) < f_2(q)$$

$$t < f(t) < f_2(t) < f_3(t) < f_3(q)$$

⋮

$$t < f(t) < f_2(t) < f_3(t) < \cdots < f_n(t) < f_n(q),$$

además  $f_n(q) = q$ . Para ver esto último, sabemos que  $f_n(q) = f_{n-1}(f(q))$ , y como  $q$  es la más pequeña raíz de  $f(t) = t$ , tenemos  $f_n(q) = f_{n-1}(f(q)) = f_{n-1}(q) = f_{n-2}(f(q)) = f_{n-2}(q) = \cdots = q \forall n \geq 1$ . También por lo anterior tenemos que  $f_n(t) \uparrow L \leq q$ , como  $f$  es continua nosotros podemos tomar

$$L = \lim_{n \rightarrow \infty} f_{n+1}(t) = \lim_{n \rightarrow \infty} f(f_n(t)) = f(\lim_{n \rightarrow \infty} f_n(t)) = f(L)$$

entonces  $L = f(L)$  pero como  $q$  es la más pequeña raíz en  $[0, 1]$  entonces  $L = q$ . Si  $q < t < 1$ , argumentando similarmente se muestra que  $1 > f_n(t) \downarrow L \geq q$  donde  $L = f(L)$ , por lo tanto  $L = q$  por el inciso (iv) del Lema (2.2.1). La tercera parte de este Lema es evidente. □

**OBSERVACIÓN 2.2.1** *La convergencia  $f_n(t) \uparrow q$  para  $t \in [0, q]$  es uniforme dado que  $f_n(0) \leq f_n(t) \leq q$ .*

## 2.3 PROBABILIDAD DE EXTINCIÓN

Como se mencionó en la primera sección de este capítulo

$$P[Z_i = 0 \text{ para alguna } i \geq 1] = P\left[\bigcup_{i=1}^{\infty} Z_i = 0\right],$$

es llamada la probabilidad de *extinción* del proceso y dado que

$$Z_k = 0 \Rightarrow Z_n = 0 \text{ para } n \geq k,$$

la sucesión  $\{[Z_n = 0]\}_{n \geq 0}$  es creciente y por lo tanto tenemos las siguientes igualdades:

$$\begin{aligned} P[Z_n = 0 \text{ para alguna } n \geq 1] &= \\ &= P[(Z_1 = 0) \cup (Z_2 = 0) \cup \dots] \\ &= \lim_{n \rightarrow \infty} P[(Z_1 = 0) \cup (Z_2 = 0) \cup \dots \cup (Z_n = 0)] \\ &= \lim_{n \rightarrow \infty} P[Z_n = 0] = \lim_{n \rightarrow \infty} q_n = q_{\infty} \end{aligned} \quad (2.3.1)$$

pero por las ecuaciones (2.2.4) y (2.2.9) tenemos

$$\begin{aligned} \lim_{n \rightarrow \infty} q_n &= \lim_{n \rightarrow \infty} f(q_{n-1}) = \lim_{n \rightarrow \infty} f(f(q_{n-2})) \\ &= \lim_{n \rightarrow \infty} f(f \cdots f(q_0)) = \lim_{n \rightarrow \infty} f_n(q_0) = \lim_{n \rightarrow \infty} f_n(0) \end{aligned}$$

hemos obtenido entonces que

$$q_{\infty} = \lim_{n \rightarrow \infty} q_n = \lim_{n \rightarrow \infty} f_n(0)$$

mostrando también que el último límite debe existir.

**TEOREMA 2.3.1** Si  $m = E[Z_1] \leq 1$ , la probabilidad de extinción es  $q = 1$ . Si  $m = E[Z_1] > 1$ , entonces la probabilidad de extinción  $q$  es la única solución no negativa menor que 1 de la ecuación  $f(s) = s$ .

**Demostración:** ([28], pág. 7)

Sabemos que  $q_{\infty} = \lim_{n \rightarrow \infty} f_n(0)$ , y por el Lema (2.2.2)  $f_n(0) \rightarrow q$  tomando  $t = 0$ . De aquí que  $q_{\infty} = q$ . Si  $m \leq 1$  por el inciso (iii) del Lema (2.2.1) tenemos que  $q = 1$ . Si

$m > 1$  por el inciso (iv) del mismo Lema que  $q$  es la única solución no negativa menor que uno en el intervalo  $[0, 1)$  de la ecuación  $q = f(q)$ .

□

De aquí en adelante podemos escribir  $q_\infty = q$  para la probabilidad de extinción, la más pequeña raíz de la ecuación  $f(t) = t$  para  $t \in [0, 1]$ .

WATSON [3] dedujo que  $q$  era raíz de la ecuación  $t = f(t)$ , pero falló al dar la noticia que si  $m > 1$ , la raíz relevante es menor que 1.

**DEFINICIÓN 1** *En un proceso de Galton-Watson podemos distinguir tres casos:*

(i) **CRÍTICO:** cuando la media de reproducción es igual a uno.

(ii) **SUB-CRÍTICO:** cuando la media de reproducción es menor a uno.

(iv) **SUPER-CRÍTICO:** cuando la media de reproducción es mayor a uno.

Todavía podemos decir más sobre el comportamiento de nuestro proceso con el siguiente Teorema:

**TEOREMA 2.3.2** *Independientemente de cual sea el valor finito de la media de reproducción tenemos que:*

$$\lim_{n \rightarrow \infty} P[Z_n = k] = 0, \quad k = 1, 2, \dots$$

Mas aún  $Z_n \rightarrow \infty$  con probabilidad  $1 - q$  y  $Z_n \rightarrow 0$  con probabilidad  $q$ .

Este teorema dice que el proceso de Galton- Watson es inestable. La población se extingue o crece indefinidamente. De hecho para el caso *subcrítico* y *crítico*, sabemos que el proceso se extingue seguramente, que es uno de los dos casos descritos en el enunciado del Teorema. Pero para el caso *supercrítico* lo que tenemos es una probabilidad de extinción menor que uno, y por el enunciado del Teorema vemos que esta probabilidad de extinción  $q$ , es muy cercana a cero o muy cercana a uno, ya que la población a la larga nunca se encontrara con un número fijo o estable de individuos mayor que cero, es decir es una cadena de Markov con un sólo estado absorbente.

**Demostración:** ([28], pág 8)

Primero mostraremos que cada uno de los estados  $k = 1, 2, \dots$  es transitorio; esto es, si nosotros tomamos  $R_k = P[Z_{n+j} = k, \text{ p.a } j \geq 1 | Z_n = k]$  entonces  $R_k < 1$ . Ya que por

lo visto en el Capítulo (1) si un estado es transitorio, la probabilidad de que a la larga encontremos a la cadena en este estado es cero. Y si para todo estado  $k$  tenemos esta propiedad, el proceso de Galton-Watson será inestable. Para ver esto, observamos que si  $p_0 = 0$ , entonces

$$\begin{aligned} R_k &= P[Z_{n+j} = k, \text{ p.a. } j \geq 1 | Z_n = k] \\ &= P[Z_{n+1} = k | Z_n = k] + P[Z_{n+1} \neq k, Z_{n+j} = k \text{ p.a. } j \geq 2 | Z_n = k] \end{aligned}$$

dado que  $p_0 = 0$

$$\begin{aligned} &= P[Z_{n+1} = k | Z_n = k] + P[Z_{n+1} > k, Z_{n+j} = k \text{ p.a. } j \geq 2 | Z_n = k] \\ &= P[Z_{n+1} = k | Z_n = k] \text{ porque } p_0 = 0 \\ &= P(k, k) = p_1^k < 1. \end{aligned}$$

Si  $p_0 > 0$  sabemos que:

$$\begin{aligned} P[Z_{n+j} \neq k, j \geq 1 | Z_n = k] &\geq P[Z_{n+j} = 0, j \geq 1 | Z_n = k] \\ &= P[Z_{n+1} = 0 | Z_n = k] \\ &= P(k, 0) \\ &= p_0^k. \end{aligned}$$

donde en el segundo paso se usó el hecho de que

$$\{Z_{n+1} = 0 | Z_n = k\} \subset \{Z_{n+2} = 0 | Z_n = k\} \subset \{Z_{n+3} = 0 | Z_n = k\} \subset \dots$$

Entonces

$$\begin{aligned} R_k &= P[Z_{n+j} = k, \text{ p.a. } j \geq 1 | Z_n = k] \\ &= 1 - P[Z_{n+j} \neq k, j \geq 1 | Z_n = k] \\ &\leq 1 - p_0^k < 1 \end{aligned}$$

de aquí que  $k = 1, 2, \dots$  son estados transitorios. Dado que  $\{Z_n\}$  es una cadena de Markov, por lo obtenido en el capítulo 1, tenemos que:

$$\lim_{n \rightarrow \infty} P[Z_n = k] = 0.$$

Y como sabemos que  $q = P[Z_n \rightarrow 0]$  por la ecuación y Teorema (2.3.1) la prueba queda concluida. □

Este comportamiento inestable descrito por éste último Teorema es diferente al comportamiento en poblaciones biológicas, las cuales tienden a alcanzar un estado de balance con su medio ambiente. Sin embargo el proceso de Galton-Watson se ha usado para representar tempranos estados en el desarrollo familiar, antes de que el ambiente esté saturado.

## 2.4 EJEMPLOS IMPORTANTES

Algunas funciones generadoras de probabilidad son fáciles de iterar, el siguiente ejemplo nos muestra uno de estos casos. Cabe señalar que la función generadora de probabilidad  $f(s)$  que presentamos a continuación no es estrictamente convexa ya que no se cumple el supuesto de que  $p_0 + p_1 < 1$ .

*Ejemplo 1:* Sea  $f(s) = q + ps$ , donde  $0 < p < 1$  y  $p + q = 1$ . El proceso de ramificación asociado es un proceso de muerte pura. Dentro de cada período cada individuo muere con probabilidad  $q$  y sobrevive con probabilidad  $p$ . Las iteradas  $f_n(s)$  en este caso son completamente determinadas como sigue:  $f_2(s) = q + p(q + ps) = 1 - p^2 + p^2s$  e iterando, obtenemos  $f_n(s) = 1 - p^n + p^n s$ . Si nosotros seguimos la ecuación de la f.g.p para la  $n$ -ésima generación dado que se tiene una población inicial de tamaño  $k$ , tenemos:  $[f_n(s)]^k = [1 - p^n + p^n s]^k$ . La distribución del tiempo  $T$  de extinción puede ser determinada de la f.g.p como sigue:

$$\begin{aligned} P\{T = n | Z_0 = k\} &= P\{Z_n = 0 | Z_0 = k\} - P\{Z_{n-1} = 0 | Z_0 = k\} \\ &= [f_n(0)]^k - [f_{n-1}(0)]^k \\ &= (1 - p^n)^k - (1 - p^{n-1})^k \end{aligned} \quad (2.4.1)$$

ya que

$$P\{T = n\} = P\{T \leq n\} - P\{T \leq n - 1\}.$$

La probabilidad de extinción  $q = \lim_{n \rightarrow \infty} f_n(0) = 1$ .

*Ejemplo 2:* Distribución de descendientes geométrica

Consideremos un proceso de ramificación cuyos descendientes siguen una distribución geométrica modificada definida por

$$p_k = bc^{k-1} \quad \text{para } k = 1, 2, \dots \quad (2.4.2)$$

y  $p_0 = 1 - \sum_{k=1}^{\infty} p_k$ , donde  $b, c > 0$  y  $b + c \leq 1$ . Cuando  $b = c(1-c)$ , esta es la distribución geométrica usual  $p_k = c^k(1-c)$ , para  $k = 0, 1, 2, \dots$ . Cuando  $b = (1-c)$ , entonces esta es la distribución geométrica con  $p_0 = 0$  y  $p_k = c^{k-1}(1-c)$ , para  $k = 1, 2, \dots$ . Es posible determinar explícitamente un número de cantidades relevantes para un proceso de ramificación, utilizando una distribución de descendientes geométrica generalizada.

Primero,

$$p_0 = 1 - b \sum_{k=1}^{\infty} c^{k-1} = 1 - \frac{b}{1-c}$$

la correspondiente f.g.p es:

$$f(s) = 1 - \frac{b}{1-c} + bs \sum_{k=1}^{\infty} (cs)^{k-1} = \frac{1 - (b+c)}{1-c} + \frac{bs}{1-cs} = \frac{(1-b-c) + s(c^2 + b-c)}{(1-c) + s(c^2 - c)}$$

la media  $f'(1) = b/(1-c)^2$ .

Notemos que  $f(s)$  tiene la forma de una transformación lineal fraccional:

$$f(s) = \frac{\alpha + \beta s}{\gamma + \delta s} \quad (2.4.3)$$

Es importante recalcar las propiedades de una transformación lineal fraccional:

(i) Las iteradas de una transformación lineal fraccional son de nuevo transformaciones lineales fraccionales, para  $f(s)$  definida por (2.4.3), simple álgebra da:

$$f(f(s)) = \frac{\alpha(\gamma + \beta) + (\alpha\delta + \beta^2)s}{\alpha\delta + \gamma^2 + \delta(\gamma + \beta)s} \quad (2.4.4)$$

(ii) Siempre existen dos soluciones finitas (posiblemente idénticas) para la ecuación  $f(s) = s$ . Las soluciones son llamadas puntos fijos de  $f$ . Si  $f(s)$  es una f.g.p, entonces  $s_1 = 1$  es uno de los puntos fijos y nosotros podríamos ver que otro punto fijo  $s_0$  es menor que uno, igual a uno, o mayor que uno, de acuerdo a si  $f'(1)$  es mayor que, igual o menor que uno.

Para la f.g.p dada en (2.4.3), se puede verificar por álgebra directa que el segundo punto para  $c > 0$  y  $c + b < 1$  es:

$$s_0 = \frac{1 - b - c}{c(1 - c)}$$

(iii) Para cualquiera dos puntos  $s_i$ ,  $i = 0, 1$ , es fácil ver que

$$\frac{f(s) - f(s_i)}{s - s_i} = \frac{\gamma\beta - \alpha\delta}{(\gamma + \delta s)(\gamma + \delta s_i)}$$

de aquí que

$$\frac{f(s) - f(s_0)}{f(s) - f(s_1)} = \left( \frac{\gamma + \delta s_1}{\gamma + \delta s_0} \right) \left( \frac{s - s_0}{s - s_1} \right) \quad (2.4.5)$$

si nosotros ahora tomamos  $s_0$  y  $s_1$  los dos puntos fijos distintos de  $f(\cdot)$  y escribimos  $w = f(s)$ , tenemos:

$$\frac{w - s_0}{w - s_1} = k \left( \frac{s - s_0}{s - s_1} \right) \quad (2.4.6)$$

donde  $k$  puede ser calculada de (2.4.5).

Usando (2.4.6) nosotros facilmente obtenemos las iteradas  $f_n(s) = w_n$  de  $f(s)$ :

$$\frac{w_2 - s_0}{w_2 - s_1} = k \frac{w_1 - s_0}{w_1 - s_1} = k \left( k \frac{s - s_0}{s - s_1} \right)$$

y en general

$$\frac{w_n - s_0}{w_n - s_1} = k^n \left( \frac{s - s_0}{s - s_1} \right) \quad (2.4.7)$$

para la f.g.p de una distribución geométrica dada por (2.4.4), notando que los puntos fijos son  $s_0 = (1 - b - c)/(1 - c)$  y  $s_1 = 1$  obtenemos:

$$k = \frac{(1 - c)^2}{b} = \frac{1}{m}$$

donde  $m$  es la media de una distribución geométrica. Para  $m \neq 1$  los dos puntos fijos  $s_0$  y  $1$  son diferentes; de aquí que, resolviendo para  $w_n$  en (2.4.7) tenemos

$$w_n = \frac{s_0 - \left(\frac{1}{m^n}\right) \left[ (s - s_0)/(s - 1) \right]}{1 - \left(\frac{1}{m^n}\right) \left[ (s - s_0)/(s - 1) \right]}, \quad m \neq 1$$

la cual puede ser escrita en la forma:

$$f_n(s) = 1 - m^n \left( \frac{1 - s_0}{m^n - s_0} \right) + \frac{m^n \left[ (1 - s_0)/(m^n - s_0) \right]^2 s}{1 - \left[ (m^n - 1)/(m^n - s_0) \right] s} \quad (2.4.8)$$

entonces las probabilidades de extinción de la  $n$ -ésima generación son:

$$P[Z_n = 0] = f_n(0) = 1 - m^n \left( \frac{1 - s_0}{m^n - s_0} \right)$$

note que esta expresión converge a  $s_0$  cuando  $n \rightarrow \infty$  si  $m > 1$  y a  $1$  si  $m < 1$ .

La probabilidad de que en la  $n$ -ésima generación la población tenga un tamaño, i.e  $P[Z_n = k]$ ,  $k = 1, 2, \dots$  puede ser calculada por simple expansión en serie de potencias

de (2.4.8). Podemos calcular también la probabilidad de que el tiempo de extinción  $T$  sea menor ó igual a  $n$ , como sigue:

$$\begin{aligned} P[T \leq n] &= P[Z_n = 0] = f_n(0) \quad \text{y} \\ p[T = n] &= P[T \leq n] - P[T \leq n-1] = f_n(0) - f_{n-1}(0) \end{aligned}$$

en el caso  $m \neq 1$ , tenemos:

$$\begin{aligned} P[T = n] &= 1 - m^n \left( \frac{1-s_0}{m^n - s_0} \right) - 1 + m^{n-1} \left( \frac{1-s_0}{m^{n-1} - s_0} \right) \\ &= m^{n-1} s_0 \frac{(m-1)(1-s_0)}{(m^n - s_0)(m^{n-1} - s_0)} \quad \text{para } n = 1, 2, \dots \end{aligned}$$

si  $m = 1$ , entonces  $b = (1-c)^2$  y la ecuación  $f(s) = s$  tiene una raíz doble  $s = 1$  y no hay otra raíz. De hecho

$$f(s) = c + \frac{(1-c)^2 s}{1-cs} = \frac{c - (2c-1)s}{1-cs}$$

entonces

$$\begin{aligned} f_2(s) &= f(f(s)) \\ &= \frac{c - (2c-1) \left[ \frac{c - (2c-1)s}{1-cs} \right]}{1 - c \left[ \frac{c - (2c-1)s}{1-cs} \right]} \\ &= \frac{2c - (3c-1)s}{1+c-2cs} \end{aligned}$$

y por inducción

$$f_n(s) = \frac{nc[(n+1)c-1]s}{1+(n-1)c-ncs}$$

en el caso  $m=1$  nosotros tenemos las probabilidades de extinción como sigue:

$$P[Z_n = 0] = f_n(0) = \frac{nc}{1+(n-1)c} \quad \text{para } n = 1, 2, \dots$$

y el tiempo de extinción  $T$  tiene una distribución

$$\begin{aligned} P[T = n] &= f_n(0) - f_{n-1}(0) \\ &= \frac{nc}{1+(n-1)c} - \frac{(n-1)c}{1+(n-2)c} \\ &= \frac{c(1-c)}{[1+(n-1)c][1+(n-2)c]} \end{aligned}$$

*Ejemplo 3:* El más importante proceso de ramificación dentro de la biología son aquellos de descendencia binaria. En particular un proceso de Galton-Watson es binario si la ley de reproducción satisface  $p_0 = 1 - p$ ,  $p_2 = p$  para  $0 \leq p \leq 1$ .

$$P(j, k) = \begin{cases} 0 & \text{si } k \text{ es impar} \\ \binom{j}{k/2} p^{k/2} (1-p)^{j-k/2} & \text{si } k \text{ es par} \end{cases}$$

$\binom{j}{i}$  se interpretan como cero si  $i > j$ . Todo esto ya que:

$$P(j, k) = P\{Z_{n+1} = k | Z_n = j\} = P\left[\sum_{i=1}^j X_{n,i} = k\right] = P\{Y = k\}$$

donde

$$Y = \sum_{i=1}^j X_{n,i} \text{ la suma de } j \text{ v.a.i.i.d}$$

$$X_{n,i} = \begin{cases} 2 & \text{con probabilidad } p \\ 0 & \text{con probabilidad } 1 - p \end{cases}$$

y ahora tomemos  $Y' = \sum_{i=1}^j \frac{X_{n,i}}{2}$

$$\frac{X_{n,i}}{2} = \begin{cases} 1 & \text{con probabilidad } p \\ 0 & \text{con probabilidad } 1 - p \end{cases}$$

la cual resulta ser una v.a Bernoulli, de aquí que:

$$P\{Y = k\} = P\{Y' = k/2\} = \binom{j}{k/2} p^{k/2} (1-p)^{j-k/2} \text{ si } k \text{ es par}$$

ya que la  $Y'$  es una suma de  $j$  v.a Bernoulli con parámetro  $p$  y su distribución corresponde a una v.a Binomial  $(j, p)$ .

Por la ecuación (1.1.10) en el Capítulo 1, tenemos:

$$P\{Z_0 = i_0, Z_1 = i_1, \dots, Z_n = i_n\} = P\{Z_0 = i_0\}P\{i_0, i_1\}P\{i_1, i_2\} \cdots P\{i_{n-1}, i_n\}$$

de aquí que

$$P\{Z_1 = 2k_1, Z_2 = 2k_2, \dots, Z_n = 2k_n\} =$$

$$\begin{aligned}
&= \sum_{i=1}^{\infty} P\{Z_0 = i, Z_1 = 2k_1, Z_2 = 2k_2, \dots, Z_n = 2k_n\} \\
&= P\{Z_0 = 1, Z_1 = 2k_1, Z_2 = 2k_2, \dots, Z_n = 2k_n\} \\
&= P(Z_0 = 1)P(2k_0, 2k_1)P(2k_1, 2k_2) \cdots P(2k_{n-1}, 2k_n) \\
&= \binom{2k_0}{2k_1/2} p^{2k_1/2} (1-p)^{2k_0-2k_1/2} \cdots \binom{2k_{n-1}}{2k_n/2} p^{2k_n/2} (1-p)^{2k_{n-1}-2k_n/2} \\
&= \prod_{i=1}^n \binom{2k_{i-1}}{k_i} p^{\sum_{i=1}^n k_i} (1-p)^{2 \sum_{i=0}^{n-1} k_i} (1-p)^{-\sum_{i=1}^n k_i} \text{ con } 2k_0 = 1 \\
&= \prod_{i=1}^n \binom{2k_{i-1}}{k_i} \left(\frac{p}{1-p}\right)^{\sum_{i=1}^n k_i} (1-p)^{2 \sum_{i=0}^{n-1} k_i} \tag{2.4.9}
\end{aligned}$$

## 2.5 TEOREMAS LÍMITES

Dado que  $Z_n = i$ , el proceso  $\{Z_{n+k}; k = 0, 1, \dots\}$  es la suma de  $i$  independientes copias de los procesos  $\{Z_0 = 1, Z_1, Z_2, \dots\}$ . Usando la propiedad de Markov tenemos:

$$\begin{aligned} E\{Z_{n+k}|Z_n = i_n, Z_{n-1} = i_{n-1}, \dots, Z_1 = i_1, Z_0 = i_0\} \\ = E\{Z_{n+k}|Z_n = i_n\} = i_n E\{Z_k|Z_0 = 1\} = i_n m^k. \end{aligned} \quad (2.5.1)$$

Sea

$$W_n \equiv Z_n m^{-n},$$

entonces

$$\begin{aligned} H(i_n) &= E\{W_{n+k}|W_n = i_n\} = \frac{1}{m^{n+k}} E\{Z_{n+k}|(Z_n/m^n) = i_n\} \\ &= \frac{1}{m^{n+k}} E\{Z_{n+k}|Z_n = i_n m^n\} = \frac{i_n m^n m^k}{m^{n+k}} = i_n. \end{aligned} \quad (2.5.2)$$

Y de aquí que:

$$E\{W_{n+k}|W_n\} = H(W_n) = W_n \text{ c.s.} \quad (2.5.3)$$

Es decir el proceso es una martingala. Lo cual nos lleva a enunciar el siguiente Teorema:

**TEOREMA 2.5.1** Si  $0 < m < \infty$ ,  $W_n = Z_n m^{-n}$ , y  $\mathcal{F}_n$  es la  $\sigma$ -álgebra generada por  $Z_0, Z_1, Z_2, \dots, Z_n$ , entonces la sucesión  $\{W_n; \mathcal{F}_n; n = 0, 1, 2, \dots\}$  es una martingala, además como  $W_n \geq 0$ , existe una v.a.  $W$  tal que

$$\lim_{n \rightarrow \infty} W_n = W \quad \text{c.s.} \quad (2.5.4)$$

**Demostración:** ([48], pág. 9)

La prueba es clara por (2.5.3) y el hecho de que cualquier martingala no negativa converge con probabilidad 1. □

De la definición de  $W_n$  y (2.5.4) observamos que  $Z_n(w)$  crece como  $m^n W(w)$ , esto es el análogo estocástico de lo que se llama Ley Maltusiana del crecimiento geométrico de poblaciones. Aunque el teorema anterior da un claro resultado con sólo una hipótesis

muy débil; esto no es muy satisfactorio en el sentido de que no estamos diciendo nada acerca de la v.a.  $W$ . Es claro por el Teorema de Fatou que

$$E[W] \leq \liminf E[W_n] = E[Z_0]$$

pero esto no elimina la posibilidad de que  $W \equiv 0$ , en cuyo caso el teorema no dice nada excepto que  $m^n$  es un factor de normalización muy grande. De hecho, si  $m \leq 1$  sabemos que con probabilidad 1  $Z_n = 0$  para una  $n$  suficientemente grande y de aquí que  $W$  es desde luego degenerada en 0. Este teorema realmente adquiere significancia en el caso supercrítico  $m > 1$ , incluso en este caso puede pasar que  $P(W = 0) = 1$ . Sin embargo, si  $\sigma^2 < \infty$  entonces nosotros podemos afirmar que  $W$  es no degenerada. Así tenemos las siguientes caracterizaciones.

## 2.5.1 CASO SUBCRÍTICO

Cuando  $m < 1$  la probabilidad de que el proceso se extinga es 1. Por lo cual, para describir un comportamiento asintótico no trivial en esos casos Kolmogorov[36] y Yaglom[37] propusieron un mecanismo de condicionamiento,  $Z_n$  dado el evento  $\{Z_n > 0\}$ . El principal resultado de esta sección es el Teorema (2.5.1.2), el cual menciona que cuando  $m < 1$  la distribución de  $\{Z_n | Z_n > 0\}$  converge a una distribución propia.

Este resultado fue probado primero por Yaglom[37] bajo restricciones de momentos. La prueba fue simplificada y las suposiciones de momentos fueron removidas por Joffe, Seneta y Vere-Jones[38].

Dado que la f.g.p es derivable podemos aplicar la expansión en serie de Taylor alrededor de uno y obtener un residuo el cual jugará un papel crucial en el análisis del caso subcrítico. La expansión en serie de Taylor para una función  $f(x)$ , está dada por

$$f(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(x_0)(x - x_0)^k}{k!} + R_n(x)$$

donde

$$R_n(x) = \frac{f^{(n)}(c)(x - x_0)^n}{n!}; \quad c \in (x, x_0),$$

entonces

$$f(s) = f(1) + \frac{df(1)}{ds}(s - 1) + R_2$$

$$= 1 - m(1 - s) + r(s)(1 - s) \quad (2.5.1. 1)$$

para alguna función  $r$  tal que  $\lim_{s \rightarrow 1} r(s) = 0$ .

$$\begin{aligned} r(s) &= m - \frac{1 - f(s)}{1 - s} \\ r(0) &= m - (1 - p_0) \geq 0, \\ r(q) &= m - 1 > 0, \text{ si } q < 1, \\ r(1-) &= 0, \text{ y} \\ r'(s) &\leq 0, \quad 0 \leq s < 1. \end{aligned}$$

de aquí que  $r : [0, 1) \rightarrow [0, m]$  decreciente.

**LEMA 2.5.1.1** Para cualquier  $\delta, 0 < \delta < 1$ ,

$$\sum_{k=1}^{\infty} r(1 - \delta^k) < \infty \Leftrightarrow \sum_{k=1}^{\infty} p_k k \log k < \infty$$

la demostración de este lema será omitida ver ([39], pág. 27).

Si reemplazamos  $s$  por  $f_k(s)$  en la relación (2.5.1.1), obtenemos

$$\begin{aligned} f(f_k(s)) &= 1 - m(1 - f_k(s)) + r \circ f_k(s)(1 - f_k(s)) \\ f_{k+1}(s) &= 1 + [1 - f_k(s)](r \circ f_k(s) - m) \\ \frac{1 - f_{k+1}(s)}{1 - f_k(s)} &= m \left\{ 1 - \frac{r \circ f_k(s)}{m} \right\} \end{aligned} \quad (2.5.1. 2)$$

y el producto de esas ecuaciones para  $0 \leq k < n$  es

$$\frac{(1 - f(s))(1 - f_2(s)) \cdots (1 - f_n(s))}{(1 - s)(1 - f(s)) \cdots (1 - f_{n-1}(s))} = m^n \prod_{k=0}^{n-1} \left\{ 1 - \frac{r \circ f_k(s)}{m} \right\}$$

por lo tanto

$$\frac{1 - f_n(s)}{1 - s} = m^n \prod_{k=0}^{n-1} \left( 1 - \frac{r \circ f_k(s)}{m} \right) \quad (2.5.1.3)$$

dado que  $0 \leq r/m \leq 1$ ,  $m^{-n} \{1 - f_n(s)\} / (1 - s)$  decrece a algun límite  $\varphi(s) \geq 0$  cuando  $n \rightarrow \infty$ . En particular

$$P[Z_n > 0] = 1 - f_n(0) \sim m^n \varphi(0). \quad (2.5.1.4)$$

Pero por relación de convergencia de sumas y productos,  $\varphi(0) > 0$  si y sólo si

$$\sum r \circ f_k(0) < \infty.$$

Ahora por el Teorema del Valor Medio sabemos que existe  $\xi \in [s, 1]$  tal que

$$\frac{f(1) - f(s)}{1 - s} = \frac{1 - f(s)}{1 - s} = f'(\xi) \leq f'(1)$$

lo cual implica que  $1 - f(s) \leq m(1 - s)$  y por inducción se sigue que

$$1 - f_k(s) \leq m^k(1 - s) \text{ para cualquier } k. \quad (2.5.1.5)$$

Similarmente, si  $s \in [s_0, 1]$  por el Teorema del Valor Medio existe  $\xi \in [s, 1]$  tal que

$$\frac{1 - f(s)}{1 - s} = f'(\xi) \geq f'(s_0),$$

ya que  $s_0 \leq s \leq \xi \leq 1$  y  $f'(s)$  es creciente y convexa. Por inducción se sigue que

$$1 - f_k(s) \geq (f'(s_0))^k(1 - s); \quad (2.5.1.6)$$

y con  $s_0 = p_0$  (siendo positiva en el caso subcrítico) obtenemos, escribiendo  $a = f'(p_0) > 0$ ,

$$1 - m^k \leq f_k(0) = f_{k-1}(p_0) \leq 1 - a^{k-1}(1 - p_0) \leq 1 - b^k \quad (2.5.1.7)$$

donde  $b = \min\{a, (1 - p_0)\}$ . De la ecuación (2.5.1.7), el Lema anterior y el hecho de que  $r$  es una función decreciente tenemos que

$$\sum r \circ f_k(0) \leq \sum r(1 - m^k) < \infty \Leftrightarrow \sum p_k k \log k < \infty$$

y lo siguiente ha sido probado ([39], pág 29):

**TEOREMA 2.5.1.1** *En el caso subcrítico,*

$$\lim_{n \rightarrow \infty} m^{-n} P[Z_n > 0] = \begin{cases} 0 & \text{si } \sum p_k k \log k = \infty \text{ o } p_0 = 1 \\ \varphi(0) > 0 & \text{en otro caso} \end{cases}$$

**TEOREMA 2.5.1.2** En el caso subcrítico (con  $p_0 < 1$ )

$$\lim_{n \rightarrow \infty} P\{Z_n = k | Z_n > 0\} = b_k$$

existe para  $k \in \mathbb{N}$ ,

$$\begin{aligned} \sum_{k=1}^{\infty} b_k &= 1, \\ \sum_{k=1}^{\infty} k b_k &= 1/\varphi(0) < \infty \Leftrightarrow \sum_{k=1}^{\infty} p_k k \log k < \infty \end{aligned}$$

y si  $g(s) = \sum b_k s^k$ , entonces

$$g \circ f = mg + 1 - m$$

**Demostración:** ([39], pág. 29)

Sea  $g_n(s)$  la f.g.p de  $Z_n$  dado el evento  $Z_n > 0$ , entonces

$$\begin{aligned} g_n(s) &= E[s^{Z_n} | Z_n > 0] = \sum_{k=1}^{\infty} P\{Z_n = k | Z_n > 0\} s^k \\ &= \sum_{k=1}^{\infty} \frac{P\{Z_n = k; Z_n > 0\}}{P\{Z_n > 0\}} s^k \\ &= \sum_{k=1}^{\infty} \frac{P\{Z_n = k\}}{P\{Z_n > 0\}} s^k \\ &= \frac{1}{P\{Z_n > 0\}} \sum_{k=1}^{\infty} P\{Z_n = k\} s^k \\ &= \frac{f_n(s) - f_n(0)}{1 - f_n(0)} = 1 - \frac{1 - f_n(s)}{1 - f_n(0)} \quad (2.5.1. 8) \\ &= 1 - (1 - s) \prod_{k=0}^{n-1} \{1 - r \circ f_k(s)/m\} \{1 - r \circ f_k(0)/m\}^{-1}. \end{aligned}$$

Obsérvese que  $f_k(s) \geq f_k(0)$  y  $r$  no se incrementa, los términos en el producto son mayores que uno. Como  $g_n(s)$  está acotada por abajo por cero y de la última igualdad tenemos que esta función decrece, tenemos pues que  $g_n(s)$  converge. digamos a una función  $g$ , i.e  $g_n \downarrow g$ . Y dado que el límite de  $g_n(s)$  para cada  $s \in [0, 1]$  existe.

el Teorema de continuidad ([50], pág. 289) dice que esta es condición necesaria y suficiente para que exista

$$\lim_{n \rightarrow \infty} P\{Z_n = k | Z_n > 0\} = b_k,$$

cumpliendo automáticamente que

$$\lim_{n \rightarrow \infty} g_n(s) = g(s) = \sum_{k=1}^{\infty} b_k s^k,$$

para  $b_k \geq 0$  y  $\sum b_k \leq 1$ . Evidentemente  $g(0) = 0$ .

De las ecuaciones (2.5.1.8) y (2.5.1.5) con  $s = f_k(0)$  cuando  $n \rightarrow \infty$  tenemos

$$g_n \circ f_k(0) = 1 - \frac{1 - f_k \circ f_n(0)}{1 - f_n(0)} \rightarrow 1 - m^k, \quad (2.5.1.9)$$

que converge a uno si  $k \rightarrow \infty$ .

Como  $g_n(s) \rightarrow g(s)$  cuando  $n \rightarrow \infty$ ,  $f_k(0) \rightarrow q$  cuando  $k \rightarrow \infty$  y  $q = 1$  por estar en el caso subcrítico, se tiene que

$$g(1-) = 1 = \sum b_k.$$

Además, de la ecuación (2.5.1.9) y (2.5.1.4) tenemos

$$\begin{aligned} \sum k b_k = g'(1-) &= \lim_{k \rightarrow \infty} \frac{1 - g \circ f_k(0)}{1 - f_k(0)} \\ &= \lim_{k \rightarrow \infty} \frac{m^k}{(1 - f_k(0))} = \frac{1}{\varphi(0)}, \end{aligned}$$

y

$$g_n \circ f = 1 - \frac{1 - f_{n+1}}{1 - f_{n+1}(0)} \cdot \frac{1 - f \circ f_n(0)}{f - f_n(0)} \rightarrow 1 - (1 - g)m.$$

□

## 2.5.2 CASO CRÍTICO

El caso  $m=1$  es en un sentido el más interesante. El Teorema (2.3.1) dice que  $Z_n \rightarrow 0$  con probabilidad 1. Y también sabemos que el límite de probabilidades  $b_j$  de la sucesión de distribuciones condicionales de  $\{Z_n | Z_n > 0\}$  son cero y de aquí que este proceso diverge a infinito. Podremos ver más adelante que una normalización es necesaria para que el proceso condicionado converja a un límite no degenerado. El proceso modificado podrá entonces converger a una ley límite universal (la cual será de la misma forma para todas las funciones generadoras de probabilidad).

Nuestro acercamiento dependerá de un muy cuidadoso análisis del comportamiento asintótico de  $f_n(t)$ . Vimos también en la sección (3) de este capítulo que  $f_n(t) \uparrow 1$ , y ahora necesitamos conocer la razón de convergencia. El siguiente lemma será básico para este estudio.

**LEMA 2.5.2.1** *Suponga  $m = 1$  y  $\sigma^2 < \infty$ . Entonces*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \left\{ \frac{1}{1 - f_n(s)} - \frac{1}{1 - s} \right\} = \frac{\sigma^2}{2}$$

*uniformemente en  $0 \leq s < 1$ .*

**Demostración:** ([39], pág. 24)

Sea  $0 \leq s < 1$ . Vimos que la f.g.p. la podemos expresar como una expansión en serie de Taylor como sigue:

$$\begin{aligned} f(s) &= f(1) + \frac{df(1)}{ds}(s-1) + \frac{d^2 f(1)}{ds^2} \frac{(s-1)^2}{2} + R_3(s) \\ &= 1 + m(s-1) + \frac{d^2 f(1)}{ds^2} \frac{(s-1)^2}{2} + R_3(s) \end{aligned}$$

ya que

$$\begin{aligned} f(s) &= \sum_{k=0}^{\infty} p_k s^k \\ \frac{df(s)}{ds} \Big|_{s=1} &= m = 1 \\ \frac{d^2 f(s)}{ds^2} \Big|_{s=1} &= \sum_{k=0}^{\infty} k(k-1)p_k \\ &= E[Z_1^2] - E[Z_1] = E[Z_1^2] - 1 = \text{Var}[Z_1] = \sigma^2 \end{aligned}$$

esto último porque si  $m = 1$ ,  $m = E[Z_1] = (E[Z_1])^2$ . De aquí que

$$\begin{aligned} f(s) &= 1 + (s-1) + \frac{\sigma^2(s-1)^2}{2} + R_3(s) \\ &= s + \frac{\sigma^2(1-s)^2}{2} + \frac{f^3(c)(s-1)^3}{3!} \\ &= s + \frac{\sigma^2(1-s)^2}{2} + (1-s)^2 \underbrace{\left[ \frac{f^3(c)(s-1)}{6} \right]}_{r(s)} \\ &= s + \frac{\sigma^2(1-s)^2}{2} + (1-s)^2 r(s) \end{aligned}$$

para alguna función  $r$  tal que  $\lim_{s \rightarrow 1} r(s) = 0$ .

$$\begin{aligned} \frac{1}{1-f(s)} - \frac{1}{1-s} &= \frac{f(s)-s}{(1-f(s))(1-s)} \\ &= \frac{\frac{\sigma^2}{2}(1-s)^2 + r(s)(1-s)^2}{(1-f(s))(1-s)} \\ &= \frac{1-s}{1-f(s)} \left\{ \frac{\sigma^2}{2} + r(s) \right\} \end{aligned}$$

sumando y restando  $\sigma^2/2$  tenemos

$$\begin{aligned} &= \frac{\sigma^2}{2} - \frac{\sigma^2}{2} + \frac{1-s}{1-f(s)} \left\{ \frac{\sigma^2}{2} + r(s) \right\} \\ &= \frac{\sigma^2}{2} + \frac{\sigma^2}{2} \left[ \frac{1-s}{1-f(s)} - 1 \right] + r'(s) \\ &= \frac{\sigma^2}{2} + \frac{\sigma^2}{2} \underbrace{\left[ \frac{f(s)-s}{1-f(s)} \right]}_{r''(s)} + r'(s) \\ &= \frac{\sigma^2}{2} + \rho(s) \end{aligned}$$

esta última igualdad del hecho de que  $\lim_{s \rightarrow 1} r''(s) = 0$  como veremos a continuación:

$$\begin{aligned} \lim_{s \rightarrow 1} \frac{f(s)-s}{1-f(s)} &\stackrel{\text{L'HÔPITAL}}{=} \lim_{s \rightarrow 1} \frac{f'(s)-1}{-f'(s)} \\ &= \frac{f'(1)-1}{-f'(1)} \equiv 0 \text{ ya que } f'(1) = m = 1. \end{aligned}$$

y definimos  $\rho(s) = r''(s) + r'(s)$  la cual se va a cero cuando  $s \uparrow 1$ .

Iterando esto, tenemos:

$$\begin{aligned} & \frac{1}{n} \left\{ \frac{1}{1-f_n(s)} - \frac{1}{1-s} \right\} = \\ & = \frac{1}{n} \left\{ \frac{1}{1-f(f_{n-1}(s))} - \frac{1}{1-f_{n-1}(s)} \right\} + \left\{ \frac{1}{1-f_{n-1}(s)} - \frac{1}{1-s} \right\} \\ & = \dots = \frac{1}{n} \sum_{j=0}^{n-1} \left\{ \frac{1}{1-f \circ f_j(s)} - \frac{1}{1-f_j(s)} \right\} \\ & = \frac{n\sigma^2}{2} \frac{1}{n} + \frac{1}{n} \sum_{j=0}^{n-1} \rho \circ f_j(s), \end{aligned}$$

tomando

$$\lim_{n \rightarrow \infty} \frac{1}{n} \left\{ \frac{1}{1-f_n(s)} - \frac{1}{1-s} \right\} = \frac{\sigma^2}{2}$$

ya que  $f_n(0) \leq f_n(s) \leq 1$  y  $f_n(0) \uparrow 1$  la convergencia  $f_n \rightarrow 1$  es uniforme y como  $\rho$  tiende a cero, la prueba queda concluida.

□

**TEOREMA 2.5.2.1** Si  $m = 1$  y  $\sigma^2 < \infty$ , entonces

- (i)  $\lim_{n \rightarrow \infty} nP[Z_n > 0] = 2/\sigma^2$ ;
- (ii)  $\lim_{n \rightarrow \infty} E \left[ \frac{Z_n}{n} | Z_n > 0 \right] = \sigma^2/2$ ;
- (iii)  $\lim_{n \rightarrow \infty} P \left[ \frac{Z_n}{n} \leq u | Z_n > 0 \right] = 1 - e^{-\frac{2nu}{\sigma^2}}$ ,  $u \geq 0$

**Demostración:** ([39], pág. 25)

(i)

$$\begin{aligned} nP[Z_n > 0] &= n\{1 - f_n(0)\} \\ &= \left\{ \frac{1}{n} \left( \frac{1}{1-f_n(0)} - 1 \right) + \frac{1}{n} \right\}^{-1} \rightarrow \frac{2}{\sigma^2} \end{aligned}$$

(ii)

$$\begin{aligned} E\left[\frac{Z_n}{n} \mid Z_n > 0\right] &= \frac{E[Z_n]}{n} \{1 - f_n(0)\}^{-1} \\ &= \frac{1}{nP[Z_n > 0]} \rightarrow \frac{1}{n \frac{\sigma^2}{n^2}} = \frac{\sigma^2}{2} \end{aligned}$$

ya que

$$\begin{aligned} \frac{1}{n} &= E\left[\frac{Z_n}{n}\right] = E\left(\frac{Z_n}{n} \mid Z_n > 0\right) P[Z_n > 0] + 0 \cdot P[Z_n = 0] \\ &\Rightarrow E\left[\frac{Z_n}{n} \mid Z_n > 0\right] = \frac{1}{nP[Z_n > 0]} \end{aligned}$$

(iii) De la ecuación (2.5.1.8) y la relación entre la f.g.p y la f.g.m vista en la sección (2.2) tenemos

$$\begin{aligned} E[e^{-u \frac{Z_n}{n}} \mid Z_n > 0] &= 1 - \frac{1 - f_n(\exp(\frac{-u}{n}))}{1 - f_n(0)} \\ &= 1 - \frac{1}{n[1 - f_n(0)]} \left\{ \frac{1}{1 - f_n(\exp(\frac{-u}{n}))} - \frac{1}{1 - \exp(\frac{-u}{n})} \right\} \\ &\quad + \frac{1}{n(1 - \exp(\frac{-u}{n}))} \}^{-1} \rightarrow 1 - \left(\frac{\sigma^2}{2}\right) \left(\frac{\sigma^2}{2} + \frac{1}{u}\right)^{-1} = \frac{1}{(1 + u\sigma^2/2)} \end{aligned}$$

donde  $1/(1 + u\sigma^2/2)$  es la función generadora de momentos o la transformada de Laplace de una distribución exponencial con parámetro  $\lambda = 2/\sigma^2$ . Luego entonces el Teorema de Continuidad, dice que si una sucesión de funciones generadoras de momentos converge a una función generadora de momentos, entonces la sucesión de funciones de distribución, convergen a la función de distribución caracterizada por la función generadora límite. De aquí que (iii) se cumple. □

### 2.5.3 CASO SUPERCRÍTICO

Sea  $W_n = m^{-n} Z_n$ . Por el Teorema (2.5.1) sabemos que existe una v.a.  $W$  tal que

$$\lim_{n \rightarrow \infty} W_n = W \quad \text{c.s.}$$

pero si  $m \leq 1$  la extinción es segura, de aquí que no tiene sentido estudiar la v.a.  $W_n$  pues su límite es una v.a. degenerada. No así para el caso supercrítico, dado que la extinción no es segura; por lo cual el siguiente Teorema será enunciado:

**TEOREMA 2.5.3.1** Si  $m > 1$ ,  $\sigma^2 < \infty$  y  $Z_0 = 1$ , entonces:

- (i)  $\lim_{n \rightarrow \infty} E[(W_n - W)^2] = 0$
- (ii)  $E(W) = 1$  y  $Var(W) = \frac{\sigma^2}{m^2 - m}$
- (iii)  $P(W = 0) = q \equiv P(Z_n = 0)$  p.a.  $n$

**Demostración:** ([48], pág 9)

(i)

$$\begin{aligned} E(W_n^2) &= \frac{1}{m^{2n}} E(Z_n^2) \\ &= \frac{1}{m^{2n}} [Var(Z_n) + (E(Z_n))^2] \\ &= \frac{1}{m^{2n}} \left[ \frac{\sigma^2 m^n (m^n - 1)}{m^2 - m} + m^{2n} \right] \\ &= \left( \frac{\sigma^2 (1 - m^{-n})}{m^2 - m} + 1 \right) \end{aligned}$$

y

$$E[(W_{n+k} - W_n)^2] = E[W_{n+k}^2] + E[W_n^2] - 2E[W_{n+k}W_n]$$

pero

$$\begin{aligned} E[W_{n+k}W_n] &= E[E[W_{n+k}W_n|W_n]] \\ &= E[W_n E[W_{n+k}|W_n]] = E[W_n^2] \end{aligned}$$

entonces

$$\begin{aligned} E[(W_{n+k} - W_n)^2] &= E[W_{n+k}^2] - E[W_n^2] \\ &= \frac{\sigma^2 m^{-n}}{m^2 - m} (1 - m^{-k}) \quad n, k = 0, 1, 2, \dots \end{aligned}$$

se sigue que  $W_n$  converge en la media cuadrática a una v.a.  $W$ , ya que en la igualdad anterior probamos que  $\{W_n\}$  es una sucesión de Cauchy, en donde para un número  $k$

fijo y  $n$  muy grande dicha media cuadrática es tan pequeña como se quiera. Y por el teorema anterior se cumple también que

$$\lim_{n \rightarrow \infty} E[(W_n - W)^2] = 0$$

□

(ii) Sabemos que si converge en media cuadrática converge en la media de aquí que

$$\lim_{n \rightarrow \infty} E[W_n - W] = 0$$

y

$$E[W_n] = \frac{E[Z_n]}{m^n} = \frac{m^n}{m^n} = 1$$

por lo tanto  $E[W] = E[W_n] = 1$

$$E[W_n^2] = \frac{\sigma^2(1 - m^{-n})}{m^2 - m} + 1$$

y de aquí que

$$\sup_n E[W_n^2] = \lim_n E[W_n^2] = \left\{ \frac{\sigma^2}{m^2 - m} \right\} + 1 < \infty$$

y por teoría estandar de martingalas se sigue que

$$\text{Var}(W) = \frac{\sigma^2}{m^2 - m}$$

□

(iii) Si  $r = P(W = 0)$  entonces  $E(W) = 1$  implica  $r < 1$ , además, por análisis del primer paso, vemos que  $r$  debe satisfacer:

$$\begin{aligned} r &= \sum_k P(W = 0 | Z_1 = k) P(Z_1 = k) \\ &= \sum_k P(W_n \rightarrow 0 | Z_1 = k) P(Z_1 = k) \\ &= \sum_k p_k [P(W_n \rightarrow 0)]^k \\ &= \sum_k p_k [P(W = 0)]^k \\ &= f(r) \end{aligned}$$

por lo tanto  $r = f(r)$  y de aquí que  $r$  debe de ser igual a  $q$ .

## 2.6 LA PROGENIE TOTAL

Estimadores para la media de reproducción y la probabilidad de extinción, así como propiedades asintóticas de estos, que serán tratadas en el siguiente capítulo; están dadas en términos de sumas parciales del proceso  $\{Z_n\}$ , así como de la progenie total del proceso. De allí la importancia y estudio de dichas cantidades en esta sección.

La progenie total de un proceso de Galton-Watson se define como:

$$Y_\infty = \sum_{n=0}^{\infty} Z_n \quad (2.6.1)$$

lo cual define a  $Y_\infty$  como un valor entero o infinito de una variable aleatoria. Obviamente, por el Teorema (2.3.2)

$$P[Y_\infty < \infty] = q$$

y

$$Y_n = \sum_{k=0}^n Z_k \uparrow Y_\infty \text{ c.s.}$$

La función generadora de probabilidad de  $Y_n$  se puede obtener recursivamente como sigue:

$$\begin{aligned} h_{n+1}(s) &= E[s^{Y_{n+1}}] = E[E[s^{Z_0 + \dots + Z_{n+1}} | Z_1]] \\ &= sE[E[s^{Z_1 + \dots + Z_{n+1}} | Z_1]] = sE[\{h_n(s)\}^{Z_1}] \\ &= sf \circ h_n(s). \end{aligned}$$

ahora bien, sea

$$p'_k = P[Y_\infty = k] \quad k = 1, 2, \dots$$

se sigue que  $h_\infty(s) = E[s^{Y_\infty}]$ . Es claro que si  $m \leq 1$   $Z_n \rightarrow 0$  con probabilidad uno, lo cual implica que  $Y_\infty < \infty$  y por lo tanto  $h_\infty(s)$  es finita. Para el caso supercrítico ( $m > 1$ ), sabemos que  $Z_n \rightarrow 0$  con probabilidad  $q$  por lo cual  $Y_\infty$  podría no ser finita, pero  $h_\infty$  es siempre finita como se puede observar a continuación: Sea  $0 \leq s \leq 1$ , entonces

$$h_\infty(s) = E[s^{Y_\infty}] \leq \sum_{k=0}^{\infty} s^k p'_k \leq \sum_{k=0}^{\infty} p'_k < 1$$

de aquí que

$$h_{\infty}(s) = sf \circ h_{\infty}(s) \quad (2.6.2)$$

Ahora veremos que  $h_{\infty}(s)$  es solución única a esta ecuación. Si  $g$  es alguna función entre cero y uno que resuelve (2.6.2) y  $g(s) > h_{\infty}(s)$ , entonces por el T.V.M. tenemos que existe  $\xi \in (h_{\infty}(s), g(s))$  tal que:

$$\begin{aligned} |h_{\infty}(s) - g(s)| &= |sf \circ h_{\infty}(s) - sf \circ g(s)| \\ &= s |f(h_{\infty}(s)) - f(g(s))| = sf'(\xi) |h_{\infty}(s) - g(s)| \\ &\leq sm |h_{\infty}(s) - g(s)|, \end{aligned}$$

entonces

$$|h_{\infty}(s) - g(s)| \leq sm |h_{\infty}(s) - g(s)|,$$

y como  $ms < 1$  para  $0 \leq s < m^{-1}$  esto implica que  $|h_{\infty}(s) - g(s)| < |h_{\infty}(s) - g(s)|$ , lo cual es una contradicción. Por lo tanto,  $h_{\infty}(s) = g(s)$  para  $0 \leq s < m^{-1}$ ; así la ecuación (2.6.2) sólo puede tener una solución sobre el intervalo  $(0, 1)$ .

En el caso  $m \leq 1$ , sea el mapeo  $s \mapsto s/f(s)$ , obteniendo la derivada de la función, esto es

$$\frac{d[s/f(s)]}{ds} = \frac{f(s) - sf'(s)}{[f(s)]^2} \geq \frac{f(s) - s}{[f(s)]^2} > 0,$$

ya que  $m = f'(1) \leq 1$  y  $f$  es creciente. Lo cual no dice que el mapeo es estrictamente creciente. Además se observa que es la inversa de  $h_{\infty}(s)$  esto es:

$$I(h_{\infty}(s)) = \frac{h_{\infty}(s)}{f(h_{\infty}(s))} = s \frac{f(h_{\infty}(s))}{f(h_{\infty}(s))} = s.$$

En el caso supercrítico con  $q > 0$  consideremos un nuevo proceso con función generadora de reproducción  $g(s) = f(qs)/q$ , esto es:

$$\begin{aligned} g(s) &= \frac{f(qs)}{q} \\ &= \sum_{k=0}^{\infty} \frac{(qs)^k p_k}{q} \\ &= \sum_{k=0}^{\infty} (q^{k-1} p_k) s^k \end{aligned}$$

$$= \sum_{k=0}^{\infty} p_k^* s^k,$$

donde  $q^{k-1} p_k = p_k^*$  y

$$\begin{aligned} \sum_{k=0}^{\infty} q^{k-1} p_k &= \frac{1}{q} \sum_{k=0}^{\infty} q^k p_k \\ &= \frac{1}{q} f(q) = \frac{q}{q} = 1. \end{aligned}$$

Dado que  $g'(1) = f'(q) < 1$  este proceso es subcrítico. Denotemos la función generadora de la progeñie total por  $h$  entonces:

$$h(s) = s \frac{f(qh(s))}{q}$$

$qh$  debe de satisfacer la (2.6.2) y además  $h$  es  $q$  multiplicada por la inversa de  $qs/f(qs)$ .

$$\begin{aligned} E[Y_n] &= E\left[\sum_{k=0}^n Z_k\right] = \sum_{k=0}^n E[Z_k] \\ &= \sum_{k=0}^n m^k = \begin{cases} \frac{1-m^{n+1}}{1-m} & \text{si } m \neq 1 \\ n+1 & \text{si } m = 1 \end{cases} \end{aligned}$$

y también se verifica que

$$\text{Var}[Y_n] = \begin{cases} \frac{\sigma^2\{(1-m^{2n+1})/(1-m) - (2n+1)m^n\}/(1-m)}{6} & \text{si } m \neq 1 \\ \text{si } m = 1 \end{cases}$$

Para resultados significativos de esta sección usaremos una simple consecuencia del Teorema de Ballot ([39], pág. 40): Sea  $X_1, X_2, \dots$  v.a.i.i.d enteras no negativas y sea  $S_k$  su suma parcial

$$S_k = \sum_{j=1}^k X_j$$

entonces, si  $E[X_1] < 1$

$$P[S_k < k \forall k] = 1 - E[X_1] \quad (2.6.3)$$

**TEOREMA 2.6.1** Sea  $P_{jk} = P[Z_{n+1} = k | Z_n = j]$ . Entonces  $P\{Y_\infty = k\} = \frac{1}{k} P_{k,k-1}$   $k \in \mathbf{N}$ . Más generalmente, si  $Z_0 = j \in \mathbf{N}$ , entonces

$$P\{Y_\infty = k\} = \binom{j}{k} P_{k,k-j} \quad k \geq j.$$

**Demostración:** ([39], pág. 40)

Para  $0 \leq s \leq 1$  sea  $X_1, X_2, \dots$  v.a independientes con distribución  $\{p_k\}$  y sean  $X_1(s), X_2(s), \dots$  v.a independientes con

$$P\{X(s) = k\} = p_k s^k / f(s), \quad (2.6.4)$$

y

$$\sum_{k=0}^{\infty} P\{X(s) = k\} = 1.$$

Sea  $F(t)$  la f.g.p de  $X(s)$  entonces

$$\begin{aligned} F(t) &= E[t^{X(s)}] = \sum_{j=0}^{\infty} P\{X(s) = j\} t^j \\ &= \sum p_j s^j t^j / f(s) = f(st) / f(s) \end{aligned} \quad (2.6.5)$$

Definimos las sumas parciales  $S_n(s) = \sum_{j=1}^n X_j(s)$ ,  $S_n = \sum_{j=1}^n X_j$ , y  $S_0(s) = 0$ .

Suponga que comenzamos con  $m \leq 1$  entonces

$$\begin{aligned} E\{X(s)\} &= \sum_{k=0}^{\infty} k P\{X(s) = k\} = \sum_{k=0}^{\infty} k p_k s^k / f(s) \\ &= \frac{1}{f(s)} \sum_{k=0}^{\infty} k p_k s^k = \frac{s}{f(s)} \sum_{k=0}^{\infty} k p_k s^{k-1} \\ &= \frac{s f'(s)}{f(s)} < 1, \quad \text{para } s \in [0, 1), \end{aligned} \quad (2.6.6)$$

ya que  $s \leq f(s)$ , lo que implica que  $s/f(s) \leq 1$  y como  $f'(s) < f'(1) \leq 1$  para  $s \in [0, 1)$  tenemos que  $s f'(s)/f(s) < 1$ , y

$$E\{X(1)\} = \frac{f'(1)}{f(1)} = f'(1) = m,$$

además por (2.6.3)

$$P[S_n(s) < n, \text{ para toda } n \in \mathbf{N}] = 1 - sf'(s)/f(s),$$

La Ley de Grandes Números dice que:

$$\lim_{n \rightarrow \infty} \frac{S_n(s)}{n} = sf'(s)/f(s) < 1, \text{ para } 0 \leq s < 1.$$

Además, para cualquier  $k \in \mathbf{N}$ ,  $S_n(s)$  puede ser mayor o igual a  $n - k$  sólo para un número finito de índices  $n$ , ya que si fuera mayor o igual para una infinidad de índices  $n$ , tendríamos

$$\frac{S_n(s)}{n} \geq \frac{n - k}{n} \xrightarrow{n \rightarrow \infty} 1,$$

pero

$$\lim_{n \rightarrow \infty} \frac{S_n(s)}{n} < 1,$$

lo cual es una contradicción. De esto último, el hecho de que  $S_0(s) = 0$  y de la primera ecuación en el siguiente arreglo, obtenida por P. Jagers ([39], pág. 41) tenemos:

$$\begin{aligned} 1 &= \sum_{n=k}^{\infty} P[S_n(s) = n - k, S_{n+j}(s) < n - k + j, \forall j \in \mathbf{N}] \\ &= \sum_{n=k}^{\infty} P[S_{n+j}(s) < n - k + j \forall j \in \mathbf{N} | S_n(s) = n - k] P[S_n(s) = n - k] \\ &= \sum_{n=k}^{\infty} P \left[ S_n(s) + \sum_{i=n+1}^{n+j} X_i(s) < n - k + j \forall j \in \mathbf{N} | S_n(s) = n - k \right] P[S_n(s) = n - k] \\ &= \sum_{n=k}^{\infty} P \left[ n - k + \sum_{i=n+1}^{n+j} X_i(s) < n - k + j \forall j \in \mathbf{N} \right] P[S_n(s) = n - k] \\ &= \sum_{n=k}^{\infty} P \left[ \sum_{i=n+1}^{n+j} X_i(s) < j \forall j \in \mathbf{N} \right] P[S_n(s) = n - k] \\ &= \sum_{n=k}^{\infty} P \left[ \sum_{i=1}^j X_i(s) < j \forall j \in \mathbf{N} \right] P[S_n(s) = n - k] \\ &= P[S_j(s) < j, \forall j \in \mathbf{N}] \sum_{n=k}^{\infty} P[S_n(s) = n - k] \end{aligned}$$

$$= \{1 - sf'(s)/f(s)\} \sum_{n=k}^{\infty} P[S_n = n - k] s^{n-k} / \{f(s)\}^n \quad (2.6.7)$$

la última ecuación se obtiene de (2.6.5) y de observar que la f.g.p de la v.a.  $S_n(s)$  está dada por

$$\begin{aligned} F_{S_n(s)}(t) &= E[t^{S_n(s)}] = E[t^{X_1(s)+\dots+X_n(s)}] \\ &= (E[t^{X(s)}])^n = \left(\frac{f(st)}{f(s)}\right)^n \\ &= E[st^{S_n}]/(f(s))^n = \sum_{k=0}^{\infty} P[S_n = k] s^k t^k / \{f(s)\}^n \\ &= \sum_{k=0}^{\infty} \underbrace{\frac{P[S_n = k] s^k}{[f(s)]^n}}_A t^k, \end{aligned}$$

donde  $A = P[S_n(s) = k] = P[S_n = k] s^k / [f(s)]^n$  para  $k = 0, 1, \dots$ .

Ahora sea  $u = h_{\infty}^{-1}(s) = s/f(s)$  entonces

$$\frac{sf'(s)}{f(s)} = \frac{h_{\infty}(u)f'(h_{\infty}(u))}{f(h_{\infty}(u))} \stackrel{\text{Ec. (2.6.2)}}{=} f'(h_{\infty}(u))u \quad (2.6.8)$$

y diferenciando (2.6.2) en  $u$  tenemos:

$$\begin{aligned} h'_{\infty}(u) &= f \circ h_{\infty}(u) + uf'(h_{\infty}(u))h'_{\infty}(u) \\ &= f(s) + \left\{ \frac{sf'(s)}{f(s)} \right\} h'_{\infty}(u), \end{aligned}$$

o

$$1 - \frac{sf'(s)}{f(s)} = \frac{f(s)}{h'_{\infty}(u)} = \frac{f(h_{\infty}(u))}{h'_{\infty}(u)} \stackrel{\text{Ec. (2.6.2)}}{=} \frac{h_{\infty}(u)}{uh'_{\infty}(u)}.$$

Sustituyendo  $u = s/f(s)$  y  $s = h_{\infty}(u)$  dentro de (2.6.6) tenemos:

$$1 = \frac{h_{\infty}(u)}{uh'_{\infty}(u)} \sum_{n=k}^{\infty} P[S_n = n - k] u^n [h_{\infty}(u)]^{-k},$$

y esto implica

$$\frac{uh'_{\infty}(u)[h_{\infty}(u)]^k}{h_{\infty}(u)} = \sum_{n=k}^{\infty} P[S_n = n - k] u^n$$

$$[h_\infty(u)]^{k-1} h'_\infty(u) = \sum_{n=k}^{\infty} P\{S_n = n - k\} u^{n-1}.$$

Integrando de cero a  $s$  obtenemos para  $k \geq 1$

$$[h_\infty(s)]^k = \sum_{n=k}^{\infty} P\{S_n = n - k\} \frac{s^n(k)}{n}.$$

Para  $k = 1$  tenemos la f.g.p. de la variable aleatoria  $Y_\infty$ , esto es:

$$\begin{aligned} h_\infty(s) &= \sum_{n=1}^{\infty} \frac{P\{S_n = n - 1\} s^n}{n} \\ &= \sum_{n=1}^{\infty} \frac{P\left\{\sum_{j=1}^n X_j = n - 1\right\} s^n}{n} \\ &= \sum_{n=1}^{\infty} \underbrace{\frac{P_{n,n-1}}{n}}_B s^n, \end{aligned}$$

donde  $B = P\{Y_\infty = k\}$ . Y dado que  $h_\infty(0) = 0$ , esto concluye la prueba para  $m \leq 1$ .

Para el proceso supercrítico con  $q > 0$ , la consecuencia del Teorema de Ballot se aplica dado que:

$$\left\{ \frac{f(qs)}{q} \right\}^j = \sum_{k=0}^{\infty} P_{jk} q^{j-k} s^k;$$

$$\begin{aligned} \{h_\infty(s)\}^{k-1} &= q^{k-1} \{h(s)\}^{k-1} \\ &= q^{k-1} \sum_{n=k}^{\infty} P_{n-1, n-k} q^{1-k} s^{n-1} (k-1)/(n-1) \\ &= \sum_{n=k}^{\infty} P_{n-1, n-k} s^{n-1} (k-1)/(n-1). \end{aligned}$$

Tomando  $k = 2$  tenemos:

$$\begin{aligned} h_\infty(s) &= \sum_{n=2}^{\infty} \frac{P_{n-1, n-2} s^{n-1}}{n-1} \\ &= \sum_{k=1}^{\infty} \frac{P_{k, k-1} s^k}{k} \end{aligned}$$

□

La función generadora de probabilidad es estrictamente creciente y tiene una inversa  $g$ , cuya  $n$ -ésima iterada la denotamos por  $g_n$ ,  $n \in \mathbb{N}$ . Definimos  $g_0(s) = s$ . La función  $g$  es creciente, cóncava, diferenciable, y mapea el intervalo  $[q, 1]$  sobre sí mismo. Definimos  $C_n(s) = -\log g_n(s)$ , lo que nos permite enunciar el siguiente

**TEOREMA 2.6.2** *En un proceso supercrítico, sean  $C_n(s)$  las constantes definidas arriba entonces, para  $q < s < 1$*

$$C_n(s)Y_n \rightarrow \frac{Y(s)m}{m-1} \text{ c.s.}$$

donde

$$Y(s) = \lim_{n \rightarrow \infty} C_n(s)Z_n$$

Similarmente, con  $W = \lim_{n \rightarrow \infty} Z_n/m^n$

$$\frac{Y_n}{m^n} \rightarrow \frac{Wm}{m-1}$$

**Demostración:** ([74, 73])

Primero observemos que dado que  $f(s) \leq s$  para  $q \leq s \leq 1$ ,  $g(s)$  debe ser mayor o igual a  $s$  y por lo tanto  $g_n \uparrow$  con  $g_\infty$  una función límite. Dado que

$$s = f_n \circ g_n(s) \leq f_n \circ g_\infty(s) \rightarrow q$$

si  $g_\infty(s) < 1$ ,  $g_\infty(s) = 1$  para  $s > q$ . Y ahora tomando la expansión en serie de Taylor de (2.5.1.1),  $1 - f(s) = \{m - r(s)\}(1 - s)$  y sustituyendo  $s$  por  $g(s)$  para  $q < s < 1$  obtenemos

$$\frac{1 - g(s)}{(1 - s)} = \frac{m^{-1}}{\left\{1 - \frac{r \circ g(s)}{m}\right\}}.$$

como hicimos antes, tomamos el producto y tenemos

$$m^n \{1 - g_n(s)\} = \frac{(1 - s)}{\prod_{k=1}^n \{1 - r \circ g(s)/m\}}. \quad (2.6.9)$$

Esta representación dice algo acerca de  $C_n(s)$  dado el  $-\log x \sim 1 - x$  cuando  $x \downarrow 1$ .

En particular

$$\lim_{n \rightarrow \infty} \frac{C_n}{C_{n-1}} = \lim_{n \rightarrow \infty} \frac{(1 - g_n)}{(1 - g_{n-1})} = \frac{1}{m} \quad (2.6.10)$$

Ignorando la dependencia sobre  $s$  vemos de (2.6.10) que  $C_{n-j}/C_n \rightarrow m^j$ , cuando  $n \rightarrow \infty$  y por lo tanto

$$C_n Y_n \geq C_n \sum_{j=0}^k Z_{n-j} = \sum_{j=0}^k \left( \frac{C_n}{C_{n-j}} \right) C_{n-j} Z_{n-j} \rightarrow Y \sum_{j=0}^k m^{-j} \text{ c.s.}$$

de aquí que

$$\lim(\inf C_n Y_n) \geq \frac{Ym}{(m-1)}$$

el otro lado es directamente obtenido de la ecuación (2.6.9) de tal forma que:

$$\frac{1-g_n}{1-g_k} = m^{k-n} \prod_{j=k+1}^n (1-r \circ g_j/m) \leq m^{k-n}$$

sea  $\epsilon > 0$ , y sea una v.a  $\nu$  tal que  $(1-g_n)Z_n < y + \epsilon$  para  $n \geq \nu$  dado que  $(C_n/(1-g_n) \rightarrow 1)$ . De aquí que

$$\begin{aligned} C_n Y_n &\leq C_n Y_\nu + \{C_n/(1-g_n)\}(Y + \epsilon) \sum_{k=\nu}^n (1-g_n)/(1-g_k) \\ &\leq C_n Y_\nu + \{C_n/(1-g_n)\}(Y + \epsilon) \sum_{k=0}^n m^{-k} \rightarrow \frac{(Y + \epsilon)m}{m-1} \end{aligned}$$

el caso con normalización  $m^{-n}$  se sigue simplemente. □

**TEOREMA 2.6.3** *Considere un proceso de ramificación supercrítico con varianza finita de reproducción  $\sigma^2$ . Cuando  $n \rightarrow \infty$ , la distribución de  $\{(Y_{n+1} - Z_0)/Y_n - m\} \sqrt{Y_n}$  dada  $Z_n > 0$ , es asintóticamente normal con media cero y varianza  $\sigma^2$ .*

La demostración de este teorema no se incluirá en este trabajo, ([39], págs. 48-49).

En este momento es importante enunciar dos nuevos y funcionales teoremas límites, que nos ayudarán a demostrar algunas propiedades asintóticas, las cuales enunciaremos más adelante. La idea de su prueba es dada en [41].

Sea  $\{\xi_n\}$  una sucesión de v.a.i.i.d sobre  $(\Omega, \mathcal{B}, P)$ , con

$$E(\xi_1) = 0 \text{ y } 0 < E(\xi_1^2) = \sigma^2 < +\infty.$$

Considere las sumas parciales  $\{S_n\}_{n \geq 1}$  y sea  $0 \leq t \leq 1$ ,

$$X_n(t, \omega) = \sigma^{-1} n^{-\frac{1}{2}} S_{[nt]}(\omega), \quad \omega \in \Omega.$$

Para cada  $n \geq 1$ , sea  $\nu_n$  una v.a. no-negativa sobre el mismo espacio y definimos

$$Y_n(t, \omega) = \begin{cases} X_{\nu_n(\omega)}(t, \omega) & \text{si } \nu_n(\omega) > 0, \\ 1 & \text{en otro caso.} \end{cases}$$

**TEOREMA 1** Suponga que existe una sucesión  $\{a_n\}$  de constantes tales que

(i)  $a_n \rightarrow +\infty$ , si  $n \rightarrow +\infty$ .

(ii)  $(\nu_n/a_n) \xrightarrow{P} \theta$ ,  $\theta$  una v.a. no-negativa

Suponga también que  $D = \{\omega : \theta(\omega) > 0\}$  es tal que  $P(D) > 0$ . Y tomemos  $P_D(\cdot) = P(\cdot|D)$ . Entonces, para cualquier medida de probabilidad  $Q$  absolutamente continua con respecto a  $P_D(Q \ll P_D)$ , y  $0 \leq t \leq 1$ ,

$$Q\{\omega : Y_n(t, \omega) \leq x\} \rightarrow Q\{\omega : W^*(t, \omega) \leq x\}$$

donde  $W^*$  es el proceso de Wiener.

**TEOREMA 2** Bajo las hipótesis del Teorema 1, definimos, para  $\omega \in D$ ,  $Y'_n$  por

$$Y'_n(t, \omega) = (a_n/\nu_n(\omega))^{\frac{1}{2}} \cdot Y_n(t, \omega), \quad 0 \leq t \leq 1.$$

Entonces, para cualquier  $Q \ll P_D$ ,

$$Q\{\omega : Y'_n(t, \omega) \leq x\} \xrightarrow{n \rightarrow \infty} Q\{\omega : W^*(t, \omega) \cdot \theta_0^{\frac{1}{2}} \leq x\}$$

donde  $W^*$  es el proceso de Wiener,  $\theta_0$  es  $Q$ -independiente de  $W^*$  y

$$Q\{\omega : \theta_0(\omega) \leq x\} = Q\{\omega : \theta(\omega) \leq x\}.$$

## Capítulo 3

# ESTIMACIÓN DE LA MEDIA DE REPRODUCCIÓN Y LA PROBABILIDAD DE EXTINCIÓN

Como un problema de estimación paramétrico para un proceso estocástico, generalmente se entiende lo siguiente. Sea  $\{X_k; k \geq 1\}$  el proceso estocástico definido sobre un espacio de probabilidad  $(\Omega, \mathcal{F}, P_\theta)$ . Suponga que el vector  $X_{(n)} = (X_1, X_2, \dots, X_n)$  de v.a. posiblemente dependientes, tienen una densidad de probabilidad conjunta  $P(X_{(n)}; \Theta)$ , que depende de parámetros desconocidos  $\Theta = (\theta_1, \dots, \theta_r)$ , donde la forma funcional de esta densidad es conocida mientras el vector  $\Theta$  es desconocido, y  $x_{(n)} = (x_1, \dots, x_n)$  denota una muestra o realización de  $n$  observaciones del proceso estocástico. Dado que no conocemos el vector de parámetros  $\Theta$ , se desea obtener una buena estimación de este, a partir de la muestra de observaciones  $(x_1, \dots, x_n)$ . Todo esto nos lleva a formular la siguiente definición:

**Un Estimador:** Es una estadística que toma valores en el espacio parametral y cuyo valor se aproxima al verdadero valor de  $\Theta$ .

Lo que se quiere es un **Estimador** para el vector de parámetros que involucra nuestro proceso. Obviamente, hablar de un estimador involucra también, el hecho de poseer criterios matemáticos de optimalidad que nos permitan ver que tan buenas o que tan malas han resultado nuestras aproximaciones. Muchos criterios de optimalidad tales como: suficiencia, eficiencia, insesgamiento, han sido desarrollados y están disponibles para seleccionar un buen estimador entre la clase de posibles estimadores y muchos

criterios asintóticos de optimalidad tales como, consistencia, eficiencia asintótica, suficiencia asintótica, insesgamiento asintótico están disponibles para seleccionar buenos estimadores de entre la clase de posibles estimadores cuando estudiamos propiedades muestrales a la larga de los estimadores.

Este capítulo no tiene la finalidad de abordar un estudio profundo acerca de las propiedades que tendrán los estimadores para la media y la probabilidad de extinción del proceso de Galton-Watson, pero sí al menos dar una idea de la calidad de las estimaciones.

## INFERENCIA ESTADÍSTICA PARA LOS PROCESOS ESTOCÁSTICOS

Hoy día existen dos enfoques en lo que a *Inferencia Estadística* se refiere, el enfoque *Clásico* y el *Bayesiano*. A continuación se presentan brevemente, métodos de inferencia estadística para el apartado de *Estimación Paramétrica*, utilizando estos dos enfoques.

**a- Estimación Clásica (Método de Máxima Verosimilitud):** Cabe señalar que existen varios métodos de estimación clásica ([55], págs 273-290), pero en este trabajo sólo hablaremos del método de Máxima Verosimilitud (MV). Propiedades de éstos estimadores como: normalidad asintótica (bajo ciertas condiciones de regularidad), completos, consistencia, el que estén basados en estadísticas suficientes, entre otras [56]; hacen que estos estimadores sean generalmente buenos, y por lo tanto objeto de nuestro estudio.

Sea  $X = (X_1, \dots, X_n)$  un vector aleatorio de observaciones, cuya distribución conjunta está descrita por una densidad  $f_n(x|\Theta)$  sobre el espacio Euclideo  $n$ -dimensional  $\mathbb{R}^n$ . El vector de parámetros desconocido  $\Theta$  está contenido en el espacio paramétrico  $\Gamma \subset \mathbb{R}^s$ . Para un  $x$  fijo definimos la *Función de Verosimilitud* de  $x$  como  $L(\Theta) = L_x(\Theta) = f_n(x|\Theta)$  considerada como una función de  $\Theta \in \Gamma$ .

**Definición.** *Cualquier  $\hat{\Theta} = \hat{\Theta}(x) \in \Gamma$  que maximice  $L(\Theta)$  sobre  $\Gamma$  es llamado un estimador Máximo Verosímil (EMV) del parámetro desconocido  $\Theta$ .*

Frecuentemente para obtener un EMV, es más fácil, para efectos de cálculo, maximizar el logaritmo de la verosimilitud  $\log L(\Theta)$  en lugar de  $L(\Theta)$ .

Como ejemplo de funcionamiento del método considere, un proceso estocástico Markoviano discreto  $\{X_n; n \geq 1\}$  definido sobre un espacio de probabilidad

$(\Omega, \mathcal{F}, P_\Theta)$  con un vector de variables aleatorias posiblemente dependientes  $X_{(n)} = (X_1, \dots, X_n)$ , teniendo una densidad conjunta  $P(X_{(n)}; \Theta)$ , donde el parámetro  $\Theta = (\theta_1, \dots, \theta_r) \in \Gamma$  donde  $\Gamma$  es un subconjunto abierto de el espacio Euclideo  $r$ -dimensional y  $x_{(n)} = (x_1, \dots, x_n)$  denota una muestra de  $n$  observaciones del proceso estocástico en cuestión. Escribimos  $P_k(\theta) = P(X_k | X_{(k-1)}; \theta)$  para la densidad condicional de  $X_k$  dado  $X_1, X_2, \dots, X_{k-1}$ . Sea  $L_n(\Theta)$  la función de verosimilitud asociada con  $(x_1, \dots, x_n)$  entonces la función de verosimilitud  $L_n(\Theta)$  está dada por

$$\begin{aligned} L_n(\Theta) &= P(X_{(n)}; \Theta) = \\ &= P(X_1; \Theta)P(X_2|X_{(1)}; \Theta) \cdots P(X_n|X_{(n-1)}; \Theta) \\ &= \prod_{k=1}^n P_k(\Theta), \end{aligned}$$

donde  $P_1(\Theta) = P(X_1|X_0; \Theta) = P(X_1; \Theta)$ .

Como lo que se quiere es maximizar la verosimilitud o la log-verosimilitud como función de  $\Theta$ , un método muy usual es derivar esta función con respecto a cada entrada del vector  $\Theta$ . de aquí que bajo ciertas condiciones el estimador máximo verosímil de  $\Theta$  puede ser obtenido de la solución del sistema

$$\frac{\partial \log L_n(\Theta)}{\partial \theta_i} = 0, \quad i = 1, 2, \dots, r.$$

**b-Estimación Bayesiana (Pérdida Esperada Mínima):** La metodología estadística clásica aparece ante el investigador como un conjunto de técnicas aisladas cada una de las cuales resulta apropiada en determinados casos y bajo ciertas condiciones. Los métodos clásicos por otra parte, limitan el concepto de probabilidad a aquéllos sucesos en los que pueden definirse frecuencias relativas, lo que reduce seriamente su utilidad práctica. Además, no dispone de una base axiomática en la que pueda fundamentarse, por lo que ocasionalmente da lugar a resultados contradictorios. Sin embargo, hoy resulta posible ofrecer una metodología estadística unificada, totalmente general, la cual recibe el nombre de Estadística Bayesiana; que se deriva de analizar cuidadosamente el proceso lógico que debe seguirse para tomar una decisión. La diferencia fundamental entre los métodos clásicos y los bayesianos, radica en que el segundo utiliza un enfoque de probabilidad diferente el *Subjetivo* ver ([54], Cap. 6), que permite ampliar el concepto clásico de probabilidad y se fundamenta en un cuerpo axiomático que elimina la posibilidad de obtener resultados contradictorios.

La teoría de inferencia estadística es bajo el marco bayesiano, un problema de Teoría de Decisión. Esto es, todo problema de inferencia estadística tendrá la estructura de un problema de decisión en presencia de incertidumbre. Un problema de decisión en ambiente de incertidumbre es una estructura de la forma  $(\mathcal{D}, \mathcal{A}, U)$ , en donde  $\mathcal{D}$  es el conjunto de opciones, el cual debe ser exhaustivo, es decir que agote todas las posibilidades que puedan, en principio, parecer razonables.  $\mathcal{A}$  es una colección de sucesos inciertos relevantes que modifican o condicionan las posibles consecuencias de las opciones en  $\mathcal{D}$ .  $U$  es una función de utilidad para el decisor particular. Si  $U$  es la utilidad,  $L = -U$  es una posible función de pérdida asociada al problema de decisión. También se incluye una medida de probabilidad  $P$  asociada a los elementos de  $\mathcal{A}$  ([71], págs. 10-15).

Antes de formular el problema de inferencia que compete a este trabajo (la estimación), como un problema de decisión, tenemos que establecer conceptos básicos que serán fundamentales no sólo en la formulación del problema, sino también en su solución.

Supongamos sin pérdida de generalidad que estamos interesados en el valor de<sup>1</sup>

---

<sup>1</sup>Si los sucesos inciertos en que estamos interesados son de la forma  $\{A_i, i = 1, 2, \dots\}$  podemos considerar en su lugar la cantidad  $\theta = 1, 2, \dots$ , de forma que  $P(A_i) = P(\theta)$

$\theta$  que denominaremos parámetro de interés. El conocimiento disponible sobre el valor de  $\theta$  deberá ser expresado mediante una medida de probabilidad que describa el grado de creencia del investigador en la ocurrencia de los distintos posibles valores de  $\theta$ . Se trata pues de una cantidad aleatoria cuya distribución de probabilidad describe la información sobre el valor de  $\theta$  que inicialmente se posee; esta distribución recibe el nombre de *Distribución Inicial de  $\theta$*  ([71], pág. 128).

Frecuentemente con objeto de mejorar la información que inicialmente se tiene sobre el valor del parámetro de interés  $\theta$ , es posible realizar un experimento cuyo resultado  $x$  es una cantidad aleatoria con una distribución  $P(x|\theta)$  que depende de  $\theta$ ; en tal caso, la observación de  $x$  proporcionará información sobre el valor de  $\theta$ . Como función de  $x$  para un valor  $\theta = \theta_0$  fijo,  $P(x|\theta_0)$  es una densidad de probabilidad que describe la probabilidad de obtener los distintos posibles valores de  $x$ , si el parámetro de interés tuviese el valor  $\theta = \theta_0$ . Como función de  $\theta$ , para un  $x = x_0$  fijo,  $P(x_0|\theta)$  es una función que describe, la verosimilitud de los distintos valores de  $\theta$  a la luz del resultado experimental observado  $x_0$ . Los valores de  $\theta$  que hacen grande  $P(x_0|\theta)$  son aquéllos que hacen más plausible la observación del resultado  $x_0$  que ha sido observado; de manera que, después de observar  $x_0$ , estos valores de  $\theta$  resultan más verosímiles que los demás. Como se mencionó en el apartado clásico, la función de  $\theta$ ,  $P(x_0|\theta)$  es la Función de Verosimilitud (de  $\theta$  dado  $x_0$ ) y se denota  $L_{x_0}(\theta)$ . Cabe señalar que la función de verosimilitud es una función positiva de  $\theta$ , pero no una función de probabilidad ([71], pág. 134).

Supongamos que, con objeto de mejorar la información de que se dispone inicialmente sobre el valor de  $\theta$ , se realiza el experimento descrito en el párrafo anterior, esto es, se lleva a cabo un experimento que da lugar a resultados  $x$  con probabilidad  $P(x|\theta)$ . Como función de  $x$  y  $\theta$ ,  $P(x|\theta)$  es el modelo probabilístico que define la relación entre los resultados experimentales  $x$  y el parámetro de interés  $\theta$ . Sea  $P(\theta)$  la distribución inicial asignada a  $\theta$ . Después de observar el resultado  $x$  del experimento, la información de que disponemos sobre el valor de  $\theta$  estará descrita por su *Distribución Final*  $P(\theta|x)$ . El *Teorema de Bayes* ([71], págs. 145-146) permite obtener la distribución la distribución final  $P(\theta|x)$  a partir de la distribución inicial  $P(\theta)$  y la función de verosimilitud del resultado obtenido

$P(x|\theta)$ , así:  $P(\theta|x) = P(x|\theta)P(\theta)/P(x)$  donde  $P(x)$  es la *Distribución Predictiva o Marginal* de  $x$  ([71], pág. 139).

Ahora bien, supongamos que el decisor está interesado en una *Estimación* del valor del parámetro de interés  $\theta \in \Gamma$ , luego entonces, el decisor está frente a un problema de *Estimación Puntual*. En consecuencia, el problema de inferencia sobre  $\theta$  puede ser descrito como un problema de decisión, cuyo espacio de opciones  $\mathcal{D}$  es el espacio paramétrico de donde se debe seleccionar el valor estimado  $\hat{\theta}$  para  $\theta$ ; cuyo espacio de sucesos inciertos  $\mathcal{A}$  es igual al conjunto  $\{A_\theta = \{\theta\}\}$  para  $\theta \in \Gamma$ . Y cuya función de utilidad o pérdida es una medida de la discrepancia entre el verdadero valor del parámetro de interés y su estimación. Frecuentemente sin embargo, especialmente en problemas de investigación, el decisor estará interesado en obtener la mayor cantidad de información posible sobre el parámetro de interés, de forma que no resulta razonable que se limite a una simple estimación puntual de su valor. Cuando se culmina una etapa de investigación se deben comunicar las conclusiones finales sobre la magnitud de interés, junto con la argumentación en que tales conclusiones descansan. En términos más técnicos, los investigadores deben comunicar su conocimiento acerca del parámetro a través de una distribución final, junto con el modelo, los datos y la distribución inicial utilizadas para obtenerla. Por parte de la comunidad científica, la utilidad de sus conclusiones puede ser medida por la cantidad de nueva información proporcionada. Dado lo anterior, el problema de inferencia sobre  $\theta$  puede ser descrito de una manera más general, nuevamente como un problema de decisión, cuyo espacio paramétrico es el conjunto  $\Gamma$  de valores posibles de  $\theta$ , cuyo espacio de decisiones es la familia de las posibles distribuciones finales de  $\theta$ , y cuya función de utilidad o pérdida es una cierta medida de la diferencia o divergencia que existe entre la verdadera distribución final para el parámetro de interés y la distribución final estimada. ([71], pág 212).

La solución de cualquier problema de inferencia estadística bajo el enfoque bayesiano, en particular la estimación paramétrica, está basada en el uso de un conjunto de axiomas llamados en la literatura *Principios de Coherencia* ([71], págs. 22-28), los cuales, si son aceptados por el decisor, permiten asignar formalmente una medida cuantitativa de sus preferencias entre las consecuencias, tal medida esta dada por la función de utilidad  $U$  o pérdida  $L$ . En consecuencia se prueba

mediante el uso de dichos axiomas que el método para resolver cualquier problema de inferencia estadística es *minimizar la Pérdida Esperada* ([71], págs. 32-36). Este valor esperado es con respecto a la distribución final del parámetro de interés  $\theta$  ([53], cap. 7).

Evidentemente tenemos dos problemas, el de obtener una distribución inicial para el parámetro en cuestión y poder dar la función de pérdida asociada al problema de decisión.

La elección de una distribución inicial y una función de utilidad para el parámetro varía de decisor a decisor, es decir del conocimiento o la información que cada uno posea a-priori sobre el comportamiento del parámetro. En particular, el decisor puede hacer uso de *Familias Conjugadas* ([71], pág. 174) para representar la información inicial que tiene sobre el parámetro de interés; tal elección conduce a unos resultados matemáticos especialmente sencillos, ya que éstas tienen la propiedad que para cualquier distribución inicial perteneciente a una familia conjugada, se obtiene una distribución final que también pertenece a ella, esto es, son familias cerradas bajo el muestreo. Por otro lado, la inferencia estadística Bayesiana ha sido repetidamente descrita como la metodología que permite combinar de forma consistente la información inicial con la información experimental. Tal descripción sugiere inmediatamente un problema importante que todavía carece de una respuesta incontrovertida: se trata de determinar la forma en que deben realizarse inferencias cuando no se dispone de información inicial o cuando tal información *no se quiere o no se puede* utilizar.

Desde un punto de vista Bayesiano, el problema quedaría solucionado si se pudiese determinar una *distribución* inicial de *referencia*, que describa la situación en que los datos experimentales contienen *toda* la información relevante, en lugar de proporcionar tan sólo una parte de ella como sucede cuando se dispone de información inicial.

Se ha trabajado mucho en la obtención de distribuciones iniciales que añadan poca información a la información experimental, empezando por los trabajos de Bayes[59] y Laplace[68] basados en consideraciones de simetría. Intentos más modernos se basan en exigencias de invarianza (Jeffreys[67], Barnard[58], Hartigan[63], Box & Tiao[62] y Jaynes[66]), en límites de funciones de distribución conjugadas (Novick[69], DeGroot[54] cap. 10) o en argumentos basados en la

teoría de la información (Jaynes[64], Good[65], Zeller[70] y Bernardo[60]). Estos trabajos han demostrado que la mayor parte de los resultados numéricos clásicos encontrados en la práctica son casos particulares de la metodología Bayesiana.

En cuanto a lo que Función de Pérdida se refiere hay toda una variedad de éstas en la literatura ([71], pág. 214). La adopción de una de estas funciones, o de cualquier otra función de pérdida propia, dependerá de las razones que motiven el problema de inferencia.

### 3.1 ESTIMACIÓN DE LA MEDIA DE REPRODUCCIÓN

Estimación Clásica (Método de Máxima Verosimilitud):

Como primer ejemplo considere que la distribución de descendientes está parametrizada por alguna familia de distribuciones. El estadístico podría elegir como caso particular, trabajar con una distribución Poisson que tiene parámetro  $\lambda$ , esto es

$$P(\# \text{ de descendientes} = k) = \frac{e^{-\lambda} \lambda^k}{k!}; \quad k = 0, 1, 2, \dots$$

Supongamos que el proceso comienza con  $Z_0 = 1$  ancestros y que  $Z_1, Z_2, \dots, Z_n$  individuos fueron observados en las  $n$  primeras generaciones. Dado que el parámetro  $\lambda$  por definición mide para esta distribución, el número promedio de descendientes, éste coincide con la media ( $m$ ), que es el número promedio de descendientes por individuo en un proceso de Galton-Watson. El problema es entonces estimar  $\lambda$ . La función de verosimilitud asociada al problema está dada por

$$\begin{aligned} L_n(\lambda; Z_0, \dots, Z_n) &= P(Z_0 = 1, Z_1 = i_1, \dots, Z_n = i_n) \\ &= P(Z_0 = 1)P(1, i_1)P(i_1, i_2) \cdots P(i_{n-1}, i_n) \end{aligned}$$

donde

$$P(j, k) = P\{Z_{n+1} = k | Z_n = j\} = P\left[\sum_{i=1}^j X_{n,i} = k\right] = \frac{e^{-j\lambda} (j\lambda)^k}{k!},$$

esto último ya que las variables  $\{X_{i,j}\}$  son independientes y con la misma distribución (Poisson $[\lambda]$ ) y que esta distribución es cerrada bajo el operador suma. De aquí, que la verosimilitud queda expresada como

$$\begin{aligned} L_n(\lambda; Z_0, \dots, Z_n) &= 1 \cdot \frac{e^{-\lambda} \lambda^{i_1}}{i_1!} \cdot \frac{e^{-i_1 \lambda} (i_1 \lambda)^{i_2}}{i_2!} \cdots \frac{e^{-i_{n-1} \lambda} (i_{n-1} \lambda)^{i_n}}{i_n!} \\ &= K \cdot EXP\{-\lambda(1 + i_1 + i_2 + \cdots + i_{n-1})\} \cdot \lambda^{(i_1 + i_2 + \cdots + i_n)} \end{aligned} \quad (3.1. 1)$$

con

$$K = i_1^{i_2} \cdots i_{n-1}^{i_n} / \prod_{k=1}^n i_k!$$

Como se mencionó al principio del capítulo, un método clásico que suele dar buenas estimaciones, es el de Máxima Verosimilitud. Si se utiliza este método para estimar el parámetro  $\lambda$ , debemos maximizar la verosimilitud como función de  $\lambda$ . Una forma muy común de hacer esto, es derivar esta función con respecto al parámetro e igualarla a cero, para encontrar el punto crítico  $\hat{\lambda}$  que maximiza y tomarlo como el estimador. Por simplicidad técnica aplicamos la función logaritmo a la verosimilitud y derivemos éste igualando a cero para encontrar el máximo.

$$\log L_n(\lambda; Z(n)) = \log K - \lambda(1 + i_1 + \dots + i_{n-1}) + (i_1 + i_2 + \dots + i_n) \log \lambda$$

donde  $Z(n)$  es el vector  $(Z_0, Z_1, \dots, Z_n)$ . Derivando la función anterior con respecto a  $\lambda$  tenemos

$$\frac{d \log L_n(\lambda; Z(n))}{d\lambda} = -(1 + i_1 + \dots + i_{n-1}) + \frac{(i_1 + i_2 + \dots + i_n)}{\lambda}$$

si igualamos a cero la derivada para obtener el punto crítico y por lo tanto el estimador máximo verosímil,  $\hat{\lambda}$  es

$$\hat{\lambda} = \frac{(i_1 + i_2 + \dots + i_n)}{(1 + i_1 + \dots + i_{n-1})}$$

y por lo tanto el estimador máximo verosímil de la media de reproducción en un proceso de Galton-Watson que tiene distribución de descendientes Poisson( $\lambda$ ) es  $\hat{m} = \hat{\lambda}$ .

Ahora consideremos problemas no parametrizados, el primero de ellos es el de descendencia binaria: Suponga que el proceso comienza con  $Z_0$  ancestros y que  $Z_1, \dots, Z_n$  individuos fueron observados en las  $n$  primeras generaciones. Excepto  $Z_0$  esos números son todos pares  $Z_j = 2X_j$  y de la ecuación (2.4.10) su función de verosimilitud esta dada por:

$$L_n(p, Z_0, \dots, Z_n) = \binom{Z_0}{X_1} \binom{Z_1}{X_2} \dots \binom{Z_{n-1}}{X_n} p^{\sum_{i=1}^n X_i} (1-p)^{\sum_{i=0}^{n-1} Z_i - \sum_{i=1}^n X_i},$$

pero

$$\sum_{i=1}^n X_i = \frac{\sum_{i=1}^n 2X_i}{2}$$

$$\begin{aligned}
&= \frac{\sum_{i=1}^n Z_i}{2} = \frac{\sum_{i=0}^n Z_i - Z_0}{2} \\
&= \frac{Y_n - Z_0}{2}
\end{aligned}$$

entonces

$$\begin{aligned}
L_n(p, Z_0, \dots, Z_n) &= \\
&= C(Z_0, Z_1, \dots, Z_n) p^{(Y_n - Z_0)/2} (1-p)^{Y_{n-1} - (Y_n - Z_0)/2}. \quad (3.1.2)
\end{aligned}$$

De aquí que la log-verosimilitud está dada por:

$$\ln L = \ln C + \frac{Y_n - Z_0}{2} \ln p + \left[ Y_{n-1} - \frac{Y_n - Z_0}{2} \right] \ln(1-p)$$

y derivando para maximizar tenemos

$$\frac{\partial \ln L}{\partial p} = \frac{Y_n - Z_0}{2p} - \frac{[Y_{n-1} - (Y_n - Z_0)/2]}{(1-p)}$$

igualando la derivada de la log-verosimilitud a cero para obtener el valor de  $p$  que maximiza esta función tenemos:

$$\frac{Y_n - Z_0}{2\hat{p}} = \frac{Y_{n-1} - (Y_n - Z_0)/2}{1 - \hat{p}}$$

desarrollando esta igualdad nos queda

$$\hat{p} = \frac{Y_n - Z_0}{2Y_{n-1}},$$

y como

$$m = E[Z_1] = E[2X_1] = 2E[X_1] = 2p$$

ya que las  $X_i$  son v.a Bernoullis, utilizando el principio de invarianza para los estimadores máximo verosímiles, tenemos que el estimador de máxima verosimilitud para la media de reproducción es

$$\hat{m} = 2\hat{p} = \frac{Y_n - Z_0}{Y_{n-1}}$$

donde  $i_k = 0, 1, 2, \dots$  y  $\sum_{k=0}^{\infty} i_k = n_j$ . Dado que la expresión anterior es para cada valor de  $Z_j = n_j$  fijo, esto lo podemos tener en general en una notación simplificada para cualquier valor de  $Z_j$ , tomando la composición de la expresión pasada con la variable aleatoria  $Z_j$  obteniendo así

$$P\{Z_{j0} = i_0, Z_{j1} = i_1, \dots | Z_j = n_j\} \circ Z_j = \frac{Z_j! \prod_{k=0}^{\infty} p_k^{i_k}}{\prod_{k=0}^{\infty} i_k!} \quad (3.1.3)$$

y

$$\sum_{k=0}^{\infty} i_k = Z_j.$$

En otro caso la probabilidad es cero. Dado que

$$Z_j = \sum_{k=0}^{\infty} k Z_{j-1k}, \quad (3.1.4)$$

la verosimilitud de un arreglo  $\{Z_{jk}, j \leq n\}$  es el producto de las densidades dadas en (3.1.2), ya que cada renglón de esa matriz es independiente de otro, porque estamos haciendo  $n$  repeticiones del ensayo de observar para cada generación cuantos descendientes del tipo  $k$  para  $k = 0, 1, 2, \dots$  se tienen. De aquí, que la verosimilitud está dada por:

$$\frac{Z_0!}{\prod Z_{0k}!} \prod p_k^{Z_{0k}} \frac{(\sum k Z_{0k})!}{\prod Z_{1k}!} \prod p_k^{Z_{1k}} \dots \frac{(\sum k Z_{n-1k})!}{\prod Z_{nk}!} \prod p_k^{Z_{nk}} \quad (3.1.5)$$

aplicando el logaritmo a la verosimilitud tenemos

$$\begin{aligned} \log L_n(p) &= \log \left( \frac{Z_0!}{\prod Z_{0k}!} \right) + \log(\prod p_k^{Z_{0k}}) + \log \left( \frac{(\sum k Z_{0k})!}{\prod Z_{1k}!} \right) \\ &\quad + \log(\prod p_k^{Z_{1k}}) + \dots + \log \left( \frac{(\sum k Z_{n-1k})!}{\prod Z_{nk}!} \right) + \log(\prod p_k^{Z_{nk}}) \\ &= C + \sum_{k=0}^{\infty} Z_{0k} \log p_k + \sum_{k=0}^{\infty} Z_{1k} \log p_k + \dots + \sum_{k=0}^{\infty} Z_{nk} \log p_k \\ &= C + \sum_{k=0}^{\infty} \left( \sum_{j=0}^n Z_{jk} \right) \log p_k, \end{aligned}$$

por lo cual tenemos que encontrar el máximo de

$$\sum_{k=0}^{\infty} \left( \sum_{j=0}^n Z_{jk} \right) \log p_k, \quad (3.1.6)$$

respecto a  $p_1, p_2, \dots$ .

Sin embargo, si  $\{a_k\}_0^\infty$  es cualquier distribución de probabilidad y todas las sumas convergen,

$$\sum_0^\infty a_k \log p_k - \sum_0^\infty a_k \log a_k = \sum_0^\infty a_k \log \frac{p_k}{a_k} \leq \log \sum_0^\infty p_k = 0, \quad (3.1.7)$$

esto último por la desigualdad de Jensen ([55], pág. 72).

Para aplicar lo anterior a nuestro problema, multiplicamos y dividimos (3.1.5) por la progenie total, la cual podemos escribir como:

$$Y_n = \sum_{j=0}^n Z_j = \sum_{k=0}^\infty \sum_{j=0}^n Z_{jk}.$$

Y lo que obtenemos es

$$\sum_{k=0}^\infty \left( \sum_{j=0}^n Z_{jk} \right) \log p_k = Y_n \left( \sum_{k=0}^\infty \left( \frac{\sum_{j=0}^n Z_{jk}}{\sum_{k=0}^\infty \sum_{j=0}^n Z_{jk}} \right) \log p_k \right),$$

donde

$$\frac{\sum_{j=0}^n Z_{jk}}{\sum_{k=0}^\infty \sum_{j=0}^n Z_{jk}} = d_k,$$

es una distribución de probabilidad. De aquí que sólo tenemos que encontrar el máximo de

$$\sum_{k=0}^\infty d_k \log p_k,$$

y aplicando la ecuación (3.1.6), esto es equivalente a encontrar el máximo de

$$\sum_{k=0}^\infty d_k \log p_k - \underbrace{\sum_{k=0}^\infty d_k \log d_k}_{\text{no depende de } p_k} = \sum_{k=0}^\infty d_k \log \frac{p_k}{d_k} \leq 0$$

por lo tanto el valor de  $p_k$  que maximiza la verosimilitud debe maximizar

$$\sum_{k=0}^\infty d_k \log \frac{p_k}{d_k}.$$

**LEMA 3.1.1** *Sea  $\{P_{\theta,\sigma}, \theta \in \Theta, \sigma \in \Sigma\}$ , la clase de densidades con respecto a alguna medida  $\sigma$ -finita  $\mu$  sobre  $(\mathcal{X}, \mathcal{A})$ . Sea  $T$  alguna estadística sobre ese espacio. Si existe un estimador Máximo Verosímil  $\hat{\theta}$  de  $\theta$  basado sobre observaciones  $\mathcal{X}$  y  $\hat{\theta} = g \circ T$  para alguna  $g$ , entonces  $g$  es el estimador M.V. de  $\theta$  basado sobre observaciones de  $T(\mathcal{X})$ ,  $\mathcal{x} \in \mathcal{X}$ , la densidad de  $T$  es con respecto a la medida  $\mu T^{-1}$ .*

para la demostración de este Lema ver ([39], pág. 46).

Para aplicar este Lema consideremos las variables aleatorias  $Z_{nk}$ =número de individuos en la  $n$ -ésima generación con exactamente  $k$  descendientes.

**OBSERVACIÓN 3.1.1** *supongamos que hay un número infinito numerable de resultados en un experimento. Denotamos los posibles resultados del experimento por  $i_0, i_1, i_2, \dots$  y sea  $p_k = P(i_k)$  para  $k = 0, 1, 2, \dots$ . Con  $\sum_{k=0}^{\infty} p_k = 1$ . Supongamos que repetimos el experimento  $n$  veces. Y sea  $I_k$  la v.a que denota el número de veces que el resultado  $i_k$  ocurre en  $n$  ensayos  $k = 0, 1, 2, \dots$ . Si los ensayos son repetidos e independientes entonces la densidad discreta de las variables  $I_1, I_2, \dots$  es una multinomial generalizada*

$$f_{I_1, I_2, \dots}(i_1, i_2, \dots) = n! \prod_{k=0}^{\infty} p_k^{i_k} / \prod_{k=0}^{\infty} i_k!$$

donde  $i_k \in \{0, \dots, n\}$  y  $\sum_{k=0}^{\infty} i_k = n$ .

**TEOREMA 3.1.1** *En un proceso de Galton-Watson con  $Z_0$  ancestros los estimadores máximo verosímil de  $p_k$  y  $m$  basados en observaciones de  $\{Z_{jk}; j \leq n\}$  son respectivamente,*

$$\hat{p}_k = \sum_{j=0}^n Z_{jk} / Y_n$$

$$\hat{m}_n = (Y_{n+1} - Z_0) / Y_n$$

**Demostración:** ([39], pág 47)

En particular en nuestro caso tenemos, para  $Z_j = n_j$  y variables aleatorias  $Z_{nk}$  tales que

$$P\{Z_{j0} = i_0, Z_{j1} = i_1, \dots | Z_j = n_j\} = \begin{cases} 0 & \text{si } \sum_{k=0}^{\infty} i_k \neq n_j \\ n_j! \prod_{k=0}^{\infty} p_k^{i_k} / \prod_{k=0}^{\infty} i_k! & \text{si } \sum_{k=0}^{\infty} i_k = n_j \end{cases}$$

esto es, hacerla igual a cero; de aquí que si tomamos  $\hat{p}_k = d_k$  la función anterior es cero. De aquí entonces el estimador máximo verosímil para  $p_k$  es:

$$\hat{p}_k = \sum_{j=0}^n Z_{jk} / Y_n$$

y por el principio de invarianza de los estimadores máximo verosímiles tenemos que:

$$\hat{m} = \sum_{k=0}^{\infty} k \hat{p}_k = \sum_{k=0}^{\infty} k \frac{\sum_{j=0}^n Z_{jk}}{Y_n} \quad (3.1.8)$$

$$\begin{aligned} &= \frac{\sum_{k=0}^{\infty} \sum_{j=1}^{n+1} k Z_{j-1,k}}{Y_n} = \frac{\sum_{j=1}^{n+1} \sum_{k=0}^{\infty} k Z_{j-1,k}}{Y_n} \\ &= \frac{\sum_{j=1}^{n+1} Z_j}{Y_n} = \frac{Y_{n+1} - Z_0}{Y_n} \end{aligned} \quad (3.1.9)$$

Del Lema(3.1.1) tenemos el siguiente Corolario:

**COROLARIO 3.1.1** *El estimador máximo verosímil de  $m$  basado sobre observaciones  $(Z_0, \dots, Z_n)$  es*

$$\hat{m}(n) = \frac{Y_n - Z_0}{Y_{n-1}}$$

**Demostración:**

Encontramos en (3.1.8) que el estimador máximo verosímil para la media de reproducción basado en observaciones  $\{Z_j k, j \leq n\}$  es

$$\hat{m} = \frac{\sum_{k=0}^{\infty} \sum_{j=0}^n k Z_{jk}}{\sum_{k=0}^{\infty} \sum_{j=0}^n Z_{jk}} = H(Z_{jk})$$

para  $k = 0, 1, \dots$  y  $j \leq n$ . Ahora sea la estadística

$$Y_n = \sum_{k=0}^{\infty} \sum_{j=0}^n Z_{jk} = \sum_{j=0}^n Z_j,$$

por (3.1.9) el estimador de la media puede ser escrito como

$$\hat{m}(n) = \frac{Y_n - Z_0}{Y_{n-1}} = g(Y_n, Y_{n-1}) = T(Z_0, \dots, Z_n),$$

donde

$$T(a_1, a_2, \dots, a_n) = \frac{\sum_{j=1}^{n-1} a_j}{\sum_{j=0}^n a_j}.$$

y por el Lema(3.1.1)  $T$  es el estimador máximo verosímil para  $m$  basado sobre observaciones  $(Z_0, \dots, Z_n)$ .

□

Dada las propiedades de los estimadores de máxima verosimilitud, que se mencionaron anteriormente, es importante el estudio de sus propiedades asintóticas, ya que éstas permiten dar una idea de que tan buenos son los estimadores en el límite. Obviamente si  $m \leq 1$ ,  $(Y_{n+1} - Z_0)/Y_n \rightarrow 1 - \frac{Z_0}{Y_\infty}$  c.s. Esto también se da en el caso supercrítico con probabilidad  $q$ . El Teorema (2.6.1) nos dice que  $P\{Y_\infty = k\} = (1/k)P_{k,k-1}$ . De aquí que este límite satisfice:

$$\begin{aligned}
P\left[1 - \frac{1}{Y_\infty} \leq u\right] &= P\left[-\frac{1}{Y_\infty} \leq (u-1)\right] = P\left[\frac{1}{Y_\infty} \geq (1-u)\right] \\
&= P\left[Y_\infty \leq \frac{1}{(1-u)}\right] = \sum_{k \leq 1/(1-u)} P[Y_\infty = k] = \sum_{k \leq 1/(1-u)} (1/k)P_{k,k-1}
\end{aligned}$$

sin embargo si  $Z_n \rightarrow \infty$ , y nosotros consideramos por simplicidad el caso

$$\sum p_k k \log k < \infty, \text{ con } Z_0 = 1,$$

que es la condición de convergencia, vista en la sección de Teoremas Límite del capítulo anterior, para los casos subcrítico y supercrítico. Entonces por el Teorema (2.6.2)

$$Y_n/m^n \rightarrow Wm/(m-1) \text{ c.s.}$$

y

$$P\left[\frac{(Y_{n+1}-1)}{Y_n} \rightarrow m | Z_n \rightarrow \infty\right] = \frac{P\{((Y_{n+1}-1)/Y_n \rightarrow m) \cap B\}}{P[B]} = \frac{P[B]}{P[B]} = 1$$

ya que si  $B = \{\omega : Z_n(\omega) \rightarrow \infty\}$  entonces para toda  $\omega \in B$ ,  $1/Y_n \rightarrow 0$  y

$$\frac{Y_{n+1}}{Y_n} = m \frac{Y_{n+1}/m^{n+1}}{Y_n/m^n} \rightarrow m \frac{Wm/(m-1)}{Wm/(m-1)} = m \text{ c.s.}$$

por lo tanto, para toda  $\omega \in B$ ,

$$\frac{Y_{n+1}}{Y_n} - \frac{1}{Y_n} \rightarrow m$$

en otras palabras el estimador es fuertemente consistente, condicionado a no extinción.

Respecto a la razón de convergencia se tiene que

$$\frac{(Y_{n+1} - Z_0)}{Y_n} - m = \frac{Y_{n+1} - Z_0 - mY_n}{Y_n} = \frac{\sum_{j=0}^n (Z_{j+1} - mZ_j)}{Y_n}$$

cualquier término  $Z_{j+1} - mZ_j$  es la suma de  $Z_j$  v.a.i cada una con distribución  $p_k$  pero trasladada a la media cero, es decir:

$$E[Z_{j+1} - mZ_j] = E[Z_{j+1}] - mE[Z_j] = m^{j+1} - m \times m^j = 0.$$

De aquí que  $Y_{n+1} - Z_0 - mY_n$  tiene la distribución de una suma

$$\sum_{j=1}^{Y_n} \eta_{nj},$$

donde las  $\eta_{nj}$  son v.a.i.i.d con media cero. El Teorema de Límite Central enunciado en el Teorema (2.6.3) puede ser aplicado en este caso.

Vimos que

$$\hat{m}_n = (Z_1 + \dots + Z_{n+1}) / (Z_0 + \dots + Z_n)$$

es el estimador máximo verosímil para  $m$ , con respecto a la muestra

$$\{Z_0, \dots, Z_{n+1}\}, \text{ sobre } (\Omega, \mathcal{B}, P).$$

Los próximos teoremas establecen las leyes límites, sobre  $(\Omega, \mathcal{B}, P)$ , de  $\hat{m}_n$ . Sea  $A = \{\omega : Z_k > 0, k = 0, 1, 2, \dots\}$ , el conjunto de no-extinción, y  $P_A(\cdot) \equiv P(\cdot | A)$

**TEOREMA 3.1.1** Sea  $m > 1$ ,  $0 < \sigma^2 < \infty$ . Para cualquier  $Q \ll P_A$ , con  $A$  el conjunto de no-extinción,

(i)

$$\lim_{n \rightarrow \infty} Q\{\omega : \sigma^{-1}(Z_0 + \dots + Z_n(\omega))^{\frac{1}{2}}(\hat{m}(\omega) - m) \leq x\} = \Phi(x)$$

(ii)

$$\lim_{n \rightarrow \infty} Q\{\omega : \sigma^{-1}(1 + m + \dots + m^n)^{\frac{1}{2}}(\hat{m}(\omega) - m) \leq x\} = \int_0^{\infty} \Phi(x \cdot \sqrt{x}) dS(w)$$

donde

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$$

y  $S(w) = P_A(W < w)$  y  $W$  es el límite casi seguro de  $\{Z_n/m^n\}$ .

**Demostración:** ([41], y [57] pág. 79)

Con la ayuda de una sucesión  $\{\xi\}_{n \geq 1}$  de v.a.i.i.d sobre  $(\Omega, \mathcal{B}, P)$  cada una distribuida como  $Z_1$ , es posible escribir para  $\omega \in A$ , la  $\sum_{i=1}^{n+1} Z_i(\omega)$  como

$$Z_0 + \dots + Z_n(\omega) = \sum_{i=1}^{n+1} \xi_i(\omega).$$

Entonces

$$\frac{1 + Z_1 + \dots + Z_n}{1 + m + \dots + m^n} \rightarrow W' \text{ c.s.}$$

como  $\omega \in A$  entonces

$$\hat{m}(\omega) = \frac{(Z_1 + \dots + Z_{n+1}(\omega))}{(Z_0 + \dots + Z_n(\omega))}$$

$$\begin{aligned} Q\{\omega : \sigma^{-1}(Z_0 + \dots + Z_n(\omega))^{\frac{1}{2}}(\hat{m}(\omega) - m) \leq x\} &= \\ &= Q\left\{\omega : \frac{\sum_{i=1}^{Z_0 + \dots + Z_n(\omega)-1} (\xi_i(\omega) - m)}{\sigma[Z_0 + \dots + Z_n(\omega)]^{1/2}}\right\} \\ &= Q\{\omega : Y_n(1, \omega) \leq x\} \rightarrow \Phi(x) \end{aligned}$$

ya que por el Teorema 1, con  $1 + m + \dots + m^n = a_n$ ,  $Z_0 + \dots + Z_n(\omega) = \nu_n$  y  $\theta = W'$  tenemos

$$Y_n(t, \omega) = \sum_{i=1}^{Z_0 + \dots + Z_n(\omega) \cdot t} (\xi_i(\omega) - m) / \sigma[Z_0 + \dots + Z_n(\omega)]^{1/2}$$

y como

$$Q\{\omega : Y_n(t, \omega) \leq x\} \rightarrow Q\{\omega : W''(t, \omega) \leq x\}$$

donde  $W''$  es el proceso de Wiener, el cual para cada  $t$  se describe con una densidad *Normal*(0,  $t$ ); y como en este caso  $t = 1$ , esto prueba (i).

De igual forma para cualquier  $Q \ll P_A$

$$\begin{aligned} Q\{\omega : \sigma^{-1}(1 + m + \dots + m^n)^{1/2}(\hat{m}(\omega) - m) \leq x\} &= \\ &= Q\left\{\omega : \sigma^{-1}(1 + m + \dots + m^n)^{1/2} \frac{\sum_{i=1}^{Z_0 + \dots + Z_n(\omega)-1} (\xi_i(\omega) - m)}{[Z_0 + \dots + Z_n(\omega)]}\right\} \\ &= Q\left\{\omega : \left[\frac{(1 + m + \dots + m^n)}{Z_0 + \dots + Z_n(\omega)}\right]^{1/2} \times \frac{\sum_{i=1}^{Z_0 + \dots + Z_n(\omega)-1} (\xi_i(\omega) - m)}{\sigma[Z_0 + \dots + Z_n(\omega)]^{1/2}}\right\} \end{aligned}$$

y bajo las mismas hipótesis del Teorema 1 y usando el Teorema 2 tenemos

$$Y_n'(1, \omega) = \left(\frac{a_n}{\nu_n}\right)^{1/2} Y_n(1, \omega)$$

lo cual nos dice que

$$\begin{aligned} Q\{\omega : \sigma^{-1}(1+m+\dots+m^n)^{1/2}(\hat{m}(\omega) - m) \leq x\} &= \\ &= Q\{\omega : Y'_n(1, \omega) \leq x\} \rightarrow Q\{\omega : W^*(1, \omega)W_0'^{-1/2} \leq x\} \end{aligned}$$

y este límite es igual a  $\int_0^\infty \Phi(x\sqrt{w})dS(w)$ , donde  $S(w) = P_A(W < w)$ .

El siguiente Lemma establece que el Teorema anterior es también válido si en la conclusión reemplazamos  $Q(\cdot)$  por la medida de probabilidad  $P(\cdot | Z_n > 0)$ .

**LEMA 3.1.2** Sea  $A = \{\omega : Z_k > 0, k = 0, 1, 2, \dots\}$  como en el teorema anterior;  $0 < P(A) < 1$ . Considere cualquier sucesión de eventos  $\{A_n(x)\}_{n \geq 1}$ . Entonces

$$P_A(A_n(x)) \xrightarrow{n \rightarrow \infty} F(x)$$

implica que

$$P(A_n(x) | Z_n > 0) \xrightarrow{n \rightarrow \infty} F(x)$$

**Demostración:** ([41]).

este Lema implica que

$$P\{\omega : \sigma^{-1}(Z_0 + \dots + Z_n(\omega))^{1/2}(\hat{m}(\omega) - m) \leq x | Z_n > 0\} \rightarrow \Phi(x),$$

y esto nos permite dar un intervalo de confianza aproximado para  $m$ . Dada  $Z_n > 0$

$$I_m = \left( \hat{m} \pm \sigma \cdot t_{(1/2)\alpha}(Z_0 + \dots + Z_n)^{1/2} \right)$$

es un intervalo de confianza aproximado para  $m$  con el  $(1 - \alpha) \times 100\%$  de confianza,

donde  $t_{\frac{1}{2}\alpha}$  es tal que  $1 - \Phi(t_{\frac{1}{2}\alpha}) = \frac{1}{2}\alpha$ .

## ESIMACIÓN BAYESIANA (PÉRDIDA ESPERADA MÍNIMA):

Consideremos nuevamente la distribución de descendientes parametrizada por la familia Poisson( $\lambda$ ). Bayesianamente el estadístico debe asignar una distribución inicial, que considere represente su conocimiento a-priori del parámetro de interés. En particular, si decide trabajar con familias conjugadas para expresar su conocimiento inicial, decidirá utilizar una distribución inicial Gamma([53], pág. 116). La densidad Gamma esta dada por

$$G_{\alpha, \beta}(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \text{EXP}\{-\beta\lambda\} I_{(0, \infty)}(\lambda), \quad \alpha, \beta > 0.$$

Esta distribución para valores de  $\alpha$  y  $\beta$ , puede tomar muchas formas: simétricas, sesgadas y en particular si  $\alpha \rightarrow \infty$  la distribución Gamma se aproxima a una distribución Normal Estandar ([72], vol. 1, pág. 170).

Así la distribución final para el parámetro  $\lambda$  está dada por

$$\begin{aligned} P(\lambda|Z_{(n)}) &= \\ &= \frac{P(Z_{(n)}|\lambda)P(\lambda)}{P(Z_{(n)})} \text{ donde } Z_{(n)} = (Z_0, \dots, Z_n) \\ &= K \cdot e^{-\lambda(1+i_1+\dots+i_{n-1})} \cdot \lambda^{(i_1+\dots+i_n)} \cdot [G_{\alpha, \beta}(\lambda)] / P(Z_{(n)}) \\ &= M \cdot e^{-(1+i_1+\dots+i_{n-1}+\beta)\lambda} \cdot \lambda^{(\alpha+i_1+i_2+\dots+i_n)-1}. \end{aligned}$$

Dado lo anterior  $P(\lambda|Z_{(n)})$  es una densidad Gamma con parámetros  $\alpha^* = \alpha + i_1 + i_2 + \dots + i_n$  y  $\beta^* = 1 + i_1 + \dots + i_{n-1} + \beta$ .

Para estimar el parámetro en cuestión, aplicamos el criterio de Pérdida Esperada Mínima. Para esto el estadístico debe elegir una función de pérdida que considere apropiada al problema. En particular, si decide trabajar con una pérdida cuadrática, digamos  $L(\lambda, d) = (\lambda - d)^2$ , la solución de Bayes, es la Media de la distribución final del parámetro y la pérdida asociada, la Varianza de la distribución final ([53], cap. 7). Esto es

$$\hat{\lambda} = E[\lambda|Z_{(n)}] = \frac{\alpha^*}{\beta^*} = \frac{\alpha + i_1 + i_2 + \dots + i_n}{1 + i_1 + \dots + i_{n-1} + \beta}.$$

Si el estadístico no dispone de información inicial, o no quiere o no puede utilizar tal información; puede optar por trabajar con distribuciones iniciales de referencia (que son utilizadas como se mencionó al inicio del capítulo. para representar casos de

poca información inicial). En particular, si decide utilizar distribuciones conjugadas de referencia, debe hacer  $\alpha$  y  $\beta$  tender a cero ([71], pág. 198) en la distribución final para el parámetro  $\lambda$  y el estimador límite que se obtiene es

$$\hat{m} = \hat{\lambda} = \frac{i_1 + \dots + i_n}{1 + i_1 + \dots + i_{n-1}},$$

que coincide con el estimador de máxima verosimilitud. Se observa en este caso que  $P(\lambda)$  no es una distribución de probabilidades porque no tiene integral finita (se dice que es impropia).

Ahora otra vez tomamos casos en donde la función de descendientes no está parametrizada. Tenemos como caso particular, considerar primeramente descendientes binarios. Si el investigador decide asignar una distribución conjugada como se mencionó al principio del capítulo, la forma casi binomial de la verosimilitud en la ecuación (3.1.2), hace que una distribución **BETA** para  $p$ , sea conjugada para el problema ([71], pág. 176).

$$B_{a,b}(p) = \frac{\Gamma(a,b)}{\Gamma(a)\Gamma(b)} p^{a-1} (1-p)^{b-1} \quad a, b > 0 \text{ y } 0 \leq p \leq 1$$

esta distribución también toma muchas formas: simétricas, sesgadas y en particular una distribución uniforme en el intervalo unitario seleccionando  $a = b = 1$ . Dado que hemos asignado una distribución inicial para  $p$  y tenemos la verosimilitud, la distribución final para  $p$  queda como:

$$\begin{aligned} P(p|Z_{(n)}) &= \\ &= \frac{P(Z_{(n)}|p)P(p)}{P(Z_{(n)})} \text{ donde } Z_{(n)} = (Z_0, \dots, Z_n) \\ &= \frac{[C(Z_0, \dots, Z_n)p^{(Y_n - Z_0)/2}(1-p)^{Y_{n-1} - (Y_n - Z_0)/2}] [B_{a,b}(p)]}{P(Z_{(n)})} \\ &= \frac{K}{P(Z_{(n)})} p^{(Y_n - Z_0)/2 + a - 1} (1-p)^{Y_{n-1} - (Y_n - Z_0)/2 + b - 1} \\ &= \underbrace{\frac{K}{P(Z_{(n)})}}_{K'} \underbrace{p^{[(Y_n - Z_0)/2 + a] - 1} (1-p)^{[Y_{n-1} - (Y_n - Z_0)/2 + b] - 1}}_{\text{Núcleo de una distribución Beta.}} \end{aligned}$$

Por lo cual la distribución final para  $p$  es una Beta( $\alpha, \beta$ ), con  $\alpha = (Y_n - Z_0)/2 + a$  y  $\beta = Y_{n-1} - (Y_n - Z_0)/2 + b$ . Dependiendo del contexto del problema que se trabaje, el estadístico debe elegir la función de pérdida que considere más apropiada, para la estimación del parámetro en cuestión. Si decidimos nuevamente utilizar una función de pérdida cuadrática, digamos  $L(p, d) = (p - d)^2$ , aplicando el criterio de Mínima Pérdida Esperada se obtiene como solución de Bayes  $d^*$ , la Media de la distribución final del parámetro y como pérdida esperada la Varianza de la distribución final para el parámetro.

Por lo cual el estimador bayesiano para  $p$  quedaría como:

$$\begin{aligned}\hat{p} &= E[p|Z_0, \dots, Z_n] = \frac{\alpha}{\alpha + \beta} \\ &= \frac{(Y_n - Z_0/2) + a}{(Y_n - Z_0/2) + a + Y_{n-1} - (Y_n - Z_0/2) + b} \\ &= \frac{Y_n - Z_0 + 2a}{2(Y_{n-1} + a + b)}\end{aligned}$$

y como  $m = E[Z_1] = 2E[X_1] = 2p$ , debido al mismo procedimiento anterior, el estimador de Bayes para la media de reproducción es:

$$\begin{aligned}\hat{m} &= E[m|Z_0, \dots, Z_n] = E[2p|Z_0, \dots, Z_n] \\ &= 2E[p|Z_0, \dots, Z_n] = \frac{Y_n - Z_0 + 2a}{Y_{n-1} + a + b}\end{aligned}$$

El estadístico puede de elegir también como distribución inicial una distribución conjugada de referencia o no informativa haciendo tender a cero los parámetros  $a$  y  $b$  de la distribución BETA inicial. Se observa también que en este caso  $P(p)$  no es una distribución de probabilidades porque no tiene integral finita y por lo tanto es impropia, la estimación está dada por

$$\hat{m} = \frac{Y_n - Z_0}{Y_{n-1}}.$$

Desde un punto de vista Bayesiano un problema de estimación queda resuelto al reportar el valor  $d^*$  y la correspondiente pérdida esperada, sin embargo, poseemos mucha más información sobre  $\theta$ , específicamente conocemos toda la distribución final para el parámetro. Utilizando algunas pérdidas adecuadas, lo que obtenemos como

estimadores bayesianos para  $m$  son cuantiles de la distribución final del parámetro y esto obviamente nos brinda mayor información. En particular, si el estadístico decide trabajar con una pérdida del estilo  $L(d, p) = |p - d|$  (Pérdida absoluta), la solución de Bayes está dada por cualquier mediana de la distribución final para el parámetro ([53], cap. 7).

Si el estadístico decide trabajar con una pérdida del tipo

$$L(d, p) = \begin{cases} 1 & \text{si } |d - p| < b \\ 0 & \text{en otro caso} \end{cases}$$

En cuyo caso el estimador de bayes es  $d^* = \text{moda}$  ([53], cap. 7) de la distribución final del parámetro. Luego entonces, si queremos utilizar estas pérdidas para estimar la media de reproducción cuando parametrizamos la distribución de descendientes con una distribución Poisson o cuando tenemos descendencia binaria, sólo tenemos que encontrar la mediana y moda para las distribuciones Gamma y Beta respectivamente. Las expresiones para la moda de distribuciones Gamma y Beta pueden encontrarse en ([72], vol. 1, pág. 168; vol. 2, pág. 44).

Lo más relevante de este método de estimación Bayesiano es que al poseer una distribución final para la media de reproducción, podemos calcular la probabilidad de que este parámetro sea menor o igual a uno, para el caso de descendencia binaria tenemos:

$$P\{m \leq 1\} = P\{p \leq 1/2\} = \int_0^{1/2} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1} dp = \gamma,$$

es decir con probabilidad  $\gamma$  estamos en el caso crítico y subcrítico, o sea con probabilidad  $\gamma$  el proceso se extingue seguramente. Análogamente para el caso de distribución de descendientes parametrizada por una distribución Poisson.

En cuanto a Estimación Clásica se refiere, el problema de estimar la media de reproducción en un proceso de Galton-Watson está totalmente resuelto en la literatura, para casos en los que la distribución de descendientes está parametrizada por alguna distribución de probabilidad conocida y para el caso no paramétrico ([57], cap. 1). En relación a Estimación Bayesiana, la literatura sólo registra resueltos el caso de descendencia binaria y el caso donde la distribución de descendientes se parametriza con una distribución Poisson. Tratando de profundizar un poco más este estudio tomaremos el mismo esquema para resolver el caso general, esto es, cuando tenemos una distribución

de descendientes dada por  $\{p_k\}$  para  $k = 0, 1, 2, \dots$ , esto es, la distribución no está parametrizada, y suponiendo que como información tenemos el número de individuos obtenidos en  $n$  generaciones, esto es  $Z_0, \dots, Z_n$ . La distribución final para el vector  $V = (p_0, p_1, \dots)$  se puede calcular como sigue:

$$\begin{aligned}
 & P\{V|Z_0 = i_0, \dots, Z_n = i_n\} \\
 &= \frac{1}{P\{Z_0 = i_0, \dots, Z_n = i_n\}} P\{Z_0 = i_0|V\} \times P\{Z_1 = i_1|Z_0 = i_0; V\} \\
 &\quad \times \dots \times P\{Z_j = i_j|Z_0 = i_0, \dots, Z_{j-1} = i_{j-1}; V\} \\
 &\quad \times \dots \times P\{Z_n = i_n|Z_0 = i_0, \dots, Z_{n-1} = i_{n-1}; V\} P\{V\} \\
 &\propto 1 \times p_{i_0} \times P\{Z_2 = i_2|Z_0 = i_0; Z_1 = i_1; V\} P\{Z_j = i_j|Z_{j-1} = i_{j-1}; V\} \\
 &\quad \times \dots \times P\{Z_n = i_n|Z_{n-1} = i_{n-1}; V\} P\{V\} \\
 &\propto 1 \cdot p_{i_0} \cdot P\left[\sum_{j=1}^{i_1} X_{1j} = i_2\right] \dots P\left[\sum_{j=1}^{i_{n-1}} X_{n-1,j} = i_n\right] P\{V\}
 \end{aligned}$$

Como la última expresión es un producto de convoluciones, esto hace que el cálculo de la distribución final para el vector  $V$  sea bastante complejo, dado que calcular la distribución de sumas de variables aleatorias independientes no es trivial, a menos que la distribución de éstas sea cerrada bajo el operador suma. Todo esto hace que la elección de una familia conjugada sea para muchos casos difícil de encontrar.

Sin embargo, si en forma alternativa, pensamos en  $E_1, E_2, \dots, E_n$ ; una sucesión de experimentos donde  $E_k$  es el experimento en el que se observa el número de descendientes para cada uno de una colección de  $h_k$  miembros de la población en la  $k$ -ésima generación, supongamos que se tiene un número finito de descendientes;  $Z_0 = 1$  y que  $P\{\# \text{ de desc.} = k\} = p_k$  para  $k = 0, 1, \dots, n$ . Entonces:

$$\begin{aligned}
 P\{V|Z_1 = k_1\} &= \frac{P\{Z_1 = k_1|V\}\Pi(V)}{P\{Z_1 = k_1\}} \\
 &\propto P\{Z_1 = k_1|V\}\Pi(V) = p_{k_1}\Pi(V).
 \end{aligned}$$

Si siguiendo el mismo procedimiento anterior, si se decide usar como distribución inicial una familia conjugada, conviene tomar una distribución Dirichlet como conjugada para el problema, esto es  $\Pi(V)$  es una distribución Dirichlet para el vector  $V = (p_1, \dots, p_n)$

con vector de parámetros  $\alpha = (\alpha_1, \dots, \alpha_n)$  ([54], pág. 49). Entonces

$$P(V|Z_1 = k_1) \propto \prod_{i=0}^n p_i^{\alpha_i^{(1)}},$$

donde

$$\alpha_i^{(1)} = \begin{cases} \alpha_i - 1 & \text{si } i \neq k_1 \\ \alpha_i - 1 + 1 & \text{si } i = k_1 \end{cases}$$

y  $\Pi(V)$  es una distribución inicial Dirichlet para el vector  $V = (p_1, \dots, p_n)$ , con vector de parámetros  $\alpha = (\alpha_1, \dots, \alpha_n)$ .

Ahora sea  $Z_{nj}$  = número de individuos en la  $n$ -ésima generación con exactamente  $j$  descendientes. De aquí que al incorporar la información en el segundo paso tenemos:

$$\begin{aligned} & P[V|Z_1 = k_1, Z_{10} = i_0, \dots, Z_{1n} = i_n] \\ & \propto P[Z_{10} = i_0, \dots, Z_{1n} = i_n | Z_1 = k_1, V] P[Z_1 = k_1 | V] \Pi(V) \\ & = \frac{k_1! \prod_{k=0}^n p_k^{i_k}}{\prod_{k=0}^n i_k!} \cdot \prod_{i=0}^n p_i^{\alpha_i^{(1)}} \\ & \propto \prod_{j=0}^n p_j^{\alpha_j^{(2)}}, \end{aligned}$$

donde

$$\alpha_i^{(2)} = \begin{cases} \alpha_i^{(1)} & \text{si } i \neq k \\ \alpha_i^{(1)} + 1 & \text{si } i = k. \end{cases}$$

De igual manera en nuestro tercer paso al incorporar la información y considerar el hecho de que  $Z_n = \sum_{k=0}^n k Z_{n-1,k}$  tenemos:

$$\begin{aligned} & P[V|Z_1 = k_1, Z_{10} = i_0, \dots, Z_{1n} = i_n; Z_{20} = i_0, \dots, Z_{2n} = i_n] \\ & \propto P[Z_{10} = i_0, \dots, Z_{1n} = i_n | Z_1 = k_1, V] P[Z_1 = k_1 | V] \Pi(V) \\ & \quad \times P[Z_{20} = i_0, \dots, Z_{2n} = i_n | Z_1 = k_1, Z_{10} = i_0, \dots, Z_{1n} = i_n, V] \\ & = P[Z_{10} = i_0, \dots, Z_{1n} = i_n | Z_1 = k_1, V] P[Z_1 = k_1 | V] \Pi(V) \\ & \quad \times P[Z_{20} = i_0, \dots, Z_{2n} = i_n | Z_2 = k_2, V] \\ & \propto \prod_{i=0}^n p_i^{\alpha_i^{(2)}} \cdot \frac{k_2! \prod_{k=0}^n p_k^{i_k}}{\prod_{k=0}^n i_k!} \\ & \propto \prod_{j=0}^n p_j^{\alpha_j^{(3)}}, \end{aligned}$$

donde

$$\alpha_i^{(3)} = \begin{cases} \alpha_i^{(2)} & \text{si } i \neq k \\ \alpha_i^{(2)} + 1 & \text{si } i = k. \end{cases}$$

De aquí que para el  $l$ -ésimo paso tenemos:

$$\begin{aligned} P[V|\text{toda la inf.}] &= P[V|Z_1 = k_1, Z_{10} = i_0, \dots, Z_{1n} = i_n; \dots; Z_{l-1,0} = i_0, \dots, Z_{l-1,n} = i_n] \\ &\propto \prod_{i=0}^n p_i^{\alpha_i^{(n-1)}} \cdot \frac{k_{n-1}! \cdot \prod_{k=0}^n p_k^{i_k}}{\prod_{k=0}^n i_k!} \\ &\propto \prod_{j=0}^n p_j^{\alpha_j^{(n)}}, \end{aligned}$$

donde

$$\alpha_i^{(n)} = \begin{cases} \alpha_i^{(n-1)} & \text{si } i \neq k \\ \alpha_i^{(n-1)} + 1 & \text{si } i = k \end{cases}$$

y como

$$\prod_{j=0}^n p_j^{(\alpha_j^{(n)}+1)-1}$$

es el núcleo de una distribución Dirichlet, se obtiene entonces que la distribución final para el vector  $V$  dado que hemos incorporado toda la información paso a paso es:

$$P[V|\text{toda la inf.}] = \frac{\Gamma(\beta)}{\Gamma(\beta_1) \cdots \Gamma(\beta_n)} \prod_{i=0}^n p_i^{\beta_i - 1} \quad (3.1.10)$$

donde  $\beta_i = \alpha_i^{(n)} + 1$  y  $\sum \beta_i = \beta$ .

Si ahora lo que queremos es obtener la distribución final de la media de reproducción como función de la distribución final del vector  $V$ , esto es:

$$m = \sum_{k=0}^n k p_k.$$

nos encontramos con algunas dificultades. Utilizando el Teorema de Cambio de Variable o Función Característica no es posible dar explícitamente la forma funcional de la distribución final para la media de reproducción, por tanto es conveniente recurrir a

algun tipo de aproximación para dicha distribución final. Cabe señalar que si sólo nos interesa dar una estimación puntual para la media de reproducción y decidimos utilizar pérdida cuadrática, sabemos que el estimador bayesiano para la media de reproducción es  $E[m|toda\ la\ inf.]$ , esto es:

$$\hat{m} = E[\sum k p_k] = \sum_{k=0}^n k E\{p_k\}$$

y dado que la distribución final del vector  $V$  es Dirichlet sabemos que  $E\{p_k\} = \beta_k/\beta_0$  de aquí que:

$$\hat{m} = \sum_{k=0}^n k \frac{\beta_k}{\beta_0}$$

Volviendo al problema de describir la distribución final para  $m$ , resulta muy conveniente utilizar una técnica de aproximación de distribuciones. Un resultado que aparece en la literatura Bayesiana, es el de aproximar la distribución final de un parámetro  $\theta$ , con una distribución Normal ([54], pág. 215). Este resultado se obtiene a partir de propiedades límite de la función de verosimilitud ([54], pág. 213), que establecen que si se cumplen ciertas condiciones de regularidad ([54], pág. 208) y tamaños de muestra grande, la función de verosimilitud puede aproximarse por cierta distribución Normal, para valores de  $\theta$  en una vecindad de  $\hat{\theta}_n$ . Donde

$$\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n),$$

y

$$\lim_{n \rightarrow \infty} \hat{\theta}_n(X_1, \dots, X_n) = \theta_0 \text{ c.s.},$$

esto es, la sucesión de estadísticas  $\{\hat{\theta}_n(X_1, \dots, X_n); n = 1, 2, \dots\}$ , es una sucesión consistente. Se sigue de esta aproximación que si la distribución inicial del parámetro  $\theta$  se representa por una función de densidad positiva y continua en una pequeña vecindad del punto  $\theta_0$ , entonces la distribución final para el parámetro  $\theta$  estará altamente concentrada en una vecindad de  $\hat{\theta}_n$ , y en esta vecindad puede ser aproximada por la misma distribución Normal que aproxima a la verosimilitud en la misma vecindad.

En general, la aproximación de una distribución poco tratable, mediante un modelo que sea más simple, se puede plantear como un problema de decisión que naturalmente requiere a su vez de una distribución inicial y una función de utilidad o pérdida. El

grado de dificultad de este problema puede ser considerable especialmente si se toma en cuenta que el espacio de opciones es un espacio de funciones (el espacio de todas las posibles distribuciones que se pueden utilizar como aproximación). Una simplificación comúnmente aceptada es la de medir la pérdida en términos de la diferencia, en una escala apropiada, entre la densidad que la aproximación asigna al valor verdadero del parámetro y la densidad que asigna al mismo punto la distribución original. En tal caso se dice que la utilidad correspondiente es local. Una función de utilidad de esta clase que ha sido muy utilizada es la llamada logarítmica

$$U(q(\cdot), p(\cdot)) = \log q(\theta) - \log p(\theta),$$

puesto que en estas condiciones el suceso incierto relevante es precisamente el valor desconocido de  $\theta$ , la pérdida esperada a minimizar es la llamada discrepancia logarítmica. La discrepancia logarítmica entre una densidad de probabilidad  $p(\theta)$  y su aproximación  $q(\theta)$  es el valor de la integral

$$\delta\{p(\theta), q(\theta)\} = \int p(\theta) \log \frac{p(\theta)}{q(\theta)} d\theta.$$

Si  $\theta$  tiene una distribución discreta basta, como siempre, sustituir en la definición las densidades de probabilidad por probabilidades y la integral por una suma.

La aproximación  $q(\theta)$  será tanto mejor cuanto menor sea la discrepancia  $\delta\{p(\theta), q(\theta)\}$ .

Si se decide usar la medida Logarítmica, para encontrar una aproximación Normal a la densidad  $p(\theta)$ , se obtiene el siguiente Teorema:

**TEOREMA 3.1.2** *La mejor aproximación normal a una distribución  $p(\theta)$ , con media  $E(\theta)$  y desviación típica  $D(\theta)$ , es*

$$N\{\theta|E(\theta), D(\theta)\},$$

*esto es aquella distribución normal con la misma media y desviación típica que la distribución original.*

Este teorema se generaliza al caso de distribuciones multivariadas, en donde la mejor aproximación Normal Multivariada es aquella distribución Normal Multivariada con el mismo vector de medias y la misma matriz de covarianzas que la distribución multivariada original.

Para la demostración del Teorema ver ([71], págs. 178-198).

Es intuitivamente inmediato que, en general, la mejor aproximación Normal no necesariamente es una buena aproximación; esto es, no siempre tiene una discrepancia pequeña. Obviamente la calidad de la aproximación dependerá de que tan parecida sea la distribución verdadera a su aproximación normal.

En cualquier caso, se puede proponer como una primera aproximación a la distribución de  $m$ , la Normal Multivariada, con vector de medias y matriz de varianzas y covarianzas iguales a los de la distribución final obtenida en (3.1.7). Como la distribución final resultó Dirichlet sabemos que:

$$E[p_i] = \frac{\beta_i}{\beta_0}$$

$$Var[p_i] = \frac{\beta_i(\beta_0 - \beta_i)}{\beta_0^2(\beta_0 + 1)}$$

$$Cov(p_i, p_j) = \frac{-\beta_i\beta_j}{\beta_0^2(\beta_0 + 1)}$$

y por lo tanto  $V^t \sim N_n(\mu, \Sigma)$ , donde:

$$\mu = \begin{pmatrix} \frac{\beta_1}{\beta_0} \\ \vdots \\ \frac{\beta_n}{\beta_0} \end{pmatrix},$$

$$\Sigma = \begin{pmatrix} \frac{\beta_1(\beta_0 - \beta_1)}{\beta_0^2(\beta_0 + 1)} & \dots & \frac{-\beta_1\beta_n}{\beta_0^2(\beta_0 + 1)} \\ \vdots & \ddots & \vdots \\ \frac{-\beta_n\beta_1}{\beta_0^2(\beta_0 + 1)} & \dots & \frac{\beta_n(\beta_0 - \beta_n)}{\beta_0^2(\beta_0 + 1)} \end{pmatrix}.$$

Ahora sea  $K = (1, 2, \dots, n)$ ; como  $m$  es combinación lineal de las entradas de  $V$  esto es:

$$m = \sum_{j=0}^n k_j p_j = KV^t,$$

entonces  $m$  se distribuye aproximadamente normal con:

$$E[m] = E[KV^t] = KE[V^t] = K\mu$$

$$VAR[m] = VAR[KV^t] = KVAR[V^t]K^t = K\Sigma K^t$$

De aquí que podremos calcular como se hizo en el caso binario  $P[m \leq 1]$ , es decir, la probabilidad de estar en el caso crítico o subcrítico.

Otra forma para aproximar la distribución final de la media de reproducción es utilizar técnicas de simulación. Dado que la distribución final para el vector  $V$  resultó una distribución Dirichlet con vector de parámetros  $\beta^* = (\beta_1, \dots, \beta_n)$  y la distribución final para  $m$  es una combinación lineal de las entradas de el vector  $V$ , lo que se debe hacer si se quiere aproximar la distribución final para  $m$  utilizando simulación, es diseñar un algoritmo que simule vectores Dirichlet dado el vector de parámetros  $\beta^*$ . Después generar una muestra de estos vectores y valuar tantas veces el valor de  $m$ , como vectores hayamos generado, utilizando las entradas de estos vectores, las cuales representan la distribución de descendientes. Como opciones pueden aplicarse técnicas descriptivas y después Bondad y Ajuste para dar una distribución aproximada para  $m$ , o aplicar nuevamente la medida logarítmica para dar una aproximación Normal a la distribución  $m$ . Todo este procedimiento nos lleva nuevamente a calcular la probabilidad aproximada de que la media corresponda al caso crítico o subcrítico.

Se observa que, de no haber supuesto un número finito de descendientes, habría que parametrizar la distribución de descendientes para obtener un número finito de parámetros a estimar, y poder aplicar alguna técnica de inferencia, como en el caso de la ley de reproducción parametrizada por una distribución Poisson, tratada al inicio de este apartado.

### 3.2 ESTIMACIÓN DE LA PROBABILIDAD DE EXTINCIÓN

Consideraremos este problema como sigue. Sean  $X_1, \dots, X_n$  variables aleatorias independientes e idénticamente distribuidas que sólo toman valores enteros. Sea  $p_k = P(X_i = k)$ , y denotemos por  $f(s) = \sum p_k s^k$  a la función generadora de probabilidad de la distribución  $\{p_k\}$ . En el contexto de procesos de ramificación la variable aleatoria  $X_i$  representa el número de descendientes producidos al final de su vida por el  $i$ -ésimo individuo de una muestra aleatoria de  $n$  individuos. Dado esto, estimaremos la probabilidad de extinción  $q$  basada en una muestra de  $X_1, \dots, X_n$ , y además abordaremos el problema de estimar otras funcionales de  $f(s)$ , tales como la derivada de la f.g.p. evaluada en  $q$ ,  $f'(q)$ , cuya relevancia se ubica en el estudio de comportamientos asintóticos acerca de la probabilidad de extinción.

Sea la v.a  $Z_i$ , el número de individuos en la muestra con  $i$  descendientes. Los valores que toma la v.a  $Z_i$  son tales que  $\sum_{i=1}^{\infty} z_i = n$ , de aquí que la distribución conjunta de  $Z_1, Z_2, \dots$  dado  $V = (p_1, p_2, \dots)$  es:

$$P[Z_1 = z_1, Z_2 = z_2, \dots] = \frac{n! \prod_{i=1}^{\infty} p_i^{z_i}}{\prod_{i=1}^{\infty} z_i!} \quad (3.2.1)$$

y

$$\sum_{i=1}^{\infty} p_i = 1$$

De aquí que la función de verosimilitud está dada por:

$$L_n(V; Z_1, Z_2, \dots) = P[Z_1 = z_1, Z_2 = z_2, \dots | V]$$

aplicamos logaritmo a la verosimilitud antes de maximizar con respecto al vector de parámetros y tenemos

$$\begin{aligned} \log L_n(V) &= \log(c!c) + z_1 \log p_1 + z_2 \log p_2 + \dots + z_k \log p_k + \dots \\ &= \sum_{i=1}^{\infty} z_i \log p_i + C. \end{aligned}$$

Maximizar esta función implica encontrar el máximo de  $\sum_{i=1}^{\infty} z_i \log p_i$  como función de

$p_i$ . Dividiendo y multiplicando por  $\sum_{k=1}^{\infty} z_k$  tenemos:

$$l(p_i) = \sum_{k=1}^{\infty} z_k \sum_{i=1}^{\infty} \frac{z_i \log p_i}{\sum_{k=1}^{\infty} z_k} = n \sum_{i=1}^{\infty} \frac{z_i}{n} \log p_i$$

y de igual forma que en la sección anterior el valor que maximiza esta verosimilitud se obtiene cuando

$$\hat{p}_k(n) = \frac{z_k}{n} = \frac{N^{\circ} \text{ de } X_i = k}{n},$$

esta probabilidad claramente depende de  $n$ . Por la ley de los grandes números  $\hat{p}_k(n)$  es fuertemente consistente y como

$$E[\hat{p}_k(n)] = \frac{E[N^{\circ} \text{ de } X_i = k]}{n} = \frac{np_k}{n} = p_k$$

nuestro estimador es insesgado. Sea  $\hat{f}_n(s) = \sum \hat{p}_k(n) s^k$ , la función generadora de probabilidad empírica. Más adelante consideraremos estimaciones de funcionales de  $f(s)$  por correspondientes funcionales de  $\hat{f}_n(s)$ .

Dado que  $\{\hat{p}_k(n)\}$  son los estimadores máximo verosímiles de  $\{p_k\}$ ,  $\hat{f}_n$  es un estimador máximo verosímil de  $f$  y cualquier funcional de  $\hat{f}_n$  es un estimador máximo verosímil de la correspondiente funcional de  $f$  (todo esto por la propiedad de invarianza en los estimadores de máxima verosimilitud). Sin embargo, para resolver el problema de estimar  $q$  y  $f'(q)$ , necesitamos saber más acerca del comportamiento de  $\hat{f}_n$  y sus derivadas como estimaciones de  $f$  y sus derivadas. Sea  $f^k(s)$  la función que denota la  $k$ -ésima derivada de  $f$  en  $s$ ,  $f^0 \equiv f$ , y similarmente para  $\hat{f}_n$ . Como una observación, se verifica fácilmente realizando un procedimiento análogo al utilizado en el Lema 2.2.1 (i) que  $\hat{f}_n$  es una función convexa. El siguiente Teorema presenta una serie de propiedades atractivas de  $\hat{f}_n^k$  como un estimador de  $f^k$  que necesitaremos posteriormente en nuestro desarrollo.

**TEOREMA 3.2.1** *Para cualquier  $k \geq 0$  y  $0 \leq s \leq s_1$ , donde  $s_1$  es cualquier número menor que uno,  $\hat{f}_n^k(s)$  es un estimador fuertemente consistente de  $f^k(s)$  en el sentido de que  $\hat{f}_n^k(s) \rightarrow f^k(s)$  c.s. cuando  $n \rightarrow \infty$  y uniformemente para  $0 \leq s \leq s_1$ . Si  $f^k(1-) < \infty$ : todo lo anterior también es válido con  $s_1 = 1$ .*

### Demostración: ([45])

Primero mostraremos que  $\hat{f}_n^k(s) \rightarrow f^k(s)$  c.s. para cualquier  $0 \leq s < 1$  y  $k \geq 0$ . Sea

$$C(j) = k! \binom{j}{k};$$

$$\begin{aligned} |\hat{f}_n^k(s) - f^k(s)| &= \left| \sum_{j=m+1}^{\infty} C(j)(\hat{p}_j - p_j)s^{j-k} + \sum_{j=0}^m C(j)(\hat{p}_j - p_j)s^{j-k} \right| \\ &\leq \sum_{j=m+1}^{\infty} C(j)s^{j-k} + \max_{0 \leq j \leq m} C(j)|\hat{p}_j - p_j|(m+1). \end{aligned}$$

Dado que el primer término en la desigualdad, es la cola de una serie convergente y el segundo seleccionando  $m$  grande y aplicando la Ley Fuerte hace que  $\max |\hat{p}_j - p_j| \rightarrow 0$  c.s. para cualquier valor fijo  $m$ , y por lo tanto  $\hat{f}_n^k(s) \rightarrow f^k(s)$  c.s. La convergencia uniforme en  $s$  se sigue dado que el primer término de la desigualdad es la cola de una serie que converge uniformemente para  $0 \leq s \leq s_1 < 1$  y el segundo término converge casi seguramente, pero su convergencia no es debida a  $s$  sino a las  $\{p_j\}$  y estas no dependen de  $s$ . Además  $f^k(1-) < \infty$ , entonces el  $k$ -ésimo momento factorial de  $X_i$  es finito, y dado que  $\hat{f}_n^k$  es justo el  $k$ -ésimo momento factorial muestral, la Ley Fuerte de los grandes números dice que  $\hat{f}_n^k(1-) \rightarrow f^k(1-)$  c.s. y la uniformidad sobre  $0 \leq s \leq 1$  se sigue como arriba.

También tenemos que:

$$\begin{aligned} E[\hat{f}_n^k(s)] &= E \left[ \frac{\sum_{k=j}^{\infty} k! \hat{p}_k(n) s^{k-j}}{(k-j)!} \right] \\ &= \sum_{k=j}^{\infty} \frac{k! E[\hat{p}_k(n)] s^{k-j}}{(k-j)!} \\ &= \sum_{k=j}^{\infty} \frac{k! p_k(n) s^{k-j}}{(k-j)!} = f^j(s). \end{aligned}$$

Para obtener una convergencia a una distribución Normal, escribimos  $\epsilon_{ik}$  la función indicadora del evento  $\{X_i = k\}$ . Para cualquier  $0 \leq s < 1$

$$\hat{f}_n(s) = \sum_{i=1}^n \frac{\left\{ \sum_{k=0}^{\infty} \epsilon_{ik} s^k \right\}}{n}.$$

Las variables aleatorias

$$\sum_{k=0}^{\infty} c_{ik} s^k$$

con  $1 \leq i \leq n$  son independientes e idénticamente distribuidas y

$$E \left[ \sum_{k=0}^{\infty} c_{ik} s^k \right] = \sum_{k=0}^{\infty} E[c_{ik}] s^k = \sum_{k=0}^{\infty} p_k s^k = f(s)$$

y

$$\begin{aligned} \text{Var} \left[ \sum_{k=0}^{\infty} c_{ik} s^k \right] &= E \left[ \left( \sum_{k=0}^{\infty} c_{ik} s^k \right)^2 \right] - f^2(s) \\ &= \sum_{k=0}^{\infty} E[c_{ik}] s^{2k} - f^2(s) = f(s^2) - f^2(s) \end{aligned}$$

dado que  $e_{ij} \cdot c_{ik}$  es cero para  $j \neq k$ . Esta varianza es cero si y sólo si la ley de reproducción es degenerada. Si este no es el caso, por el Teorema Central de Límite tenemos:

$$\frac{\sqrt{n}(\hat{f}_n(s) - f(s))}{\sqrt{f(s^2) - f^2(s)}} \xrightarrow{n \rightarrow \infty} N(0, 1) \quad (3.2.2)$$

y por lo visto acerca de  $\hat{f}_n$  se sigue que  $\hat{q}$

$$\hat{q}_n = \inf\{s \geq 0; \hat{f}_n(s) = s\}$$

es un estimador Máximo verosímil consistente para  $q$ , además

**TEOREMA 3.2.2** Si  $m > 1$  y  $p_0 > 0$ , entonces  $(\hat{q}_n - q)\sqrt{n}$  es asintóticamente Normal con media cero y varianza  $\{f(q^2) - q^2\}/\{1 - f'(q)\}^2$ . Si  $m = 1$  y  $0 < \sigma^2 < \infty$ , entonces

$$\lim_{n \rightarrow \infty} P[(1 - \hat{q}_n)\sqrt{n} \leq u] = \begin{cases} \Phi(u/2) & \text{si } u \geq 0 \\ 0 & \text{en otro caso} \end{cases}$$

Si  $m < 1$ , entonces con probabilidad uno  $\hat{q}_n$  es igual a uno de algun índice en adelante.

**Demostración:** ([39], pág. 50) Dado que  $\hat{f}_n(\hat{q}_n) = \hat{q}_n$ , Por el Teorema del Valor Medio tenemos que:

$$\hat{f}_n(q) = \hat{q}_n - (\hat{q}_n - q)\hat{f}'_n(s_n)$$

para alguna  $s_n$  entre  $\hat{q}_n$  y  $q$  y  $s_n$  depende de  $X_1, \dots, X_n$ , y dado que  $f(q) = q$  se tiene que:

$$\begin{aligned} \{\hat{f}_n(q) - f(q)\}\sqrt{n} &= \{[\hat{q}_n - (q_n - q)\hat{f}'_n(s_n)] - f(q)\}\sqrt{n} \\ &= \{[\hat{q}_n - (q_n - q)\hat{f}'_n(s_n)] - q\}\sqrt{n} \\ &= (\hat{q}_n - q)\sqrt{n}(1 - \hat{f}'_n(s_n)) \end{aligned}$$

sabemos que el lado izquierdo de esta última igualdad, por la ecuación (3.2.2) es asintóticamente  $N(0, f(q^2) - q^2)$  y  $\hat{f}'_n(s_n) \rightarrow f'(q)$  c.s ya que por el teorema anterior  $\hat{f}'_n(s) \rightarrow f'(s)$  c.s uniformemente en una vecindad de  $q$ ; esto prueba el primer inciso.

La segunda parte es una consecuencia de argumentos similares y propiedades de convexidad de  $\hat{f}_n$  y de que  $\hat{f}_n(\hat{q}_n) = \hat{q}_n$ , y  $f_n(1) = 1$ . Tenemos entonces

$$\begin{aligned} P[(1 - \hat{q}_n)\sqrt{n} \leq u] &= P[1 - u/\sqrt{n} \leq \hat{q}_n] = P[\hat{f}_n(1 - u/\sqrt{n}) \geq \hat{f}_n(\hat{q}_n)] \\ &= P[\hat{f}_n(1 - u/\sqrt{n}) \geq \hat{q}_n] = P[\hat{f}_n(1 - u/\sqrt{n}) \geq 1 - u/\sqrt{n}] \end{aligned}$$

y dado que  $\hat{f}_n''(s)$  existe para toda  $s$ , podemos expandir  $\hat{f}_n(\cdot)$  en series de Taylor alrededor de uno y quedarnos con el polinomio de grado dos, esto es:

$$\begin{aligned} \hat{f}_n(1 - u/\sqrt{n}) &= \hat{f}_n(1) + \hat{f}'_n(1)[1 - u/\sqrt{n} - 1] + \frac{\hat{f}_n''(s_n)}{2}[1 - u/\sqrt{n} - 1]^2 + R_3(s_n) \\ &= \hat{f}_n(1) + \hat{f}'_n(1)[u/\sqrt{n}] + \hat{f}_n''(s_n)u^2/2n + R_3(s_n) \end{aligned}$$

donde  $1 - u/\sqrt{n} \leq s_n \leq 1$  y  $s_n$  depende de  $X_1, \dots, X_n$  y es tal que  $R_3(s_n) \rightarrow 0$  cuando  $s_n \uparrow 1$ . De aquí que:

$$\begin{aligned} P[(1 - \hat{q}_n)\sqrt{n} \leq u] &= P[1 - \hat{f}'_n(1)u/\sqrt{n} + \hat{f}_n''(s_n)u^2/2n \geq 1 - u/\sqrt{n}] \\ &= P[\hat{f}_n''(s_n)u/n \geq 2\{\hat{f}'_n(1) - 1\}/\sqrt{n}] \\ &= P[(\hat{f}'_n(1) - 1)\sqrt{n}/\hat{f}_n''(s_n) \leq u/2] \rightarrow \Phi(u/2) \end{aligned}$$

Como  $1 - u/\sqrt{n} < s_n < 1$  esto implica que  $s_n \rightarrow 1$  y

$$\hat{f}_n''(s_n) \rightarrow \hat{f}_n''(1) = \text{VAR}(X_i) = \sigma^2 \text{ c.s}$$

Sabemos que

$$(\hat{f}'_n(1) - 1)\sqrt{n} \rightarrow N(0, \sigma^2) \text{ en Distribución}$$

ya que

$$f'_n(1) = \sum k p_k = \frac{\sum_1^n X_i}{n} y m = f'(1) = 1.$$

Finalmente, si  $m < 1$ , entonces por la Ley de los grandes números

$$\hat{f}'_n(1) = \frac{\sum_1^n X_i}{n}$$

es casi seguramente, estrictamente menor que uno para algun indice. De ese indice en adelante  $\hat{q}_n = 1$ .

Otra posible enfoque de estimación para tratar de estimar la probabilidad de extinción es utilizar inferencia Bayesiana, lo que habría que estimar aquí sería  $\hat{f}_n$ , ya que la estimación de la probabilidad de extinción  $q$  es la más pequeña solución a la ecuación  $\hat{f}_n(s) = s$ . Siguiendo el enfoque de la sección anterior habría que asignar una distribución inicial para el vector  $V$ , luego si no se parametriza la ley de reproducción hay que suponer un número finito de descendientes y se obtiene una distribución multinomial como verosimilitud asociada al vector, aplicamos el Teorema de Bayes y obtenemos la distribución final para  $V$ . Como se ha venido planteando desde la sección anterior, el estadístico puede decidir utilizar familias conjugadas para expresar su conocimiento apriori sobre el vector  $V$ . Por el contrario si no se restringe la distribución de descendientes a un número finito, se debe proceder a parametrizar dicha distribución con alguna distribución de probabilidad conocida. para obtener un número finito de parámetros a estimar; luego se obtiene la verosimilitud y puede obtenerse la distribución final. De igual manera que para la media de reproducción, la función  $\hat{f}_n$  es combinación lineal de las entradas del vector  $V$ , y utilizando Teorema de Cambio de Variable o Función Característica no es posible dar explícitamente la forma funcional de su distribución final. Podemos recurrir como antes a técnicas de simulación para encontrar una aproximación a la distribución final, y un procedimiento a seguir podría ser el mismo que ya comentamos.

## COMENTARIOS FINALES

Este trabajo se ubica en un área que aún está poco explorada. Se trata de la Inferencia Estadística para Procesos Estocásticos. Existe una gran cantidad de contribuciones en el área de procesos, especialmente en lo que se refiere a procesos de ramificación, pero en su inmensa mayoría se ocupan de tratamientos más bien deductivos y no inferenciales. Por otra parte, cuando se revisa la literatura sobre inferencia en procesos, se encuentra que el enfoque predominante es el llamado frecuentista y que muy pocos resultados han sido obtenidos mediante la alternativa Bayesiana.

Muy específicamente en los procesos de Galton-Watson, se ha mostrado que el problema fundamental es establecer la probabilidad de extinción, puede reducirse a la estimación de la media de reproducción  $m$ . Bajo diferentes hipótesis, en el modelo y en el muestreo, es posible obtener estimaciones frecuentistas, tanto puntuales como de intervalos, para  $m$ .

El enfoque Bayesiano, como se ha indicado, ha recibido mucho menos atención. Sin embargo, existen reportes en la literatura para algunos de los casos más sencillos. En este trabajo el tratamiento Bayesiano se ha aplicado a variantes más generales del proceso de Galton-Watson. En términos generales y a nivel conceptual es claro, como ocurre con muchas otras aplicaciones del paradigma de Bayes, que obtener una distribución de probabilidad para el parámetro de interés, ofrece más y mejores posibilidades de análisis del fenómeno. Concretamente, en el caso tratado en este trabajo la probabilidad final del evento  $m \leq 1$ , describe la seguridad que puede tener el investigador en la extinción del proceso.

Por otra parte, la aplicación del enfoque de Bayes, aun cuando es formalmente muy simple, no necesariamente puede llevarse acabo de manera automática. Especial cuidado requiere la especificación de la distribución inicial. En este trabajo se han empleado distribuciones conjugadas, informativas y de referencia que conducen a resultados intuitivamente aceptables y razonablemente simples. Sin embargo no es esta la única posibilidad, queda abierta la investigación en este sentido. Adicionalmente, ocurre que para una variedad de casos, la distribución final de  $m$  no puede ser determinada por completo en términos analíticos. En esas condicione, se pueden ensayar aproximaciones varias, utilizando técnicas de simulación o bien recurriendo a la normalidad asintótica de la distribución final. Estas posibilidades no se exploraron en este trabajo y quedan también planteadas como líneas de actividad futura.

En resumen, el enfoque Bayesiano promete resultados más completos para la inferencia en procesos de Galton-Watson, pero aún en casos relativamente simples entraña un grado de dificultad que requiere un estudio aún más profundo.

# Referencias

- [1] Kendall, G.D.: *Branching Processes Since 1873*. J. London Math. Soc. **41**, 385-406, 1966.
- [2] Galton, F.: *Problem 4001*. Educational Times **17**, 1873.
- [3] Watson, H.W., Galton, F.: *On The Probability of the extinction of families*. J. Anthropol. Inst. Great B. and Ireland **4**, 138-144, 1874.
- [4] Malthus, T.R.: *An Essay on the Principle of Population, as it Affects the Future Improvements of Society, with Remarks on the Speculations of Mr. Godwin, M. Condorcet, and Other Writers*. 5th ed. John Murray, London, 1817. (1st. ed 1798).
- [5] Moser, L.F.: *Die Gesetze der Lebensdauer*. Veit Berlin. 1839.
- [6] Benoiston de Châteauneuf, L.F.: *Mémoire Sur la Duréé des Families Nobles de France (Read Aug. 31, 1844 and Feb. 16, 1845)*. Memoires de l'Académie Royale des Sciences Morales et Politiques de l'Institut de France **5**, 753-794, 1847.
- [7] Fisher, R.A.: *On the dominance ratio*. Proc. Royal. Soc. Edingburgh **42**, 321-341, 1922.
- [8] Heyde, C.C., and Seneta E.: *The Simple Branching Process, a Turning Point Test and a Fundamental Identity: A Historical Note on I.J. Bienaymé*. Biometrika **59**, 680-683, 1972.
- [9] Galton, F.: *Hereditary Genius*. London, 1869.
- [10] de Candolle, A.: *Histoire des sciences et des Savants depuis deux siècles*. H. Georg, Geneva, 1873.

- [11] Schröder, E.: *Ueber iterierte Functionen*. Math. Ann. **3**, 296-322, 1871.
- [12] Kendall, D.G.: *Branching Processes since 1873*. J. London Math. Soc. **41**, 385-406, 1966.
- [13] Haldane, J.B.S.: *A Mathematical theory of Natural and Artificial Selection, V: Selection and Mutation*. Proc. Cambridge Philos. Soc. **23**, 838-844, 1927.
- [14] Steffensen, J.F.: *Om Sandsyndigheden for at Afkommet uddor*. Matematisk Tidsskrift **B**, 19-23, 1930.
- [15] Erlang, A.k.: *Opgave til Losning*. Matematisk Tidsskrift **B**, 36, 1929.
- [16] Süssmilch, J.P.: *Die göttliche Ordnung in den Veränderungen des menschlichen Geschlechts, aus der Geburt, dem Tode, und der Fortpflanzung desselben erwiesen*. Berlin, 1741.
- [17] Euler, L.: *Introductio in analysin infinitorum*. I. Bousquet, Lausanne, 1748. Available as: Leonhardi Euleri opera omnia, Series prima VIII, Teubner, Leipzig and Berlin, 1922.
- [18] Euler, L.: *Recherches générales sur la mortalité et la multiplication du genre humain*. Histoire de l'Académie Royale des Sciences et Belles-Lettres, année 1760, 144-164, Berlin, 1767. Available as: Leonhardi Euleri opera omnia, Series prima VII, 79-100. Teubner, Leipzig and Berlin, 1923.
- [19] McKendrick, A.G.: *Studies on the Theory of Continuous Probabilities, with special reference to its bearing on Natural Phenomena of a progressive nature*. Proc. London Math. Soc. Ser. II **13**, 401-416, 1914.
- [20] Yule, G.U.: *A Mathematical Theory of Evolution based on the conclusions of Dr. J.C. Willis, F.R.S.* Trans. Royal Soc. London **B** **213**, 21-87, 1924.
- [21] Furry, W.H.: *On fluctuation phenomena in the passage of high energy electrons through lead*. Phys. Rev. **52**, 569-581, 1937.
- [22] Volterra, V.: *Lecons sur la théorie mathématique de la lutte pour la vie*. Gauthier-Villars, Paris, 1931.

- [23] Feller, W.: *Die Grundlagen der Volterraschen Theorie des Kampfes ums Dasein in wahrscheinlichkeitstheoretischer Behandlung*. Acta Biotheoretica 5, 11-40, 1939.
- [24] Kendall, D.G.: *On the generalized 'birth-and-death' process*. Ann. Math. Statist. 19, 1-15, 1948.
- [25] Kendall, D.G.: *Stochastic processes and population growth*. J. Royal Statist. Soc. B 11, 230-264, 1949.
- [26] Bellman, R. and Harris, T.E.: *On the theory of age-dependent stochastic branching processes*. Proc. Nat. Acad. Sci. 34, 601-604, 1948.
- [27] Ney, P.E.: *Generalized branching processes I and II-III*. J. Math. 8, 316-350, 1964.
- [28] Harris, T.E.: *The Theory of Branching Processes*. Springer Verlag, Berlin, 1963.
- [29] Ryan, T.R. Jr.: *On age-dependent branching processes*. Ph.D. Dissertation. Cornell University, 1968.
- [30] Crump, K.S. and Mode C.J.: *A general age-dependent branching process I and II*. J. Math. Anal. Appl. 24 and 25, 494-508 and 8-17, 1968 and 1969.
- [31] Jager, P.: *A general stochastic model for population development*. Skand. Aktuarietidskr. 52, 84- 103, 1969.
- [32] Kolmogorov, A.N. and Dmitriev, N.A.: *Velvyasciesya slucaynye processy (Branching Stochastic Processes)*. Dokl. Akad. Nauk SSSR 56, 7-10, 1947.
- [33] Kolmogorov, A.N. and Sevast'yanov, B.A.: *The calculation of final probabilities of branching random processes*. Dokl. Akad. Nauk SSSR 56, 783-786, 1947.
- [34] Sevast'yanov, B.A.: *Age-dependent branching processes*. Theory. Prob. Appl. 9, 521-537, 1964.
- [35] Mode, C.J.: *Multitype Branching Processes*. Elsevier, New York, 1971.
- [36] Kolmogorov, A.: *Zur Lösung einer biologischen Aufgabe*. Izvestiya Nauchno-issledovatel'skogo instituta matematiki i mehaniki pritomskom Gosudarstvennom Universitete 2, 1-6, 1938.

- [37] Yaglom, A.M.: *Certain Limit Theorems of the Theory of Branching Processes*. Dokl. Acad. Nauk SSSR **56**, 795-798, 1947.
- [38] Joffe, A.: *On The Galton-Watson Branching Processes with Mean Less Than One*. AMS **38**, 264-266. 1967.
- [39] Jagers P.: *Branching Processes with Biological Applications*. London, J. Wiley 1975.
- [40] Hoel, P.G., Port, S.C., Stone, Ch. J.: *Introduction to Stochastic Processes*. Houghton Mifflin, 1972.
- [41] Dion, J.P.: *Estimation of the Mean and the Initial Probabilities of Branching Process*. J. Appl. Prob. **11**, 687-694, 1974.
- [42] Nagayev, A.V.: *On estimating the Expected Number of Direct Descendants of a Particle in a Branching Process*. Theory Prob. Appl. **12**, 314-320, 1967.
- [43] Shinozaki, N.: *Simultaneous Estimation of Location Parameters Under Quadratic Loss*. The Annals of Statistics **12**, 322-335, 1984.
- [44] Pakes, A.G.: *Some Limit Theorems for the Total Progeny of a Branching Process*. Adv. Appl. Prob. **3**, 176-172, 1971.
- [45] Stigler, S.M.: *The Estimation of the Probability of Extinction and Other Parameters Associated with Branching Process*. Biometrika **58**, 499-507, 1971.
- [46] Alam, K., Mitra, A.: *An Empirical Bayes Estimate of Multinomial Probabilities*. Commun. Statist. Meth. **15**, 3103-3127, 1986.
- [47] Nayak, T.K., Naik N.D.: *Estimating Multinomial Cell Probabilities Under Quadratic Loss*. The Statistician **38**, 3-10, 1989.
- [48] Athreya, K.B., Ney, P.E.: *Branching Processes*. Springer-Verlag, 1972.
- [49] Taylor, H.M., Karlin, S.: *An Introduction to Stochastic Modeling*. Academic Press, 1984.
- [50] Feller, W.: *Introducción a la Teoría de Probabilidades y sus Aplicaciones, Vol. 1*. Limusa, 1975.

- [51] Lee, P.M.: *Bayesian Statistics: An Introduction*. Edward Arnold, 1989.
- [52] Berger, J.O.: *Statistical Decision*. Springer-Verlag, 1985.
- [53] Smith J.Q.: *Decision Analysis: A Bayesian Approach*. Chapman and Hall, 1984.
- [54] DeGroot, M.: *Optimal Estistical Decisions*. McGraw Hill, 1970.
- [55] Mood, A., Graybill, F.: *Introduction to the Theory of Statistics*. McGraw Hill, 1986.
- [56] Kotz, S., Johnson N.: *Encyclopedia of Statistical Sciences*. John Wiley & Sons 5, 1981.
- [57] Nanthi k.: *Statistical Estimation for Stochastic Processes*. Queen's University Kingston, 62, 1983.
- [58] Banard, G.A.: *The Frequency Justification of Certain Sequential Tests*. Biometrika 39, 144-150, 1952.
- [59] Bayes T.: *An Essay towards solving a problem in the doctrine of chances*. Biometrika 45, 293-315, 1763.
- [60] Bernardo J.M.: *Reference posterior distributions for Bayesian inference*. J. Roy. Statist. Soc. B 41, 113-147, 1979.
- [61] Bernardo J.M.: *An information-theoretical approach to aproximations in statistics*. Conferencia invitada. 12 European Meeting of Statistics, Varna(Bulgaria), 1980.
- [62] Box, G.E.P., Tiao, G.C.: *Bayesian inference in Statistical Analysis*. Addison-Wesley, 1973.
- [63] Hartigan, J.A.: *Invariant prior distributions*. Ann. Math. Statist. 35, 836-845, 1964.
- [64] Jaynes, E.T.: *Prior probabilities*. IEEE Trans. Systems, Science and Cybermetics, SCC-4, 227- 291, 1968.
- [65] Good, I.J.: *What is the use of a distribution?*. En em Multivariate Analysis (Krishnaiah, ed.), 2, 183-203, 1969. New York: Academic Press.

- [66] Jaynes, E.T.: *Marginalization and prior probabilities*. Studies in Bayesian Statistics (A. Zellner, ed.). Amsterdam: North-Holland, 1980.
- [67] Jeffreys, H.: *Theory of Probability*. Oxford: University Press, 1939/1967.
- [68] Laplace, P.S.: *Theorie analytique des probabilités*. Paris: Courcier. Reeditado en Oeuvres Completes de Laplace (1912). Paris: Gauthier-Villars, 1812/1912.
- [69] Novick, M.R.: *Multiparameter Bayesian indifference procedures*. J. Roy. Statist. Soc. B **31**, 29- 64, 1969.
- [70] Zellner, A.: *Maximal data information prior distributions*. New Developments in the Applications of Bayesian Methods, 211-132. Amsterdam: North-Holland, 1977.
- [71] Bernardo, J.M.: *Bioestadística. Una perspectiva Bayesiana*. Vicens-Vives, 1981.
- [72] Johnson N., Kotz S.: *Distributions in Statistics (Continuous Univariate Distributions)*. John Wiley & Sons, 1970.
- [73] Seneta, E.: *Functional Equations and the Galton- Watson Process*. Adv. Appl. Prob. **1**, 1-42, 1969.
- [74] Heyde, C.C.: *Extension of the result of Seneta for the Supercritical Galton- Watson Process*. Ann. Math. Statist. **41**, 739- 742, 1970.