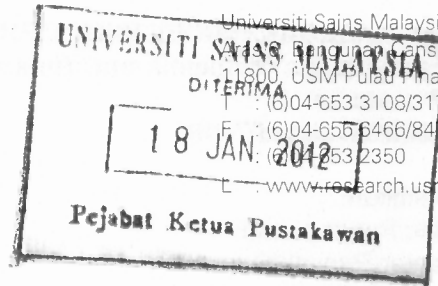




Canselori,

No. Fail : F0333
Tarikh : 2 Disember 2011

Dr. Husna Hasan
Pusat Pengajian Sains Matematik
Universiti Sains Malaysia



Puan,

LAPORAN AKHIR SKIM GERAN PENYELIDIKAN FUNDAMENTAL (FRGS)

Tajuk Projek : **Exceedance Problems and Binary Splitting in Branching Stochastic Processes**

No. Akaun : 203/PMATHS/671191

Dengan hormatnya perkara di atas dirujuk.

2. Terlebih dahulu saya ucapkan ribuan terima kasih di atas satu salinan laporan akhir untuk projek penyelidikan seperti tajuk di atas.

3. Adalah dimaklumkan walaupun projek ini telah selesai, kerjasama Jabatan Bendahari dipohon untuk menguruskan penutupan akaun projek pada selewat-lewatnya **31 Disember 2011**. Tempoh ini bertujuan untuk menyelesaikan semua urusan tuntutan dan bayaran yang telah dibelanjakan di dalam tempoh projek. Walau bagaimanapun, puan dinasihatkan supaya tidak mengeluarkan borang-borang pesanan baru di dalam tempoh ini.

4. Selanjutnya sila ambil perhatian terhadap perkara-perkara berikut sekiranya berkaitan:

- (i) Semua penerbitan harus merakamkan penghargaan kepada **Skim Geran Penyelidikan Fundamental (FRGS)** dan puan dipohon mengemukakan satu salinan ke Pejabat ini.
- (ii) Bahagian Penyelidikan & Inovasi boleh/akan mengagihkan semula peralatan yang telah dibeli menggunakan peruntukan geran ini seandainya terdapat penyelidik lain yang memerlukan peralatan tersebut.

5. Akhir sekali, tahniah di atas usaha dan kejayaan pihak puan dapat menyelesaikan projek ini dengan jayanya.

Sekian, terima kasih.

“BERKHIDMAT UNTUK NEGARA”
‘Memastikan Kelestarian Hari Esok’

Yang menjalankan tugas,

(AMRA OTHMAN)
Penolong Pendaftar
Unit Pengurusan Geran & Kontrak

HAA, HAR, SM

LAPORAN AKHIR SKIM GERAN PENYELIDIKAN FUNDAMENTAL (FRGS)


Tajuk Projek : Exceedance Problems and Binary Splitting in Branching Stochastic Processes

No. Akaun : 203/PMATHS/671191

s.k. Dekan Penyelidikan
Pelantar Sains Fundamental
Pejabat Pelantar Penyelidikan
Universiti Sains Malaysia

Dekan
Pusat Pengajian Sains Matematik
Universiti Sains Malaysia

Timbalan Dekan
(Pengajian Siswazah & Penyelidikan)
Pusat Pengajian Sains Matematik
Universiti Sains Malaysia

 Ketua Pustakawan
Perpustakaan Hamzah Sendut
Universiti Sains Malaysia

Penolong Bendahari Kanan
Unit Kumpulan Wang Penyelidikan
Jabatan Bendahari
Universiti Sains Malaysia

Pegawai Sains
Pelantar Sains Fundamental
Pejabat Pelantar Penyelidikan
Universiti Sains Malaysia

} Disampaikan satu salinan laporan akhir projek untuk simpanan Perpustakaan

} Mohon kerjasama pihak puan untuk menguruskan penutupan akaun projek selewat-lewatnya pada **31 Disember 2011** dan mohon kemukakan satu salinan penyata kewangan terakhir ke Pejabat ini untuk tujuan rekod

Kod Projek :

FRGS/FASA2-
1/2007/(STATISTIK)/(USM)/(203/PMATHS/671191)

FINAL REPORT FUNDAMENTAL RESEARCH GRANT SCHEME (FRGS)

Laporan Akhir Skim Geran Penyelidikan Asas (FRGS) IPT

Pindaan 1/2009

A RESEARCH TITLE : Exceedance Problems and Binary Splitting in Branching Stochastic Processes
Tajuk Penyelidikan

PROJECT LEADER : Husna Binti Hasan
Ketua Projek

PROJECT MEMBERS : 1. Prof. George P. Yanev
(including GRA) 2. Siti Sabariah Bt. Abas & Nurzafirah Bt. A. Khadar
Ahli Projek

PROJECT ACHIEVEMENT (*Prestasi Projek*)

ACHIEVEMENT PERCENTAGE			
Project progress according to milestones achieved up to this period	0 - 50%	51 - 75%	76 - 100%
Percentage			
RESEARCH FINDINGS			
Number of articles/ manuscripts/ books	Indexed Journal	<i>Non-Indexed Journal</i>	
	-	2	
Paper presentations	International	<i>National</i>	
	1	1	
Others (Please specify)			
HUMAN CAPITAL DEVELOPMENT			
Human Capital	Number		Others (Please specify):
	On-going	Graduated	
PhD Student			
Masters Student	1		
Undergraduate Students			
Temporary Research Officer	1		
Temporary Research Assistant	1		
Total	3		

EXPENDITURE (Perbelanjaan)

C	Budget Approved (Peruntukan diluluskan)	: RM 44,200.00
	Amount Spent (Jumlah Perbelanjaan)	: RM 44,026.10
	Balance (Baki)	: <u>RM 173.90</u>
	Percentage of Amount Spent (Peratusan Belanja)	: 99.61 %

ADDITIONAL RESEARCH ACTIVITIES THAT CONTRIBUTE TOWARDS DEVELOPING SOFT AND HARD SKILLS
 (Aktiviti Penyelidikan Sampingan yang menyumbang kepada pembangunan kemahiran insaniah)

D

International		
Activity	Date (Month, Year)	Organizer
(e.g : Course/ Seminar/ Symposium/ Conference/ Workshop/ Site Visit)	1. Jun 2009, Asian Mathematical Conference	1. Universiti Sains Malaysia & Persatuan Sains Matematik Malaysia
	2. Disember 2009, 10th Islamic Countries Conference on Statistical Sciences (ICSS-X)	2. American University in Cairo & Islamic Society of Statistical Sciences
National		
Activity	Date (Month, Year)	Organizer
(e.g : Course/ Seminar/ Symposium/ Conference/ Workshop/ Site Visit)	1. November 2009, SPSS Annual Conference	1. SPSS User 's Association
	2. Disember 2009, Simposium Kebangsaan Sains Matematik ke-17	2. Persatuan Sains Matematik Malaysia & Universiti Putra Malaysia

PROBLEMS / CONSTRAINTS IF ANY (Masalah/ Kekangan sekiranya ada)

E

-

RECOMMENDATION (Cadangan Penambahbaikan)

F

-

RESEARCH ABSTRACT – Not More Than 200 Words(*Abstrak Penyelidikan – Tidak Melebihi 200 patah perkataan*)

G We explore the relationship between branching processes and random sums of indicators. As a tool for obtaining the results, we look into a simple lemma about the distributional properties of a sequence of random variables with members falling in certain random interval. Then, we consider the number of families in Bienayme-Galton-Watson branching processes where size is within a random interval. We obtain limit theorems for subcritical, critical and supercritical processes with and without immigration. In the proofs, we make use of the result about the number of observations in a sample with random sample size falling within a random interval. In addition, the limit theorems for the families and total progeny in the first $(n+1)$ generations for all three cases of criticality of the processes with and without immigration are also obtained.

Date : 17/5/10
Tarikh

Project Leader's Signature: 
Tandatangan Ketua Projek

COMMENTS, IF ANY/ ENDORSEMENT BY RESEARCH MANAGEMENT CENTER (RMC)
(*Komen, sekiranya ada/ Pengesahan oleh Pusat Pengurusan Penyelidikan*)

H
.....
.....

Name:
Nama:

Signature:
Tandatangan:

Date:
Tarikh:

UNIVERSITI SAINS MALAYSIA
 JABATAN BENDAHARI
 KUMPULAN WANG FRGS (203)
 PENYATA PERBELANJAAN PADA 30 APRIL 2010

Tarikh Laporan : 04/05/2010

NAMA PROJEK :

EXCEEDANCE PROBLEMS AN BINARY SPLITTING IN BRANCHING STOCHASTIC PROCESSES

TEMPOH : 2 TAHUN

KETUA PROJEK : DR. HUSNA HASSAN

P.P.SAINS MATEMATIK

AKAUN	PTJ	PROJEK	DONOR	PERUNTUKAN PROJEK	PERBELANJAAN TERKUMPUL SEHINGGA THN LALU	PERUNTUKAN SEMASA	TANGUNGAN SEMASA	BAYARAN SEMASA	BELANJA SEMASA	BAKI PROJEK
111	PMATHS	671191		13,570.00	575.75	12,994.25	0.00	8,618.06	8,618.06	4,376.19
221	PMATHS	671191		8,000.00	15,960.00	-7,960.00	0.00	-140.44	-140.44	-7,819.56
223	PMATHS	671191		650.00	0.00	650.00	0.00	0.00	0.00	650.00
227	PMATHS	671191		2,000.00	4,184.50	-2,184.50	0.00	378.60	378.60	-2,563.10
228	PMATHS	671191		1,000.00	0.00	1,000.00	0.00	300.00	300.00	700.00
229	PMATHS	671191		10,000.00	5,707.98	4,292.02	0.00	1,022.50	1,022.50	3,269.52
335	PMATHS	671191		8,980.00	7,288.00	1,692.00	0.00	0.00	0.00	1,692.00
552	PMATHS	671191		0.00	40.00	-40.00	0.00	91.15	91.15	-131.15
				44,200.00	33,756.23	10,443.77	0.00	10,269.87	10,269.87	173.90

SENARAI JUMLAH JUMLAH KECIL :

110	EMOLUMEN			13,570.00	575.75	12,994.25	0.00	8,618.06	8,618.06	4,376.19
220	PERKHIDMATAN DAN BEKALAI			21,650.00	25,852.48	-4,202.48	0.00	1,560.66	1,560.66	-5,763.14
330	ASET			8,980.00	7,288.00	1,692.00	0.00	0.00	0.00	1,692.00
550	PERBELANJAAN LAIN			0.00	40.00	-40.00	0.00	91.15	91.15	-131.15
				44,200.00	33,756.23	10,443.77	0.00	10,269.87	10,269.87	173.90

On Randomly Bounded Size Families in Branching Processes

G. P. Yanev

Department of Mathematics, University of Texas - Pan American, Edinburg, TX 78539, U.S.A.

Email: gyanev@utpa.edu

H. Hasan

School of Mathematical Sciences, Universiti Sains Malaysia, 11800 Minden, Pulau Pinang, Malaysia.

Email: husna@cs.usm.my

Abstract

We consider the number of families in Bienaymé-Galton-Watson branching processes whose size is within a random interval. We obtain limit theorems for subcritical, critical, and supercritical processes with and without immigration. In the proofs, we make use of a result about the number of observations in a sample with random sample size falling within a random interval.

Keywords: Family size, order statistics, records, branching processes.

2000 Mathematics Subject Classification: 60J80, 62G30.

1 Introduction

Branching processes are stochastic population growth models. They have very diverse applications in both mathematics and other sciences. Central problems in the study of branching processes are criteria for extinction and non-extinction, time to extinction, limit theorems for the generation size and the total progeny. Other characteristics of the structure of the process' genealogy and the branching tree attract attention of the researchers as well. Examples of those include maximum family size, number of pairs of individuals having the same number of descendants in a future generation, distance to the closest common ancestor, existence of an infinite complete N -ary subtree, among others.

In this paper we focus our attention on the families with randomly bounded size in simple branching processes with and without immigration. This characteristic is closely related to the number of large families, exceedances, and reduced branching processes studied by other authors, see [12], [13] and the references therein.

In our analysis we explore the relationship between branching processes and random sums of indicators which has been studied in details by Rahimov starting in [10], see also [11]. More specifically, as a tool for obtaining the results we use a simple lemma about the

distributional properties of a sequence of random variables with members falling in certain random interval. In general, this is related to the theory of exceedance statistics (see [14]), placement statistics and tolerance limits, for more details see [1] and the references therein.

The rest of the paper is organized as follows. In Section 2 we state and prove two lemmas about the number of observations in a sample with random sample size falling in a random interval. We also illustrate the applicability of these lemmas with two examples when the endpoints of the interval are either order statistics or record values. Section 3 deals with simple branching processes without immigration. We consider all three cases of criticality of the processes. In Section 4 we extend the results in Section 3 to branching processes with immigration. In the last section we present results for the families living in the first $n + 1$ generations of both processes with and without immigration.

2 Preliminaries

Let us have three sets of random variables defined on a common probability space as follows.

- A random vector $X = (X_L, X_U)$ where $X_L < X_U$ a.s.
- A triangular array $Y_i(n)$, $i, n = 1, 2, \dots$ of iid variables with a common cdf G ; $Y_i(n)$ are independent from X .
- A sequence of nonnegative integer-valued variables Z_n , $n = 1, 2, \dots$, not necessarily independent; $Y_i(n)$ are independent from Z_n .

Consider the number of observations $S_{Z_n}^X$ in the sample $Y_1(n), \dots, Y_{Z_n}(n)$ falling into the random interval (X_L, X_U) . Let us introduce the following indicator variables.

Definition 2.1. For $i = 1, 2, \dots$ and $n = 1, 2, \dots$, we define

$$\delta_i(n) = \begin{cases} 1 & \text{when } Y_i(n) \in (X_L, X_U) \\ 0 & \text{when } Y_i(n) \notin (X_L, X_U). \end{cases}$$

Then the number of Y s, among the first Z_n members in the n th row, falling in (X_L, X_U) is given for $Z_n > 0$ by

$$S_{Z_n}^X = \sum_{i=1}^{Z_n} \delta_i(n) \quad (n = 1, 2, \dots). \quad (2.1)$$

We set $S_0^X = 0$ and denote

$$\Delta_X = G(X_U) - G(X_L) \quad \text{and} \quad \bar{\Delta}_X = 1 - \Delta_X.$$

Further on we assume that $0 < \Delta_X < 1$. Below we find the exact distribution of $S_{Z_n}^X$ as well as its mean and variance.

Lemma 2.1. For any nonnegative integer n and j

$$P(S_{Z_n}^X = j) = E \left\{ \binom{Z_n}{j} \Delta_X^j \bar{\Delta}_X^{Z_n-j} \right\},$$

$$ES_{Z_n}^X = E \{ Z_n \Delta_X \} \quad \text{and} \quad \text{Var} S_{Z_n}^X = \text{Var} \{ Z_n \Delta_X \} + E \{ Z_n \Delta_X \bar{\Delta}_X \}$$

provided that the moments exist.

Proof. It is clear by (2.1) that for fixed (Z_n, X) the random variable $S_{Z_n}^X$ is binomial. Thus,

$$P(S_{Z_n}^X = j) = E \{ P(S_{Z_n}^X = j | Z_n, X) \} = E \left\{ \binom{Z_n}{j} \Delta_X^j \bar{\Delta}_X^{Z_n-j} \right\},$$

which proves the probability formula. For the expected value of $S_{Z_n}^X$ we have

$$ES_{Z_n}^X = EE \left\{ \sum_{j=0}^{Z_n} j \binom{Z_n}{j} \Delta_X^j \bar{\Delta}_X^{Z_n-j} | Z_n, X \right\} = E \{ Z_n \Delta_X \}.$$

Finally, since

$$\begin{aligned} E (S_{Z_n}^X)^2 &= EE \left\{ \sum_{j=0}^{Z_n} j^2 \binom{Z_n}{j} \Delta_X^j \bar{\Delta}_X^{Z_n-j} | Z_n, X \right\} \\ &= E \{ Z_n(Z_n - 1) \Delta_X^2 \} + E \{ Z_n \Delta_X \}, \end{aligned}$$

we have

$$\begin{aligned} \text{Var} S_{Z_n}^X &= E \{ Z_n(Z_n - 1) \Delta_X^2 \} + E Z_n \Delta_X - (ES_{Z_n}^X)^2 \\ &= \text{Var} \{ Z_n \Delta_X \} + E \{ Z_n \Delta_X \bar{\Delta}_X \}. \end{aligned}$$

□

Corollary 2.1. If Z_n and X are independent, then

$$\text{Var} S_{Z_n}^X = D(EZ_n, \text{Var} Z_n)$$

provided the moments exist, where

$$D(x, y) = xE\Delta_X[1 - xE\Delta_X] + [y + x(x - 1)]E\Delta_X^2. \quad (2.2)$$

To illustrate the applicability of the above results, let us look at two important special cases when X_L and X_U are certain statistics based on a sample from a random variable X_G with cdf $G(x)$, i.e., $Y = X_G$ in distribution. In the two examples below we assume that $G(x)$ is continuous. For related results we refer to [2] and [3].

Example 2.1. Let $X = (X_{r:m}, X_{r+d:m})$ be the random interval with endpoints the r th and $(r + d)$ th order statistics in a sample of size m with $1 \leq r < r + d \leq m$, from a random variable with continuous cdf $G(x)$. It is not difficult to see that

$$G(X_{k:m}) = (G(X_G))_{k:m} = U_{k:m} \quad k = 1, 2, \dots, m,$$

where $U_{k:m}$ is the k th order statistic from Uniform $(0,1)$ distribution. Therefore, Δ_X is the d th spacing from Uniform $(0,1)$ distribution and we have

$$\Delta_X \sim \text{Beta}(d, m - d + 1), \quad E\Delta_X = \frac{d}{m + 1}, \quad E\Delta_X^2 = \frac{d(d + 1)}{(m + 1)(m + 2)}.$$

Assuming independence of Z_n and X , Lemma 2.1 implies

$$ES_{Z_n}^X = \frac{d}{m + 1}EZ_n, \quad E(S_{Z_n}^X)^2 = \frac{d(d + 1)}{(m + 1)(m + 2)}EZ_n^2 + \frac{d(m - d + 1)}{(m + 1)(m + 2)}EZ_n,$$

and

$$\begin{aligned} P(S_{Z_n}^X = j) &= \frac{1}{B(d, m - d + 1)} E \left\{ \binom{Z_n}{j} \int_0^1 x^j (1 - x)^{Z_n - j} x^{d-1} (1 - x)^{m-d} dx \right\} \\ &= E \left\{ \binom{Z_n}{j} \frac{B(j + d, Z_n - j + m - d + 1)}{B(d, m - d + 1)} \right\}, \end{aligned}$$

where $B(x, y)$ is the Beta function.

Example 2.2. Let $X = (R_{r-1}, R_r)$ be the random interval with endpoints the $(r - 1)$ st and r th (upper) records in a sequence of random variables with continuous cdf $G(x)$. It is not difficult to see (e.g. [1, p.240]) that $G(R_{r-1})$ and $G(R_r)$ are the corresponding records from Uniform $(0,1)$ distribution and for $\Delta_X = G(R_r) - G(R_{r-1})$ we have

$$\frac{d}{dx}P(\Delta_X \leq x) = \frac{(-\ln x)^{r-1}}{(r - 1)!}, \quad E\Delta_X^k = (k + 1)^{-r}.$$

Assuming independence of Z_n and X , Lemma 2.1 implies

$$ES_{Z_n}^X = 2^{-r}EZ_n, \quad E(S_{Z_n}^X)^2 = 3^{-r}EZ_n^2 + (2^{-r} - 3^{-r})EZ_n,$$

and

$$\begin{aligned} P(S_{Z_n}^X = j) &= \frac{1}{(r - 1)!} E \left\{ \binom{Z_n}{j} \int_0^1 x^j (1 - x)^{Z_n - j} (-\ln x)^{r-1} dx \right\} \\ &= \frac{1}{(r - 1)!} E \left\{ \binom{Z_n}{j} \int_0^\infty e^{-(j+1)y} (1 - e^{-y})^{Z_n - j} y^{r-1} dy \right\}. \end{aligned}$$

In the rest of this section, we prove a limit result for $S_{Z_n}^X$ as $n \rightarrow \infty$. Note that more general results for random sums of indicators can be found in [10] and [12].

Lemma 2.2. (1). Let $c_n \rightarrow \infty$ and $\lim_{n \rightarrow \infty} P(Z_n/c_n \leq x | Z_n > 0) = P(Z \leq x)$. Then for $x \in R$

$$\lim_{n \rightarrow \infty} P\left(\frac{S_{Z_n}^X}{c_n} \leq x | Z_n > 0\right) = P(Z \Delta_X \leq x).$$

(2). Assume $\lim_{n \rightarrow \infty} P(Z_n \leq x | Z_n > 0) = P(Z \leq x)$. Then

$$\lim_{n \rightarrow \infty} P(S_{Z_n}^X = j | Z_n > 0) = r_j, \quad (2.3)$$

where $\{r_j\}$ is a probability distribution with generating function

$$\sum_{j=0}^{\infty} r_j u^j = E[1 - \Delta_X + \Delta_X u]^Z.$$

Proof. (1). Considering the Laplace transform of $S_{Z_n}^X/c_n$, we obtain

$$\begin{aligned} E\left\{e^{-\lambda S_{Z_n}^X/c_n}\right\} &= E\left\{\sum_{k=0}^{\infty} \sum_{j=0}^k e^{-\lambda j/c_n} \binom{k}{j} \Delta_X^j \bar{\Delta}_X^{k-j} P(Z_n = k | B_n)\right\} \\ &= E\left\{\sum_{k=0}^{\infty} (1 - \Delta_X + \Delta_X e^{-\lambda/c_n})^k P(Z_n = k | B_n)\right\} \\ &= E\left\{\sum_{k=0}^{\infty} (1 - \Delta_X + \Delta_X e^{-\lambda/c_n})^{c_n(k/c_n)} P\left(\frac{Z_n}{c_n} = \frac{k}{c_n} | B_n\right)\right\} \\ &\rightarrow E\left\{\int_0^{\infty} e^{-\lambda \Delta_X y} dP(y \leq y)\right\} = E\left\{e^{-\lambda Z \Delta_X}\right\}, \end{aligned}$$

where we have used that $\lim_{x \rightarrow \infty} (1 - a + ae^{b/x})^x = e^{ab}$. Indeed, as $x \rightarrow \infty$

$$\begin{aligned} x \log(1 - a + ae^{b/x}) &= x \log(1 + ab/x + o(1/x)) \\ &= x(ab/x + o(1/x)) \\ &\rightarrow ab. \end{aligned}$$

(2). Using the probability generation function of $S_{Z_n}^X$, we obtain

$$\begin{aligned} E\left\{u^{S_{Z_n}^X}\right\} &= E\left\{\sum_{k=0}^{\infty} \sum_{j=0}^k u^j \binom{k}{j} \Delta_X^j \bar{\Delta}_X^{k-j} P(Z_n = k | Z_n > 0)\right\} \\ &= E\left\{\sum_{k=0}^{\infty} (1 - \Delta_X + \Delta_X u)^k P(Z_n = k | Z_n > 0)\right\} \\ &\rightarrow E\left\{(1 - \Delta_X + \Delta_X u)^Z\right\}. \end{aligned}$$

3 Simple Branching Processes

The Bienamé-Galton-Watson (BGW) branching process is defined by the recurrence

$$Z_{n+1} = \sum_{i=1}^{Z_n} Y_i(n), \quad (n = 0, 1, 2, \dots), \quad Z_0 \equiv 1, \quad (3.1)$$

where the offspring variables $Y_i(n)$ are independent with respect to both indexes i and n and have a common cdf $G(x)$. The random variable $S_{Z_n}^X$ from (2.1) can be interpreted as the number of families living in the $(n+1)$ st generation (i.e., the parents are from the n th generation), who have family size within the random interval (X_L, X_U) .

3.1 Critical processes

Let us consider the critical branching process, i.e., assume $EY_i(n) = 1$ with offspring pgf $f(s)$ given by

$$f(s) = s + (1-s)^{1+\alpha}L(1/(1-s)) \quad (3.2)$$

for $0 < \alpha \leq 1$, where $0 \leq s \leq 1$ and $L(x)$ is a slowly varying at infinity function (s.v.f). It is well-known (see e.g. [5, Th. 8.12.3]) that (3.2) is a necessary and sufficient condition for

$$\lim_{n \rightarrow \infty} P(Q_n Z_n > y \mid Z_n > 0) = P(Z > y), \quad y \geq 0, \quad (3.3)$$

where $Q_n = P(Z_n > 0)$ and the limit Z has Laplace transform

$$\varphi(u) = Ee^{-uZ} = 1 - (1 + u^{-\alpha})^{-1/\alpha} \quad u > 0. \quad (3.4)$$

Moreover, if (3.2) holds then $Q_n = n^{-1/\alpha}M(n)$ and $\lim_{x \rightarrow \infty} M^\alpha(x)L(x^{1/\alpha}/M(x)) = 1/\alpha$, where $M(n)$ and $L(x)$ are s.v.f.'s.

Proposition 3.1. *Assume (3.2). Then*

$$\lim_{n \rightarrow \infty} P(Q_n S_{Z_n}^X \leq z \mid Z_n > 0) = P(Z \Delta_X \leq z), \quad (3.5)$$

where Z has Laplace transform (3.4). Furthermore, if Z_n and X are independent, then

$$ES_{Z_n}^X = E\Delta_X$$

and if $\text{Var}Y_i(n) = \sigma^2 < \infty$ then

$$\text{Var}S_{Z_n}^X = E\Delta_X E\bar{\Delta}_X + n\sigma^2 E\Delta_X^2.$$

Proof. The limit (3.5) follows from Lemma 2.2, taking into account (3.3). To prove the moment results, recall that Z_n and X are independent and for the critical process $EZ_n = 1$ and $\text{Var}Z_n = n\sigma^2$. Thus, Lemma 2.1 implies

$$ES_{Z_n}^X = EZ_n \Delta_X = EZ_n E\Delta_X = E\Delta_X$$

and the formula for $Var S_{Z_n}^X$ simplifies to

$$\begin{aligned} Var S_{Z_n}^X &= E\Delta_X[1 - \Delta_X] + EZ_n^2 E\Delta_X^2 - (EZ_n)^2 (E\Delta_X)^2 \\ &= E\Delta_X(1 - E\Delta_X) + n\sigma^2 E\Delta_X^2. \end{aligned}$$

□

Example 3.1. Let $\alpha = 1$ and $1 - G(x) \sim x^{-2} \log x$. By Prop. A(ii) in [4] it follows that $L(x) \sim \int_1^x (\log u)/u du = (\log x)^2/2$. Since $\lim_{x \rightarrow \infty} M(x)L(x/M(x)) = 1$, we have $1/M(n) \sim (\log n)^2/2$. Now, $1/Q_n \sim n(\log n)^2/2$ and for $x \geq 0$,

$$\lim_{n \rightarrow \infty} P\left(\frac{2S_{Z_n}^X}{n(\log n)^2} \leq z \mid Z_n > 0\right) = P\left(Z \left(\frac{\log X_L}{X_L^2} - \frac{\log X_U}{X_U^2}\right) \leq z\right),$$

where Z is a unit exponential random variable.

3.2 Non-critical processes

It is known (see e.g. [5, Th. 8.12.5]) that in the supercritical case, i.e., when $m = EY_i(n) > 1$, there exists a sequence of constants $\{C_n\}$, with $\lim_{n \rightarrow \infty} C_n = \infty$ such that $\{Z_n/C_n\}$ converges almost surely to a non-degenerate limit W . The Laplace transform $\psi(u) = E \exp\{-uW\}$, $u > 0$ of the limiting variable is the unique (up to a scale factor) solution of the functional equation $\psi(u) = f(\psi(u/m))$. The constants C_n take on the form $C_n = m^n/L_2(m^n)$, where $L_2(x) = \int_0^x P(W > y)dy$ is a s.v.f..

In the subcritical case, i.e., when $0 < m < 1$, it is known (see e.g. [6, Th. 2.6.2]) that

$$\lim_{n \rightarrow \infty} P(Z_n = j \mid Z_n > 0) = p_j, \quad j = 0, 1, \dots, \tag{3.6}$$

where $\{p_j\}$ is a probability distribution whose generating function $\gamma(u)$ is the unique generation-function solution of the equation $\gamma(f(u)) = m\gamma(u) + 1 - m$, $\gamma(0) = 0$, where $f(u)$ is the offspring pgf.

Proposition 3.2. (1). Assume $1 < m < \infty$. Then

$$\lim_{n \rightarrow \infty} P\left(\frac{S_{Z_n}^X}{C_n} \leq x\right) = P(W\Delta_X \leq x), \tag{3.7}$$

where W has the Laplace transform $\psi(u)$ given above.

(2). If $0 < m < 1$, then

$$\lim_{n \rightarrow \infty} P(S_{Z_n}^X = j \mid Z_n > 0) = r_j, \tag{3.8}$$

where $\{r_j\}$ is a probability distribution with $\sum_{j=0}^{\infty} r_j u^j = E\{\gamma(1 - \Delta_X + \Delta_X u)\}$.

(3). Assume that Z_n and X are independent. If $m \neq 1$ and $Var Y_i(n) = \sigma^2 < \infty$, then

$$ES_{Z_n}^X = EZ_n E\Delta_X \quad \text{and} \quad Var S_{Z_n}^X = D(EZ_n, Var Z_n),$$

where $D(x, y)$ is given in (2.2), $EZ_n = m^n$ and $Var Z_n = \sigma^2 m^{n-1}(m^n - 1)/(m - 1)$.

Proof. Parts (1) and (2) follow directly from Lemma 2.2. For the moments, under the assumption of independence, we have

$$ES_{Z_n}^X = EZ_n E\Delta_X = m^n E\Delta_X$$

and

$$\begin{aligned} \text{Var}S_{Z_n}^X &= EZ_n E\Delta_X [1 - \Delta_X] + EZ_n^2 E\Delta_X^2 - (EZ_n)^2 [E\Delta_X]^2 \\ &= EZ_n E\Delta_X [1 - EZ_n E\Delta_X] + [\text{Var}Z_n + EZ_n(EZ_n - 1)] E\Delta_X^2. \end{aligned}$$

□

Example 3.2. Consider a branching process with zero-modified geometric offspring distribution, i.e., $Y_i(n)$ in (3.1) has a pmf given for $0 < a < 1$ and $0 < p < 1$ by

$$g(y) = \begin{cases} 1 - a & y = 0, \\ ap(1 - p)^{y-1} & y = 1, 2, \dots \end{cases}$$

Thus, $EY_i(n) = a/p$. Let (X_L, X_U) be the random interval (X_{\min}, X_{\max}) with endpoints the minimum and the maximum of two independent geometric variables each distributed as $Y_i(n)$. Then, one can see that $G(X_{\min})$ has pmf

$$h_{G(\min)}(k) = \begin{cases} 1 - a^2 & k = 0, \\ \frac{a^2 p(2 - p)}{(1 - p)^2} (1 - p)^{2k} & k = 1, 2, \dots \end{cases}$$

and $G(X_{\max})$ has pmf

$$h_{G(\max)}(k) = \begin{cases} (1 - a)^2 & k = 0, \\ 2ap(1 - p)^{k-1} - a^2 p(2 - p)(1 - p)^{2k-2} & k = 1, 2, \dots \end{cases}$$

Hence,

$$\begin{aligned} E\Delta_X &= EG(X_{\max}) - EG(X_{\min}) \\ &= \left(1 - a + \frac{2a^2}{2 - p} - \frac{a^3}{3 - 3p + p^2}\right) - \left(1 - a + \frac{a^3}{3 - 3p + p^2}\right) \\ &= 2a^2 \left(\frac{1}{1 + q} - \frac{a}{1 + q + q^2}\right), \end{aligned}$$

where $q = 1 - p$. Notice that Propositions 3.1 and 3.2 imply

$$ES_{Z_n}^X = \begin{cases} 2(1 - q)^2 \left(\frac{1}{1 + q} - \frac{1 - q}{1 + q + q^2}\right) & \text{when } a/p = 1 \text{ (critical case),} \\ 2a^2 \left(\frac{a}{p}\right)^n \left(\frac{1}{1 + q} - \frac{a}{1 + q + q^2}\right) & \text{when } a/p \neq 1 \text{ (noncritical case).} \end{cases}$$

4 Processes with Immigration

Assume at the time of birth of the n th generation, I_n individuals immigrate into the population. Then the BGW process with immigration is defined by

$$Z_{n+1}^* = \sum_{i=1}^{Z_n^*} Y_i(n) + I_{n+1}, \quad Z_0^* = 1 \quad (n = 0, 1, \dots)$$

where I_1, I_2, \dots are independent and identically distributed with generating function $h(u)$ and I_i 's are independent of the offspring variables $Y_i(n)$ which have common pgf $f(u)$. The pgf of Z_{n+1}^* is given by

$$g_{n+1}(u) = h(u)g_n[f(u)]. \quad (4.1)$$

The random variable $S_{Z_n^*}^X$ from (2.1) can be interpreted as the number of families (including families with immigrated parents) living in the $(n+1)$ st generation who have family size within the random interval (X_L, X_U) .

4.1 Critical processes

In the critical process with immigration, the same normalization as for that without immigration leads to the following limit distribution (see [6]). If

$$f'(1) = 1, \quad f''(1) = \sigma^2 < \infty \quad \text{and} \quad 0 < h'(1) = \mu < \infty, \quad (4.2)$$

then $2Z_n^*/(\sigma^2 n)$ converges in distribution to a random variable with the gamma density function

$$w(x) = \frac{x^{(2\mu/\sigma^2)-1} e^{-x}}{\Gamma(2\mu/\sigma^2)}, \quad x \in (0, \infty). \quad (4.3)$$

Proposition 4.1. Assume (4.2). Then

$$\lim_{n \rightarrow \infty} P\left(\frac{2S_{Z_n^*}^X}{\sigma^2 n} \leq z\right) = P(Z\Delta_X \leq z), \quad (4.4)$$

where Z has density (4.3). Denote $\tau^2 = \text{Var} I_n < \infty$ and assume that Z_n^* and X are independent. Then

$$ES_{Z_n^*}^X = EZ_n^* E\Delta_X \quad \text{and} \quad \text{Var} S_{Z_n^*}^X = D(EZ_n^*, \text{Var} Z_n^*),$$

where $D(x, y)$ is given in (2.2), $EZ_n^* = \mu n + 1$ and $\text{Var} Z_n^* = n(\sigma^2 + \tau^2) + n(n-1)\mu\sigma^2/2$.

Proof. The limit in (4.4) follows from Lemma 2.2 taking into account (4.3). To calculate the moments, recall that Z_n^* and X are independent. Differentiating (4.1) at $u = 1$, we obtain $E(Z_n^*) = \mu n + 1$. Thus, from Lemma 2.1,

$$ES_{Z_n^*}^X = E[Z_n^* \Delta_X] = EZ_n^* E\Delta_X = (\mu n + 1)E\Delta_X$$

and

$$\begin{aligned} \text{Var} Z_n^* &= \text{Var} Z_{n-1}^* + \mu\sigma^2(n-1) + \sigma^2 + \tau^2 \\ &= n(\sigma^2 + \tau^2) + \frac{n(n-1)}{2}\mu\sigma^2 \end{aligned}$$

and as in Proposition 3.2, one can obtain the formula for $\text{Var} S_{Z_n^*}^X$. \square

Example 4.1. Consider the pgf's $f(u) = u + (a/\delta)(1-u)^\delta$, and $h(u) = \exp\{c(u-1)\}$ where $1 < \delta < 2$, $0 < a < 1$ and $0 < c < \infty$. Letting $a \rightarrow 0$ and $c \rightarrow 0$ such that $a/c = k$, we have in the limit,

$$g(u) = \exp\{-\alpha(1-u)^\beta\} \quad (\alpha > 0, 0 < \beta < 1), \quad (4.5)$$

where $\alpha = \delta/k(2-\delta)$ and $\beta = 2-\delta$. The pgf (4.5) is that of the Poisson distribution with mean α , compounded by the distribution with pgf $1 - (1-u)^\beta$. Applying Proposition 4.1, we see that $2S_{Z_n^*}^X/(\sigma^2 n)$ has a limit $Z\Delta_X$ with pgf

$$E s^{Z\Delta_X} = E \exp\{-\alpha(1-s^{\Delta_X})^\beta\} = E g(s^{\Delta_X}).$$

4.2 Non-critical processes

In the supercritical case, when $1 < m < \infty$, there exists a sequence of constants C_n such that $\{Z_n^*/C_n\}$ converges with probability one to a random variable V (see [6]). Furthermore, if $E(\log I_1)^+ < \infty$, then $P(V < \infty) = 1$ and V has an absolutely continuous distribution on $(0, \infty)$; if $E(\log I_1)^+ = \infty$, then $P(V < \infty) = 0$. It is known (see [6]) that in the subcritical case, i.e., when $0 < m < 1$, if $0 < h'(1) = \mu < \infty$, then Z_n^* has a proper limit distribution

$$\lim_{n \rightarrow \infty} P(Z_n^* = k) = \rho_k, \quad k = 0, 1, 2, \dots$$

where $\{\rho_k, k \geq 0\}$ is a probability distribution with pgf

$$g(u) = h(u) \prod_{n=1}^{\infty} h(f_n(u)),$$

where $f_n(u)$ is the n -fold iteration of $f(u)$.

Proposition 4.2. (1). Assume $1 < m < \infty$. Then

$$\lim_{n \rightarrow \infty} P\left(\frac{S_{Z_n^*}^X}{C_n} \leq x\right) = P(V\Delta_X \leq x),$$

where V is the limiting variable for the supercritical case given above.

(2). If $0 < m < 1$ and $0 < \mu < \infty$, then

$$\lim_{n \rightarrow \infty} P(S_{Z_n^*}^X = j) = r_j,$$

where r_j is a probability distribution with pgf $E[g(1 - \Delta_X + \Delta_X u)]$.

(3). Assume that Z_n^* and X are independent. If $m \neq 1$, then

$$ES_{Z_n^*}^X = EZ_n^* E\Delta_X \quad \text{and} \quad \text{Var}S_{Z_n^*}^X = D(EZ_n^*, \text{Var}Z_n^*),$$

where $D(x, y)$ is given in (2.2),

$$EZ_n^* = \mu \frac{1 - m^n}{1 - m} + m^n, \quad (4.6)$$

$$\text{Var}Z_n^* = m^{2(n-1)}(\sigma^2 + \tau^2) + (\tau^2 - \mu + \mu^2) \frac{1 - m^{2(n-1)}}{1 - m^2} + \sum_{i=0}^{n-2} m^{2i} S_{n-i}, \quad (4.7)$$

where $\tau^2 = \text{Var}I_n < \infty$ and

$$S_n = EZ_{n-1}^*(m^2 EZ_{n-1}^* + \sigma^2 - m + 2\mu m) + (1 - EZ_n^*)EZ_n^*.$$

Proof. Parts (1) and (2) follow from Lemma 2.2. Differentiating (4.1) at $u = 1$ and iterating, we obtain (4.6) for $m \neq 1$. To find $\text{Var}S_{Z_n^*}^X$, we observe from (4.1) that

$$\text{Var}Z_n^* = m^2 \text{Var}Z_{n-1}^* + (\tau^2 - \mu + \mu^2) + S_n.$$

Now, after iterating, we obtain (4.7). □

Example 4.2. Consider the case $f(u) = 1/(m + 1 - mu)$ and $h(u) = [f(u)]^\nu, \nu > 0$. If $m < 1$ and $EI_n \log I_n < \infty$, then (see [7]) the limiting distribution is negative binomial with pgf

$$g(u) = \left(\frac{1 - m}{1 - mu} \right)^\nu.$$

Therefore, by Proposition 4.2(2), the limiting distribution of $S_{Z_n^*}^X$ has pgf

$$\sum_{j=0}^{\infty} r_j u^j = E \left(\frac{1 - m}{1 - m(1 - \Delta_X + \Delta_X u)} \right)^\nu.$$

5 Families living in the first $(n + 1)$ generations

Let T_n be the total number of individuals who have been present up to and including the n th generation. In this section, we study the number $S_{T_n}^X$ of families that have lived up to and including the $(n + 1)$ st generation and whose family size is within the random interval (X_L, X_U) .

5.1 Processes without immigration

Recall the following limit theorem for the total progeny of the critical process obtained by Pakes [8]. If $m = 1$ and $\sigma^2 < \infty$, then

$$\lim_{n \rightarrow \infty} P\{T_n \leq vn^2 | X_n > 0\} = F(v) \quad (0 \leq v < \infty),$$

where $F(v)$ is absolutely continuous and

$$\int_0^\infty e^{-\lambda v} dF(v) = \sqrt{(2\sigma^2\lambda)} \operatorname{cosech}(\sqrt{(2\sigma^2\lambda)}), \quad (0 \leq \lambda < \infty). \quad (5.1)$$

Proposition 5.1. *If $m = 1$ and $\sigma^2 < \infty$, then*

$$\lim_{n \rightarrow \infty} P\left(\frac{S_{T_n}^X}{n^2} \leq z | Z_n > 0\right) = P(Z\Delta_X \leq z),$$

where Z has the Laplace transform (5.1). Furthermore, if T_n and X are independent, then

$$ES_{T_n}^X = (n+1)E\Delta_X \quad \text{and} \quad \operatorname{Var} S_{T_n}^X = D(ET_n, \operatorname{Var} T_n),$$

where $D(x, y)$ is given in (2.2), $ET_n = n+1$ and $\operatorname{Var} T_n = (2n+1)(n+1)n\sigma^2/6$.

Proof. The limit follows from Lemma 2.2, taking into account the limit theorem above. The formulas for the moments follow from Lemma 2.1 and the fact that (see [8]) $ET_n = n+1$ and $\operatorname{Var} T_n = (2n+1)(n+1)n\sigma^2/6$. \square

It is known that in the subcritical case (see [8]) with $\sigma^2 < \infty$, the sequence $\{T_n/n\}$ given $Z_n > 0$ converges in probability to $C+1$ where

$$C = \frac{f''(1)}{m(1-m)}. \quad (5.2)$$

In the supercritical process, there exists a sequence of positive constants $\{C_n, n \geq 1\}$ with $C_n \rightarrow \infty$ and $C_{n+1}/C_n \rightarrow m$ as $n \rightarrow \infty$ such that

$$\begin{aligned} \lim_{n \rightarrow \infty} P\{T_n \leq zC_n | Z_n > 0\} &= \lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} \{T_n \leq zC_n | Z_{n+k} > 0\} \\ &= P\left\{W \leq \frac{z(m-1)}{m} | W > 0\right\} \end{aligned} \quad (5.3)$$

for $k = 0, 1, \dots$ and $0 \leq z < \infty$, where W is a non-degenerate random variable having a continuous distribution on the set of positive real numbers. It is known that the Laplace transform $\psi(\lambda) = E(e^{-\lambda W})$ satisfies the equation $\psi(m\lambda) = f[\psi(\lambda)]$.

Proposition 5.2. (1). *If $m < 1$ and $\sigma^2 < \infty$, then*

$$\lim_{n \rightarrow \infty} P\left(\frac{S_{T_n}^X}{n} \leq z | Z_n > 0\right) = P((C+1)\Delta_X \leq z),$$

where C is from (5.2).

(2). Assume $1 < m < \infty$ and $\sigma^2 < \infty$, then

$$\lim_{n \rightarrow \infty} P\left(\frac{S_{T_n}^X}{C_n} \leq z | Z_n > 0\right) = P\left(\frac{m}{m-1} W \Delta_X \leq z | W > 0\right),$$

where W is as in (5.3).

(3). Assume that T_n and X are independent. If $m \neq 1$ and $f''(1) < \infty$, then

$$ES_{T_n}^X = ET_n E \Delta_X \quad \text{and} \quad \text{Var} S_{T_n}^X = D(ET_n, \text{Var} T_n),$$

where $D(x, y)$ is given in (2.2), $ET_n = (1 - m^{n+1}) / (1 - m)$ and

$$\text{Var} T_n = \frac{\theta^2}{(1 - m)^2} \left(\frac{1 - m^{2n+1}}{1 - m} - (2n + 1)m^n \right),$$

where $\theta^2 = f''(1) + m(1 - m)$.

Proof. Parts (1) and (2) follow from Lemma 2.2, taking into account the limit theorems for subcritical and supercritical processes. To calculate the moments, we refer to Lemma 2.1 and the expressions for ET_n and $\text{Var} T_n$ given in [8]. □

5.2 Families and total progeny in processes with immigration

In this section, let T_n^* be the total number of individuals as before but allowing the immigration into the process. Assuming $Z_0 \equiv 1$, it is known (see [9]) that

$$E(u^{T_n^*}) = g_n^T(u) \prod_{k=0}^{n-1} h(g_k^T(u)), \tag{5.4}$$

where $g_n^T(u)$ is the pgf of the total number of individuals up to the n th generation of a process without immigration. Let $m = 1$ and $\sigma^2 < \infty$. It is proved in [9] that $T_n^*/n^2 \rightarrow W$ in distribution, where W has the Laplace transform

$$E(e^{-\lambda W}) = [\text{sech} \sqrt{\sigma^2 \lambda / 2}]^{2\mu / \sigma^2}. \tag{5.5}$$

Proposition 5.3. *If $m = 1$ and $\sigma^2 < \infty$, then*

$$\lim_{n \rightarrow \infty} P\left(\frac{S_{T_n^*}^X}{n^2} \leq z\right) = P(W \Delta_X \leq z),$$

where W has the Laplace transform (5.5). Furthermore, if T_n^* and X are independent, then

$$ES_{T_n^*}^X = ET_n^* E \Delta_X \quad \text{and} \quad \text{Var} S_{T_n^*}^X = D(ET_n^*, \text{Var} T_n^*),$$

where $D(x, y)$ is given in (2.2), $ET_n^* = (n + 1) + \mu n(n + 1) / 2$, and for $\tau^2 = \text{Var} I_n < \infty$

$$\text{Var} T_n^* = \text{Var}(T_n) + \frac{\mu n(n + 1)}{2} + \tau^2 \sum_{k=0}^{n-1} (k + 1)^2 + \mu \sum_{k=0}^{n-1} (k + 1) \left(\frac{(2k + 1)k\sigma^2}{6} - 1 \right). \tag{5.6}$$

Proof. The limit result follows from Lemma 2.2. One can find the first moment using the fact (see [9]), that $ET_n^* = (n+1) + \mu n(n+1)/2$. After logarithmic differentiation of (5.4), we arrive at

$$\begin{aligned} \text{Var}T_n^* &= \left((n+1) + \frac{\mu n(n+1)}{2} \right)^2 + \text{Var}(T_n) - (n+1) + (\tau^2 - \mu) \sum_{k=0}^{n-1} (k+1)^2 \\ &+ \mu \sum_{k=0}^{n-1} \left(\frac{(2k+1)(k+1)k\sigma^2}{6} - (k+1) + (k+1)^2 \right) \\ &+ (n+1) + \frac{\mu n(n+1)}{2} - \left((n+1) + \frac{\mu n(n+1)}{2} \right)^2, \end{aligned}$$

which in turn leads to (5.6) and the expression for $\text{Var}S_{T_n^*}^X$. Note that, if there is no immigration (5.6) reduces to $\text{Var}T_n$. \square

If $m < 1$ and $\mu < \infty$, then it is known that $T_n^*/n \rightarrow \mu/(1-m)$ a.s. If $m > 1$ and $\mu < \infty$, then

$$\frac{T_n^*}{m^n} \rightarrow \frac{mU}{m-1} \quad \text{a.s.} \quad (5.7)$$

where U has a finite mean and either $P(U=0) = 0$ or $P(U=0) = 1$ (see [7]). In the former case U has a continuous distribution concentrated on $(0, \infty)$ and this occurs iff $EY_n \log Y_n < \infty$.

Proposition 5.4. (1). If $m < 1$ and $\mu < \infty$, then

$$\lim_{n \rightarrow \infty} P\left(\frac{S_{T_n^*}^X}{n} \leq z\right) = P\left(\frac{\mu \Delta_X}{1-m} \leq z\right).$$

(2). If $1 < m < \infty$ and $\mu < \infty$, then

$$\lim_{n \rightarrow \infty} P\left(\frac{S_{T_n^*}^X}{m^n} \leq z\right) = P\left(\frac{mU \Delta_X}{m-1} \leq z\right),$$

where U is as in (5.7).

(3). Assume that T_n^* and X are independent. If $m \neq 1$ and $\tau^2 = \text{Var}I_i < \infty$, then

$$ES_{T_n^*}^X = ET_n^* E\Delta_X \quad \text{and} \quad \text{Var}S_{T_n^*}^X = D(ET_n^*, \text{Var}T_n^*),$$

where $D(x, y)$ is given in (2.2),

$$ET_n^* = \frac{\mu}{1-m} \left(n - \frac{m(1-m^n)}{1-m} \right) + \frac{1-m^{n+1}}{1-m} \quad (5.8)$$

and

$$\text{Var}T_n^* = \text{Var}T_n + \mu \sum_{k=0}^{n-1} [\text{Var}T_k - ET_k - \mu ET_k] + (\tau^2 + \mu^2) \sum_{k=0}^{n-1} (ET_k)^2.$$

Proof. Parts (1) and (2) follow from Lemma 2.2, taking into account the limit theorems above. Formula (5.8) is given in [7]. Using (5.4), it is not difficult to obtain the formulas for $VarT_n^*$ and $VarS_{T_n^*}^X$. \square

Example 5.1. Consider the process from Example 4.2 where $f(u) = 1/(m+1-mu)$ and $h(u) = [f(u)]^\nu$, $\nu > 0$. Let $m > 1$. The limit law for T_n^*/m^n is Gamma with parameters $(\alpha - 1/\alpha)^2$ and ν (see [7]). Applying Proposition 5.4(2), we see that $S_{T_n^*}^X$ has a limit $mU\Delta_X/(m-1)$ with Laplace transform

$$E\left(e^{-\lambda mU\Delta_X/(m-1)}\right) = \left[\frac{m-1}{m-1 + \lambda m^2 \Delta_X/(m-1)}\right]^\nu.$$

Acknowledgment

We thank the referee for the valuable comments which led to a significant improvement in the presentation of the results.

References

- [1] M. Ahsanullah and V. B. Nevzorov, (2001). *Ordered Random Variables*, Nova Sci. Publ., Nuntington, New York.
- [2] I. G. Bairamov, (1997). Some distribution free properties of statistics based on record values and characterizations of the distributions trough a record, *J. Appl. Statist. Sci.* **5**, 17–25.
- [3] I. G. Bairamov and S. N. Eryilmaz, (2000). Distributional properties of statistics based on minimal spacing and record exceedance statistics, *J. Statist. Planning Infer.* **90**, 21–33.
- [4] N. H. Bingham and R. A. Doney, (1974). Asymptotic properties of super-critical branching processes I: Galton-Watson process, *Adv. Appl. Probab.*, **6**, 711–731.
- [5] N. H. Bingham, C. M. Goldie and J. L. Teugels, (1987). *Regular Variation*, Cambridge University Press, Encyclopedia of Mathematics and Its Applications 27.
- [6] P. Jagers, (1975) *Branching Processes with Biological Applications*, John Wiley, London.
- [7] A. G. Pakes, (1971a). Branching processes with immigration, *J. Appl. Prob.* **8**, 32–42.
- [8] A. G. Pakes, (1971b). Some limit theorems for the total progeny of a branching processes, *Adv. Appl. Prob.* **3**, 176–192.
- [9] A. G. Pakes, (1972). Further results on the critical Galton-Watson process with immigration, *J. Aust. Math. Soc.* **13**, 277–290.
- [10] I. Rahimov, (1987). A limit theorem for random sums of dependent indicators and its applications in the theory of branching processes, *Theory Probab. Appl.* **32**, 290–298.
- [11] I. Rahimov, (1995). *Random Sums and Branching Stochastic Processes*, Lecture Notes in Statistics 96, Springer, New York.

- [12] I. Rahimov, (2003). Random sums of independent indicators and generalized reduced processes, *Stoch. Analysis Appl.* **21**, 205–221.
- [13] I. Rahimov, (2008). Large families and exceedances in branching processes. In: Eds. Ahsanullah, M. and Yanev, G.P., *Records and Branching Processes*, NOVA Sci. Publ., Nuntington, NY, in press.
- [14] J. Wesolowski and M. Ahsanullah, (1998). Distributional properties of exceedance statistics, *Ann. Inst. Statist. Math.* **50**, 543–565.

Rekabentuk Ukuran Ulangan dan Proses Bercabang Dalam Biologi

Husna Binti Hasan

*Pusat Pengajian Sains Matematik,
Universiti Sains Malaysia, 11800 Minden, Penang.
E-mel: husna@cs.usm.my*

ABSTRAK

Tercetusnya konsep Biologi Bermatematik boleh melahirkan usaha kerjasama antara penyelidik bidang sains hayat, matematik dan juga penyelidik dari bidang sains lain yang berkaitan. Data yang terkumpul dari hasil penyelidikan kajian kes sains hayat perlu dianalisa dan ditafsir dengan betul supaya kesimpulannya adalah benar bagi semua kes secaman. Kewujudan pelbagai pakej statistik yang boleh mengeluarkan keputusan analisis dengan segera telah membuka peluang kepada pengguna yang mempunyai pengetahuan terhad dalam bidang statistik khususnya analisis data dan berkemungkinan menghasilkan kesimpulan yang kurang bermakna. Kertas ini membincangkan bagaimana penyelidik sains hayat dan statistik boleh bekerjasama dalam menjayakan analisis data yang melibatkan rekabentuk ukuran ulangan supaya kesimpulan yang terhasil adalah lebih bermakna, munasabah dan realistik. Masalah yang dihadapi dalam proses bercabang yang merupakan satu bidang statistik tulen juga dibangkitkan.

PENGENALAN

Matematik adalah sesuatu yang menggerunkan bagi sesetengah pelajar sains hayat (atau Biologi). Apatah lagi statistik yang merlukan latar belakang matematik untuk memahaminya. Kedua-dua subjek ini adalah antara jurang dalam pendidikan matematik bagi pelajar yang berorientasikan biologi. Walaupun terdapat kursus asas yang boleh diikuti, namun kursus-kursus statistik yang ditawarkan oleh Pusat Pengajian Sains Matematik (atau Statistik) kerap kali tidak memenuhi kehendak pelajar atau bakal penyelidik biologi kerana dikatakan terlalu *theoretical*. Pelajar pula perlu mengaplikasikan kaedah-kaedah statistik ke atas masalah atau kajian kes yang mereka minati.

Mengikut takrif yang diberikan dalam [11],

'Mathematical biology or biomathematics is an interdisciplinary field of academic study which aims at modeling natural, biological processes using mathematical techniques and tools. It has both practical and theoretical applications in biological research.'

Kaedah statistik merupakan salah satu teknik atau alat matematik yang digunakan dalam penyelidikan sains. Tidak dapat disangkal bahawa sesuatu teori menjadi lebih jelas apabila diaplikasikan kepada masalah tertentu. Bayangkan dari suatu teorem matematik yang terdiri dari hanya persamaan atau pernyataan matematik dapat dihuraikan dalam bentuk masalah harian atau kajian kes semasa.

Kebelakangan ini, aplikasi matematik dan statistik dalam bidang biologi timbul dengan kadar yang memberangsangkan. Antara sebabnya ialah lahirnya set-set data bermaklumat misalnya data *genomics* yang susah untuk difahami tanpa alat analitik. Perkembangan dan penemuan terbaru alat matematik dapat membantu penyelidik memahami konsep-konsep biologi yang kompleks. Teknologi perkomputeran yang semakin canggih dan cekap sekarang ini mampu menjalankan pengiraan dan simulasi yang sebelumnya agak mustahil. Pada masa yang sama, bidang-bidang baru dalam teori kebarangkalian dan statistik juga berkembang untuk memenuhi keperluan bidang biologi.

Kaedah statistik dan peranannya

Seerti yang dinyatakan, statistik adalah suatu bidang yang berlandaskan matematik secara dasarnya. Ianya adalah sains matematik yang mengendalikan pengumpulan, analisis, tafsiran dan persembahan data. Objektifnya adalah untuk membuat inferen mengenai suatu populasi yang diminati berdasarkan maklumat yang diperolehi dari sampel-sampel ukuran dari populasi tersebut. Kaedah-kaedah statistik yang telah dibangun boleh digunakan untuk meringkas, menghurai koleksi data, mengenal pasti corak data agar dapat dimodelkan. Ianya kemudian diguna untuk membuat inferen mengenai proses atau populasi yang dikaji. Dari suatu ujikaji atau kajian, data dikumpul dan dianalisa supaya dapat menjawab soalan penyelidik dan memberikan suatu keputusan statistik yang bererti. Daripada keputusan ini juga, suatu teori baru boleh diuji dan dibangunkan.

Sebelum ujikaji dijalankan, rangka pengumpulan data dirancang supaya data yang sesuai dapat dikumpulkan. Ini dikenali sebagai reka bentuk ujikaji. Data yang dikumpul kemudiannya dianalisa dengan menggunakan kaedah

statistik yang mengandungi faktor atau aras yang diambil kira atau dipersetujui semasa proses merangka ujikaji untuk menghasilkan kesimpulan yang kukuh dan memenuhi objektif kajian. Pendekatan secara statistik pada reka bentuk ujikaji adalah perlu jika ingin mendapatkan kesimpulan yang bermakna daripada data (Montgomery, 1983).

Pemilihan faktor, aras, pemboleh ubah, sambutan dan reka bentuk ujikaji merupakan langkah-langkah penting dalam proses ujikaji. Dalam analisis rekabentuk ujikaji, hipotesis yang dijalankan ialah menguji kesamaan beberapa min. Tatacara yang digunakan ialah Analisis Varians (ANOVA) yang mengambil kira ralat ujikaji atau ubahan dalam data. Segala nilai ubahan yang diambil kira diringkaskan dalam Jadual ANOVA.

Reka bentuk ukuran ulangan

Kebanyakan penyelidikan sains gunaan memerlukan lebih dari satu cerapan daripada satu unit ujikaji (atau subjek) bagi penjimatan kos. Dengan itu, adalah penting untuk mendapat seberapa banyak maklumat yang boleh daripada satu unit ujikaji ini.

Salah satu rekabentuk yang boleh digunakan dalam kajian biologi ke atas data begini ialah reka bentuk ukuran ulangan. Sama seperti mana-mana ANOVA, hipotesis dalam reka bentuk ukuran ulangan ialah menguji kesamaan min. Dalam rekabentuk ukuran ulangan, data tercerap diperolehi daripada t masa yang berlainan tetapi merujuk kepada unit ujikaji (atau subjek) yang sama. Hasilnya ialah data bersandar memandangkan cerapan terkumpul bagi semua sampel atau aras rawatan adalah kepunyaan subjek yang sama.

Antara sebab lain penyelidik memerlukan lebih dari satu cerapan dari unit ujikaji adalah kerana dipercayai bahawa ubahan di kalangan atau di antara unit ujikaji adalah lebih besar jika dibandingkan dengan ubahan dalam suatu unit ujikaji (Ott, 2001). Dalam kes ini, adalah lebih baik menjadikan unit ujikaji sebagai blok dan setiap unit ujikaji diberi setiap rawatan. Kajian dengan reka bentuk ukuran ulangan mempunyai kelebihan untuk menambah baik kejituan anggaran (Agresti, 1999). Unit ujikaji atau subjek yang sama dalam setiap kumpulan boleh membantu penyelidik menghindar ralat yang berlebihan.

Jadual di bawah menunjukkan bentuk set data yang diperolehi dari suatu reka bentuk ujikaji dua faktor dengan ukuran ulangan pada satu faktor.

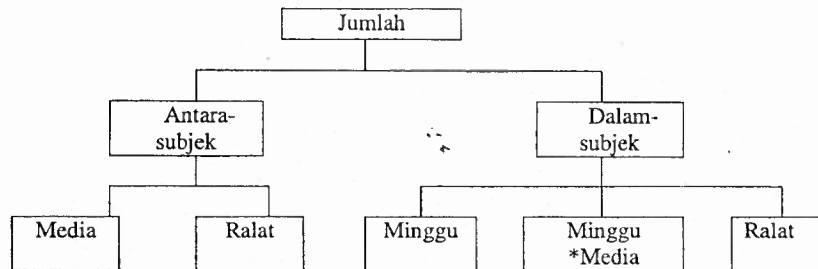
JADUAL A: Data reka bentuk ukuran ulangan

Pokok	Minggu					
	1	2	3	4	5	6
Media 1	1	...	b
	n					
...
	n					
Media a	1
n						

Dalam ujikaji di atas, subjeknya ialah pokok dan data tercerap ialah peratusan akar pokok yang keluar. Tujuan ujikaji adalah untuk mengkaji samada semua aras media memberi kesan yang sama ke atas perkembangan peratusan akar yang keluar. Faktor media terdiri dari sampel tak bersandar. Dua kesan tetap bagi ujikaji ini ialah media (1, 2, 3, ..., a) yang diwakili oleh a sampel tak bersandar dan minggu yang terdiri dari b kali cerapan. Minggu merupakan faktor dalam-subjek memandangkan perbandingan kategorinya menggunakan ukuran ulangan dalam sampel. Media adalah faktor antara-subjek kerana perbandingan kategorinya menggunakan pokok berbeza dari setiap kategori.

Walaupun kedua-dua faktor (media dan minggu) adalah kesan utama, analisisnya berbeza dari ANOVA dua-hala yang biasa digunakan. Ini adalah kerana ukuran ulangan ke atas faktor dalam-pokok menghasilkan kesan ketiga, iaitu kesan rawak bagi pokok. Setiap pokok diperhatikan pada setiap kategori minggu, tetapi hanya pada 1 kategori media. Dengan itu, pokok dikatakan tersarang dalam faktor media. Maka pembahagian ubahan bagi ANOVA dua hala ini adalah seperti dalam rajah di bawah:

JADUAL B: Rajah pembahagian ubahan



Model yang digunakan bagi ujikaji dua-faktor ini ialah

$$y_{ijk} = \mu + \alpha_i + \pi_{k(i)} + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}$$

$$i = 1, \dots, a; \quad j = 1, \dots, b; \quad k = 1, \dots, n$$

yang mana μ adalah min keseluruhan, α_i, β_j dan $\alpha\beta_{ij}$ adalah kesan tetap bagi media, minggu dan saling tindak, masing-masing. Kesan rawak yang disebabkan oleh pokok ke-k dalam media ke-i ditandai $\pi_{k(i)}$, adalah tak bersandar dan tertabur $N(0, \sigma_{\pi}^2)$ dan tak bersandar dengan ϵ_{ijk} yang juga tak bersandar dan tertabur $N(0, \sigma_{\epsilon}^2)$. Jadual ANOVA yang terhasil adalah seperti dalam jadual berikut.

JADUAL C: Jadual ANOVA ujikaji reka bentuk ukuran ulangan

Punca	SS	dk	MSE
Antara-subjek	SSM	a-1	SSM/a-1
Media	SSP(M)	a(n-1)	SSP(M)/a(n-1)
Pokok-dalam-Media			
Dalam-subjek	SSW	b-1	SSW/b-1
Minggu	SSMW	(a-1)(b-1)	SSMW/(a-1)(b-1)
Media*Minggu	SSE	(a-1)(b-1)(n-1)	SSE/(a-1)(b-1)(n-1)
Ralat			
Jumlah	SST	abn-1	

Untuk menguji kesan saling tindak dan dua kesan utama iaitu

$H_0 : \theta_{MW} = 0, H_0 : \theta_w = 0$ and $H_0 : \theta_M = 0$, statistik ujian F yang digunakan ialah $F = \frac{MSMW}{MSE}$, $F = \frac{MSW}{MSE}$ dan $F = \frac{MSM}{MSP(M)}$, masing-masing.

Apa yang diperhatikan, jika penyelidik tidak mengambil kira kesan rawak bagi pokok, penyelidik akan tersilap mengambil $F = \frac{MSM}{MSE}$ untuk menjalankan pengujian hipotesis $H_0 : \theta_M = 0$. Apatah lagi dengan keupayaan sofwer statistik masa kini yang dengan mudah mengeluarkan jadual ANOVA. Walau bagaimana pun, penyelidik perlu berhati-hati kerana sesetengah sofwer menggunakan MSE dalam semua pengiraan statistik ujian F . Dalam kes ini, jika penyelidik tidak memahami statistik ujian sepenuhnya, sudah tentulah penyelidik ini akan berpandu kepada nilai F tidak betul yang dikeluarkan oleh sofwer tersebut. Dengan itu, kesimpulan yang dihasilkan adalah kurang tepat. Di sinilah perlunya pengetahuan statistik supaya penyelidik dapat menggunakan statistik ujian yang betul dan seterusnya merumus kesimpulan yang tepat.

Terdapat juga sesetengah penyelidik yang mengabaikan terus kesan rawak yang wujud sedangkan ianya benar-benar wujud. Apabila diabaikan, maka MSE akan menjadi besar. Statistik ujian untuk menguji kesan media akan terus mengambil kira MSE . Untuk memahami keperluan dan penggunaan statistik ujian yang betul, adalah suatu perkara yang kadang kala merumitkan bagi penyelidik biologi, ditambah pula kurangnya pengetahuan dalam bidang matematik khasnya statistik. Untuk mengelak daripada melakukan kesilapan, elok kiranya penyelidik biologi bekerjasama dengan penyelidik statistik dalam menjalankan kajian supaya analisis yang betul dapat dijalankan dan matlamat penyelidikan tercapai. Kes seterusnya membincangkan keperluan penyelidik biologi dalam penyelidikan matematik pula.

Proses Bercabang

Suatu populasi terdiri dari subjek yang mampu menghasilkan keturunan dari jenis yang sama. Katakan setiap subjek, pada akhir hayatnya telah melahirkan se bilangan keturunan. Dalam proses evolusi dan kelahiran, diandaikan bahawa semua subjek bertindak secara tak bersandar antara satu sama lain dan pada akhir tempoh hayat mereka menjana suatu bilangan rawak inidividu (atau subjek) baru berdasarkan beberapa hukum

kebarangkalian. Proses $X(t)$, iaitu saiz populasi pada generasi ke- t (t diskret di sini) dikenali proses bercabang yang merupakan kelas khas bagi rantai Markov dan mempunyai aplikasi dalam bidang biologi. Juga dikenali sebagai proses stokastik bercabang kerana ia merupakan suatu model matematik yang menerangkan pembangunan proses dalam bentuk bercabang misalnya pengembangan atau pembiakan suatu populasi.

Kajian proses bercabang bermula pada tahun 1873 oleh Sir Francis Galton yang mengkaji kepupusan suatu kaum bangsawan dan masalah ini berjaya dijemakan dalam lelaran fungsi penjana oleh Rev. Henry Watson. Dalam tahun 1972, suatu kertas kerja yang telah ditulis oleh I. J. Bienayme dalam tahun 1845 yang mengkaji masalah yang serupa telah dijumpai. Sejak itu, proses bercabang masa diskret ini dikenali sebagai proses Bienayme-Galton-Watson (BGW).

Katakan pada masa kelahiran generasi ke- t , terdapat suatu imigrasi $Y(t)$ subjek ke dalam populasi tersebut. Maka, saiz populasi baru ini, $Z(t)$ dikenali sebagai proses bercabang dengan imigrasi. Dalam proses ini, setiap $Y(t)$ subjek akan memulakan suatu proses bercabang. Taburan bilangan immigrant (atau pendatang) boleh berbeza mengikut masa dan pembiakan populasi sedia ada. Faktor imigrasi agak penting dalam aplikasi proses biologi dan ekologi. Andaian ketakbersandaran proses imigrasi dan pembiakan juga adalah munasabah dalam kebanyakan bidang sains.

Proses bercabang berbeza antara satu sama lain bergantung kepada suatu parameter A iaitu min bilangan keturunan oleh satu subjek, yang dikenali sebagai keterkritikan proses. Terdapat tiga jenis proses bercabang. Jika $A < 1, A = 1$ dan $A > 1$, proses ini dikenali proses subkritikal, kritikal dan superkritikal, masing-masing. Berbagai kemungkinan boleh berlaku pada suatu proses dalam suatu tempah yang agak panjang. Salah satu kemungkinan ialah kepupusan proses. Dalam model kebarangkalian, jika $A \leq 1$ dan $A > 1$, kebarangkalian kepupusan ialah 1 dan < 1 , masing-masing. Walau bagaimana pun, tidak mustahil jika sesuatu proses akan wujud secara berterusan. Ini dikenali sebagai ketakpupusan proses. Kajian ke atas proses bercabang mengambil kira faktor kepupusan dan ketakpupusan ini.

Model proses bercabang berbeza antara satu sama lain dengan andaian tambahan ke atas proses pembiakan subjek baru dan juga boleh mengambil kira taburan tercantum bagi hayat hidup dan saiz keturunan. Objek utama dalam penyelidikan proses bercabang ialah proses rawak $X(t)$ (atau $Z(t)$ dalam proses dengan imigrasi), dan juga cirian lain yang menerangkan struktur pokok bagi suatu proses. Bilangan subjek dalam suatu generasi yang

mempunyai keturunan set tak kosong pada masa t dikenali sebagai proses bercabang terturun (Fleishman & Siegmund-Scgultz (1987), Yakimov (1980), Yeoh (1998)) dan bilangan pasangan yang mempunyai bilangan keturunan yang sama (Rahimov, 1995) adalah dua contoh pemboleh ubah rawak yang dikaji oleh penyelidik.

Masalah lebih pertama

Antara cirian yang dikaji ialah masalah lebih dalam proses bercabang. Kertas kerja ini membangkitkan masalah lebih pertama pada suatu aras yang diberi bagi sebilangan proses bercabang tak bersandar. Ditakrifkan

$$v(t, \theta) = \min\{k : X_k(t) > \theta(t)\}$$

bagi suatu fungsi aras $\theta(t)$ diberi, samada tetap atau menokok.

Proses $X_i(t); i=1,2,\dots$ boleh ditafsirkan sebagai saiz populasi yang wujud dalam bahagian yang berbeza dalam suatu kawasan yang luas. Maka, proses $v(t, \theta)$ ialah indeks proses pertama dalam famili tersebut yang mana saiznya melebihi aras $\theta(t)$. Beberapa kes boleh diambil kira untuk mengkaji masalah indeks pertama ini. Antaranya ialah apabila proses mempunyai taburan keturunan secaman, berbeza dan juga kewujudan famili imigrasi. Beberapa taburan had bagi $v(t, \theta)$ telah dibuktikan (Rahimov, 1998). Salah satu teorem yang telah dibuktikan bagi proses $X_i(t)$ dengan taburan secaman dinyatakan disini:

Teorem: Jika $A < 1$, dan $\theta(t) = \theta \in N = \{1, 2, \dots\}$, maka

$v(t, \theta) \in A[Q(t)y, E_1(x)]$, yang mana

$$E_1(x) = \begin{cases} 0 & ; x < 0 \\ 1 - e^{-q^*(\theta)x} & ; x \geq 0; \quad q^*(\theta) = \sum_{k=\theta+1}^{\infty} P_k^* \end{cases}$$

yang menggunakan Teorem Yaglom:

$$\lim_{t \rightarrow \infty} P[X(t) = k | X(t) > 0] = P_k^*; k = 1, 2, \dots$$

dan $Q(t)$ adalah kebarangkalian ketakpupusan.

Teorem tersebut telah dibuktikan menggunakan teknik jelmaan Laplace. Penghasilan teorem ini adalah satu kejayaan bagi bagi seorang penyelidik matematik atau statistik tulen kerana suatu teorem baru telah dibangunkan. Seperti yang telah dinyatakan di awal kertas kerja ini, suatu teorem akan lebih jelas dan bermakna jika dapat diaplikasikan kepada masalah nyata atau sebenar agar dapat diketahui sejauh mana ketepatan teorem yang telah dibuktikan. Penyelidik matematik kebiasaannya menggunakan contoh matematik untuk mengukuhkan teorem yang telah dibuktikan. Seperti dalam kes ini jika taburan keturunan adalah taburan geometri dengan $p > 1/2$ dan $\theta \in N$, maka $v(t)(1-A)A^t$ menumpu dalam taburan kepada suatu pemboleh ubah rawak yang mempunyai fungsi ketumpatan kebarangkalian eksponen $A^\theta e^{-A^\theta x}$ yang mana $A = q/p$, $q = 1-p$ dan $x > 0$. Bagi kes ini, $E[v(t)] \sim \left(\frac{p}{2p-1}\right)\left(\frac{p}{1-p}\right)^{\theta+t}$, $t \rightarrow \infty$.

Memandangkan masalah proses bercabang ini boleh diaplikasikan dalam proses biologi, adalah baik jika data nyata/sebenar dapat diperolehi dan dianalisa supaya ketepatan teorem dapat dilihat. Untuk mendapat data sebenar ini, sudah tentulah penyelidik statistik memerlukan bantuan kepakaran penyelidik biologi. Disebabkan kerumitan mendapat data yang boleh memenuhi kehendak masalah, penyelidik matematik biasanya kebiasaannya menggunkana data simulasi sebagai ganti kepada data sebenar.

KESIMPULAN

Daripada kedua-dua kes yang telah dinyatakan, jelas kelihatan bahawa penyelidik statistik dan biologi saling merlukan antara satu sama lain supaya hasil penyelidikan akan lebih bererti dan bermakna. Dalam kes reka bentuk ukuran ulangan, jika penyelidik biologi tidak mempunyai pengetahuan yang cukup, berkemungkinan berlakunya kesilapan dalam analisis dan seterusnya kesimpulan yang diperolehi adalah tidak kukuh. Dalam kes proses bercabang, penyelidik matematik atau statistik tulen amat memerlukan kepakaran penyelidik biologi dalam mendapatkan data sebenar bagi proses yang dapat memenuhi kehendak masalah supaya keputusan yang telah dibuktikan dapat ditentu sahkan dan diguna pakai.

RUJUKAN

- [1] Agresti A. & Finlay B. 1999, *Statistical Method for the Social Sciences*, 3rd ed. Prentice Hall.

Athreya K. B. & Ney P. E. 1972, *Branching Processes*, New York: Springer Verlag.

Fleishmann K & Siegmund-Schultze R. 1977, The structure of reduced critical G-W processes, *Math. Nachr.* **79**, 233-244.

Jagers P. 1975, *Branching Processes with Biological Applications*, London: Wiley.

Montgomery D C. 1983, *Design & Analysis of Experiments*, 2nd ed., John Wiley & sons.

Ott R. L. & Longnecker M. 2001, *An Introduction to Statistical Method & data Analysis*, 5th ed. Duxbury.

Rahimov I. 1995 *Random Sums and Branching Processes*, New York: Springer-Verlag.

Rahimov I. & Hasan H. 1998, Exceedance problems concerning a family of branching processes, *Pakistan J. Stats.* **14**(1), 37-47.

Yakimov A. 1980, Reduced Branching Processes, *Theory Prob. & Appl.* **25**, No. 3, 593-596.

Yeoh H. B. 1998, *Teorem Penghad bagi Proses Galton-Watson Terturun dengan Imigrasi*, Msc. Thesis. USM.

<http://en.wikipedia.org/wiki/Mathematical%5Fbiology>