Inference for Controlled Branching Process,

Bayesian Inference for Zero-inflated Count Data

and

Bayesian Techniques for Hairline Fracture Detection and Reconstruction

by

Archan Bhattacharya

(Under the direction of Gauri Sankar Datta and T. N. Sriram)

Abstract

Here three different problems are addressed in Statistics and its applications. The first chapter addresses a statistical estimation problem in controlled branching processes, then a Bayesian test is developed for testing zero-inflation in count and it ends with an application of Bayesian techniques in Bio-medical imaging.

Controlled branching processes (CBP) with a random control function provide a useful way to model generation sizes in population dynamics studies, where control on the growth of the population size is necessary at each generation. Motivated by the work of Wei and Winnicki (1990), we develop a weighted conditional least squares estimator of the offspring mean of the CBP and derive the asymptotic limit distribution of the estimator when the process is subcritical, critical and supercritical, respectively. The results obtained here extend those of Wei and Winnicki (1990) for branching processes with immigration and provide a unified limit theory of estimation.

A zero-inflated power series distribution is a mixture of a power series distribution and a degenerate distribution at zero, with a mixing probability p for the degenerate distribution. This distribution is useful for modeling count data that may have extra zeros. One question is whether the mixture model can be reduced to the power series portion, corresponding to $p = 0$, or whether there are so many zeros in the data that zero inflation relative to the pure power series distribution must be included in the model i.e., $p \geq 0$. The problem is difficult partially because $p = 0$ is a boundary point. Here, we present a Bayesian test for this problem. We compare our Bayesian solution to two standard frequentist testing procedures.

The next topic addresses two clinically challenging problems in the domain of virtual mandibular surgery, namely, (a) detection of minor/hairline fractures and (b) generation of target pattern (reconstructed mandible) with accompanied prognosis of fracture healing. Identification of hairline fractures in input X-ray or Computer Tomography (CT) images is very difficult, especially, in the presence of noise. We propose a Markov Random Field (MRF)-Maximum A Posteriori (MAP) probability based two-phase approach using the principles of Bayesian image restoration to solve both the aforementioned problems.

INDEX WORDS: Branching Process, Mixture Model, Computer Tomography, Bayesian Image Restoration, MRF, MAP.

Inference for Controlled Branching Process, Bayesian Inference for Zero-inflated Count Data and Bayesian Techniques for Hairline Fracture Detection and Reconstruction

by

Archan Bhattacharya

M.Sc., Indian Institute of Technology, Bombay, India, 1998

M.Sc., The University of Calcutta, India, 1995

A Dissertation Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the

Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2007

Inference for Controlled Branching Process, Bayesian Inference for Zero-inflated Count Data and Bayesian Techniques for Hairline Fracture Detection and Reconstruction

by

Archan Bhattacharya

Major Professors:	Gauri Sankar Datta
	T. N. Sriram
Committee:	William P. McCormick
	Lynne Seymour
	XiangRong Yin

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2007

Dedication

I would like to dedicate my thesis to my parents for their constant support and encouragement. It would not have been possible without that.

Acknowledgments

I would like to acknowledge the sincere help and immense support that I received from my major professors T. N. Sriram and Gauri Sankar Datta. I would also like to thank Prof. William McCormick, Dr. Lynne Seymour and Dr. XiangRong Yin for sincerely serving my committee. Also, I would like to thank other faculties in the department for their helpfulness.

I would like to acknowledge the technological and administrative support from Jesse, Jimmy, Daphne, Loretta, Julie and Connie. For moral support, I would like to thank my friends and colleagues Ross, Ellen, Amy, Anna, Guoying, Jennifer, Susanta, Jonathan and others.

Also I would like to acknowledge the support that I received from my collaborators and co-authors Ananda Chowdhury, Dr. Bertrand Clarke, Prof. Miguel González and others.

Last but not the least, I would like to express my gratitude to Prof. Uttam Bandyopadhyay, Dr. Gourangadeb Chattopadhyay from the University of Calcutta and Dr. Atanu Biswas and Dr. Sumitra Purakayastha from Indian Statistical Institute, Calcutta, for considering me worth of giving encouragement.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction and Literature Review

Three different problems are addressed in this dissertation from different fields of statistics and its applications. This chapter briefly gives introduction to all the three problems. Detailed introduction and literature review can be found in the respective chapters.

In Chapter 2, unified estimation problem in controlled branching process is considered. Branching processes have always provided a useful way to model population evolution and dynamics. Now, with the availability of high-speed computers, it has found new applications in areas such as algorithm, data structures, combinatorics, biology, demography, ecology, epidemiology, genetics, to mention a few. Branching process with immigration (BPI), the study of which dates back to Smoluchowski (1916), provides a useful growth model in many situations. There is a substantial literature on the topic of estimation of parameters associated with the BPI; see, for instance, Heyde (1970), Heyde and Seneta (1971, 1972, 1974), Klimko and Nelson (1978), Bhat and Adke (1981), Venkataraman (1982) and Venkataraman and Nanthi (1982).

Here, we consider situations in the study of population dynamics where some control on the growth of the population is necessary. In these instances, a class of models known as *Controlled Branching Processes* (CBP), introduced by Sevast'yanov and Zukov (1974), provides an interesting way to model the size of the population at a given time. For example, when it is necessary to control the size of some animal population, a reasonable methodology could be to achieve the corresponding control on the number of female animals. It can be shown that a CBP includes the BPI and Galton-Watson process (GWP) as special cases. Just as the BPI or GWP, the CBP also exhibits different growth behavior for different values of parameters associated with the process. Motivated by the work of Wei and Winnicki (1989, 1990), here we propose a conditional weighted least squares estimation approach for the

estimation of the offspring mean of a CBP with a random control function. We then derive the asymptotic limit distribution of the conditional weighted least squares estimator of the offspring mean in each of the three cases: (i) subcritical, (ii) critical and (iii) supercritical. Some results of independent interest are also established along the way, which generalize the results of Wei and Winnicki (1989, 1990).

In the next chapter, a Bayesian test for testing the presence of excess zeros in count is proposed. Standard models for count data often fail to fit in practice because of the presence of more zeros in the data than is explained by the model. This situation is often called zero inflation because the number of zeros is inflated from the baseline number of zeros that would be expected in, say, a Poisson distribution. Poisson family is a member of one-parameter discrete exponential families. Zero inflation is a special case of overdispersion that contradicts the relationship between the mean and variance in a one-parameter exponential family. Typical count data distributions, such as Poisson, cannot be used to model data containing extra zeros.

Johnson, Kotz and Kemp (1992, 312-318) discuss a simple modification of a power series (PS) distribution $f(\cdot|\theta)$ to handle extra zeros. An extra proportion of zeros, p is added to the proportion of zeros from the original discrete distribution, while decreasing the remaining proportions in an appropriate way. So the zero-inflated PS distribution is defined as

$$f^*(y|p,\theta) = \begin{cases} p + (1-p)f(0|\theta) & \text{if} \quad y = 0 \\ (1-p)f(y|\theta) & \text{if} \quad y > 0 \end{cases}, \tag{1}$$

where $\theta \in \Theta$, the parameter space and the mixing parameter p ranges over the interval $-f(0|\theta)/(1-f(0|\theta)) < p < 1$. This allows the distribution to be well defined for certain negative values of p, depending on θ. Although the mixing interpretation is lost when $p < 0$, these values have a natural interpretation in terms of zero deflation, relative to a PS model. Correspondingly, $p > 0$ can be regarded as zero inflation relative to a PS model.

The first question to be asked is whether the degenerate distribution at zero is necessary. If it is not, then no zero inflation needs to be modeled and the model simplifies to $f(y|\theta)$, which is equivalent of a hypothesis testing problem. In this work, a Bayesian test is studied and compared to other usual likelihood based tests.

In Chapter 4, a problem in the area of Bio-medical Imaging is solved using Bayesian image reconstruction paradigm. Craniofacial fractures, especially mandibular fractures, are frequently encountered. Mandibular fractures are observed to possess certain distinct patterns in X-ray or Computer Tomography (CT) images (see Ogundare et. al.(2004)). In some cases, the fractures are observed to be hairline or minor in nature. By the terms *hairline fracture* or *minor fracture* we respectively refer to those situations where the broken bone fragments are not visibly out of alignment or have incurred very little relative displacement. The presence of noise makes the detection and subsequent visualization of such types of fractures in X-ray or CT images a very challenging task. In case of a *major fracture*, i.e., fractures where the broken fragments are clearly displaced relative to each other, surgical intervention is almost mandatory. However, in the case of a *hairline/minor fracture*, the decision regarding surgical intervention is less clear as a surgeon can rely on natural bone healing. In this chapter, we propose a Markov Random Field (MRF)- Maximum A Posteriori probability (MAP) based scheme for (a) mandibular hairline fracture detection in the presence of noise and (b) target pattern generation for hairline/minor fractures.

1.1 References

[1] Bhat, B. R. and Adke, S. R. (1981). Maximum likelihood estimation for branching processes with immigration. *Adv. Appl. Prob.* vol 13. pp 498-509.

[2] Heyde, C. C., (1970). Extension of a result of Seneta for the supercritical Galton-Watson process. *Ann. Math. Statist.* vol 41. pp 739-742.

[3] Heyde, C. C. and Seneta, E., (1971). Analogous of classical limit theorems for the supercritical Galton-Watson process with immigration. *Math. Biosci.* vol 11. pp 249-259.

[4] Heyde, C. C. and Seneta, E., (1972). Estimation theory for growth immigration rates in a multiplicative process. *J. Appl. Prob.* vol 9. pp 235-258.

[5] Heyde, C. C. and Seneta, E., (1974). Notes on 'Estimation theory for growth immigration rates in a multiplicative process.' *J. Appl. Prob.* vol 11. pp 572-577.

[6] Johnson, N. L., Kotz, S., and Kemp, A. W. (1992). *Univariate Discrete Distributions.* Second edition. John Wiley & Sons Inc.

[7] Klimko, L. A. and Nelson, P. I., (1978). On conditional least squares estimation for stochastic processes. *Ann. Statist.* vol 6. pp 629-642.

[8] Ogundare, B.O., Bonnick, A. and Bayley, N. (2003) Pattern of Mandibular Fractures in an Urban Major Trauma Center, *J. of Oral and Maxillofac. Surg.*, vol. 61, no. 6, pp. 713-718.

[9] Sevast'yanov, B. A. and Zukov, A. (1974). Controlled Branching Processes. *Theor. Prob. Appl.* vol 19. Issue 1. pp 14-21.

[10] Smoluchowski, M. (1916). Drei vortrage uber diffusion Brownsche bewegung und wagulation von kolloidtelchen. *Physik. Zeits.* vol 17. pp 557-585.

[11] Venkataraman, K. N. (1982). A time series approach to the study of the simple subcritical Galton-Watson process with immigration. *Adv. Appl. Prob.* vol 14. pp 1-20.

[12] Venkataraman, K. N. and Nanthi, K. (1982). A limit theorem on a subcritical Galton-Watson process with immigration. *Ann. Prob.* vol 10. pp 1069-1074.

[13] Wei, C. Z. and Winnicki, J. (1989). Some asymptotic results for the branching process with immigration. *Stoch. Proc. Appl.* vol 31. pp 261-282.

[14] Wei, C. Z. and Winnicki, J. (1990). Estimation of the mean in the branching process with immigration. *Ann. Statist.* vol 18. pp 1757-1773.

Chapter 2

Estimation of the offspring mean in a Controlled Branching Process with a Random Control Function[1]

[1] T. N. Sriram, A. Bhattacharya, M. González, R. Martínez and I del Puerto. submitted in *Stochastic Processes and their Applications*.

2.1 Introduction

Branching processes have always provided a useful way to model population evolution and dynamics. Now, with the availability of high-speed computers, it has found new applications in areas such as algorithm, data structures, combinatorics, biology, demography, ecology, epidemiology, genetics, to mention a few. Branching process with immigration (BPI), the study of which dates back to Smoluchowski (1916), provides a useful growth model in many situations. There is a substantial literature on the topic of estimation of parameters associated with the BPI; see, for instance, Heyde (1970), Heyde and Seneta (1971, 1972, 1974), Klimko and Nelson (1978), Bhat and Adke (1981), Venkataraman (1982) and Venkataraman and Nanthi (1982). The above mentioned articles propose estimators which require prior knowledge about the growth behavior of the BPI, namely, whether the process is subcritical, critical and supercritical.

In an attempt to solve a long standing estimation problem of providing estimators of the parameters of BPI which do not require any prior knowledge about the growth behavior of the BPI, Wei and Winnicki (1990) proposed a unified estimation theory based on a conditional weighted least squares approach. The asymptotic limit distribution of Wei and Winnicki's (1990) estimator, however, changes as the growth behavior of BPI varies. Sriram, Basawa and Huggins (1991), on the other hand, showed that if one samples sequentially, then the limit distribution of the sequential estimator of the offspring mean is normal, regardless of whether the BPI is subcritical or critical. Qi and Reeves (2002) have recently extended the latter result to include the supercritical BPI.

Here, we consider situations in the study of population dynamics where some control on the growth of the population is necessary. In these instances, a class of models known as *Controlled Branching Processes* (CBP), introduced by Sevast'yanov and Zukov (1974), provides an interesting way to model the size of the population at a given time. For example, when it is necessary to control the size of some animal population, a reasonable methodology could be to achieve the corresponding control on the number of female animals. It can be shown that a CBP includes the BPI and Galton-Watson process (GWP) as special cases. Just

as the BPI or GWP, the CBP also exhibits different growth behavior for different values of parameters associated with the process. Motivated by the work of Wei and Winnicki (1989, 1990), here we propose a conditional weighted least squares estimation approach for the estimation of the offspring mean of a CBP with a random control function. We then derive the asymptotic limit distribution of the conditional weighted least squares estimator of the offspring mean in each of the three cases: (i) subcritical, (ii) critical and (iii) supercritical. Some results of independent interest are also established along the way, which generalize the results of Wei and Winnicki (1989, 1990).

In Section 1.2 we introduce the CBP along with basic notations which will be used throughout the paper. After a brief summary of estimation results for CBP, we introduce the conditional weighted least squares estimator of the offspring mean in a CBP. Asymptotic limit distribution of the estimator is derived separately for the subcritical, supercritical and the critical cases in Sections 2.3, 2.4 and 2.5, respectively. Some of the results of independent interest are stated as Lemmas in these sections, but proved in the appendix. Finally, concluding remarks are given in Section 2.6.

2.2 Literature Review

A CBP, where the number of individuals with reproductive capacity is controlled by a random control function $\phi_n(\cdot)$, is defined iteratively as (see Yanev (1975)),

$$Z_n = \sum_{j=1}^{\phi_{n-1}(Z_{n-1})} X_{n-1,j} \qquad n = 1, 2, \ldots \tag{2.1}$$

Here, Z_n denotes the size of the n-th generation of a population and $X_{n-1,j}$ is the offspring size of the j-th individual in the $(n-1)$-th generation. Throughout this section, we will assume that $\{X_{n,j}\}$ is a sequence of independent and identically distributed (i.i.d.) random variables with mean m and variance σ^2. We will also assume that the initial value Z_0 is a non-negative, integer-valued, square-integrable random variable which is independent of $\{X_{n,j}\}$.

Furthermore, we will assume that $\{\phi_n(k) : n \geq 0; k \geq 0\}$ defined by (2.1) are independent non-negative integer-valued random variables with identical one-dimensional distributions for

each k. We also assume that $\{X_{n,j}\}$ and $\{\phi_n(k)\}$ are independent. Following the notations of González, Molina and del Puerto (2002), we define

$$\begin{aligned} \varepsilon(k) &= E[\phi_n(k)], \\ \tilde{\sigma}^2(k) &= Var[\phi_n(k)], \qquad k = 0, 1, \ldots . \end{aligned} \tag{2.2}$$

It is worthwhile to note here that if $\phi_n(k) \equiv k$, then $\{Z_n\}$ is the GWP. Furthermore, if $\phi_n(k) = k + Y_n$, where $\{Y_n\}$ is a sequence of non-negative, integer-valued i.i.d. random variables with $0 < \alpha = E[Y_0] < \infty$, and is independent of $\{X_{n,j}\}$, then $\{Z_n\}$ is a BPI, where $\varepsilon(k) = k + \alpha$.

From (2.2) it follows that for $n = 1, 2, \ldots$,

$$\begin{aligned} E[Z_n|\mathcal{F}_{n-1}] &= m\varepsilon(Z_{n-1}), \\ Var[Z_n|\mathcal{F}_{n-1}] &= \sigma^2\varepsilon(Z_{n-1}) + m^2\tilde{\sigma}^2(Z_{n-1}), \end{aligned} \tag{2.3}$$

where $\mathcal{F}_n$ is the $\sigma-$algebra generated by the random variables $Z_0, Z_1, \ldots, Z_n$.

In a pioneering work, Wei and Winnicki (1989, 1990) provided a unified estimation theory for estimating the offspring mean and the immigration mean in a branching process with immigration defined by

$$Z_n = \sum_{j=1}^{Z_{n-1}} X_{n,j} + I_n, \qquad n = 1, 2, \ldots$$

where $\{I_n\}$ is the immigration process with mean λ and variance b^2, independent of the offspring random variables.

Suppose $\mathcal{F}_n$ denotes the $\sigma-$ field generated by $\{Z_0, X_{i,j}, I_i : 1 \le i \le n, j \ge 1\}$. Then, we can rewrite the BP with immigration as an autoregressive model of order 1 with appropriate error term given by

$$Z_n = mZ_{n-1} + \lambda + \epsilon_n, \qquad n = 1, 2, \ldots \tag{2.4}$$

Using this representation and the least squares principle we can construct an estimate of the offspring mean and the immigration mean given by

$$\begin{aligned} \hat{m_n} &= \frac{[\sum_{i=1}^n Z_i \sum_{i=1}^n Z_{i-1} - n\sum_{i=1}^n Z_i Z_{i-1}]}{\left[(\sum_{i=1}^n Z_{i-1})^2 - n\sum_{i=1}^n Z_{i-1}^2\right].} \\ \hat{\lambda}_n &= \frac{\left[\sum_{i=1}^n Z_i Z_{i-1} \sum_{i=1}^n Z_{i-1} - \sum_{i=1}^n Z_{i-1}^2 \sum_{i=1}^n Z_i\right]}{\left[\{\sum_{i=1}^n Z_{i-1}\}^2 - n\sum_{i=1}^n Z_{i-1}^2\right]}. \end{aligned} \tag{2.5}$$

In the subcritical case ($m < 1$), the asymptotic properties of these estimators have been studied by Venkataraman (1982). Also for $m > 1$, Heyde and Senata (1971) and Heyde (1970), respectively, showed that the ratio estimators Z_n/Z_{n-1} and $\sum_{i=1}^{n} Z_i / \sum_{i=1}^{n} Z_{i-1}$ can be used to estimate m. Note that the form of the estimators for m proposed by the above authors changes as the range of values of m changes. This does not solve the problem of how to estimate m if we do not know whether $m < 1$, $m = 1$ or $m > 1$.

In order to find a unified method of estimation, Wei and Winnick (1990) multiplied both sides of (2.4) by $(1 + Z_{n-1})^{-1/2}$ and computed a weighted conditional least squares estimate of the offspring mean given by

$$\begin{aligned} \tilde{m_n} &= \left[\sum_{i=1}^{n} Z_i \sum_{i=1}^{n} \tfrac{1}{(1+Z_{i-1})} - n \sum_{i=1}^{n} \tfrac{Z_i}{(1+Z_{i-1})}\right] \\ &\quad \times \left[\sum_{i=1}^{n} (1 + Z_{i-1}) \sum_{i=1}^{n} \tfrac{1}{(1+Z_{i-1})} - n^2\right]^{-1} \\ \tilde{\lambda}_n &= \left[\sum_{i=1}^{n} Z_{i-1} \sum_{i=1}^{n} \tfrac{Z_i}{(1+Z_{i-1})} - \sum_{i=1}^{n} Z_i \sum_{i=1}^{n} \tfrac{Z_{i-1}}{(1+Z_{i-1})}\right] \\ &\quad \times \left[\sum_{i=1}^{n} (1 + Z_{i-1}) \sum_{i=1}^{n} \tfrac{1}{(1+Z_{i-1})} - n^2\right]^{-1}. \end{aligned} \tag{2.6}$$

Using the weighted least square estimators in (2.6), Wei and Winnicki (1990) established a unified estimation theory for estimating m and λ for all values $m > 0$ by showing that in the supercritical case ($m > 1$), $\tilde{m_n} \to m$ a.s., and furthermore,

$$\left(\sum_{i=1}^{n} (Z_{i-1} + 1)\right)^{1/2} (\tilde{m_n} - m) \xrightarrow{d} N(0, \sigma^2).$$

They also showed that if $m = 1$, then $\tilde{m_n} \xrightarrow{P} m$, $\tilde{\lambda}_n \xrightarrow{P} \lambda$ and

$$\left(\sum_{i=1}^{n} (Z_{i-1} + 1)\right)^{1/2} (\tilde{m_n} - m) \xrightarrow{d} (Y(1) - \lambda) / \left(\int_0^1 Y(t)dt\right)^{1/2}.$$

where $Y(t) = \lim_{n\to\infty} Z_{[nt]}/n$ is a limiting diffusion process (refer to Wei and Winnicki(1989)). For subcritical case ($m < 1$), Wei and Winnicki (1990) showed that $\tilde{m_n}$ and $\tilde{\lambda}_n$ are strongly consistent and

$$\left(\left(\sum_{i=1}^{n} (Z_{i-1} + 1)\right)^{1/2} (\tilde{m_n} - m), \left(\sum_{i=1}^{n} (Z_{i-1} + 1)\right)^{1/2} (\tilde{\lambda}_n - \lambda)\right)' \xrightarrow{d} N_2(\mathbf{0}, V^{-1} W V'^{-1}),$$

where

$$V = \begin{bmatrix} \frac{EZ}{[Z(Z+1)]^{1/2}} & \left[E\frac{1}{Z+1}\right]^{1/2} \\ \frac{E[Z/(Z+1)]}{[E(Z+1)]^{1/2}} & \left[E\frac{1}{Z+1}\right]^{1/2} \end{bmatrix}$$

$$W = \begin{bmatrix} E(\sigma^2 Z + b^2) & E\left(\frac{\sigma^2 Z+b^2}{Z+1}\right) \\ E\left(\frac{\sigma^2 Z+b^2}{Z+1}\right) & E\left(\frac{\sigma^2 Z+b^2}{(Z+1)^2}\right) \end{bmatrix},$$

where Z is a random variable with the stationary distribution of the process $\{Z_n\}$.

While the results of Wei and Winnicki (1990) unify the estimation theory for m and λ, note that the limiting distributions of the weighted least squares estimator are drastically different for the case $m = 1$ and $m \neq 1$, respectively. In order to unify the distribution theory, Sriram, Basawa and Huggins (1991) proposed a sequential estimator and showed that the asymptotic distribution of the sequential estimator of the offspring mean m is normal when $m < 1$ and $m = 1$. In a non-sequential setting, Datta and Sriram (1995) gave a unified distribution theory using bootstrap methods for BP with observable immigration.

Recently, there has been a proliferation of literature in the area of CBP. Many foundational results for CBP have been developed in González, Molina and del Puerto (2002, 2003, 2004, 2005) and González, Martínez and del Puerto (2004, 2005). As for estimation results for the CBP defined by (2.1) with fixed control function $\phi(\cdot)$, González, Martínez and del Puerto (2004) considered the supercritical case (see Section 2.4 for definition) and carried out nonparametric estimation of the offspring distribution. Under the assumptions

1. $\tau := \lim_{n\to\infty} n^{-1}\phi(n) < \infty$
2. $\tau m > 1$ and Z_0 large enough such that $P[Z_n \to \infty] > 0$
3. $\{(\tau m)^{-n} Z_n\}_{n\geq 0}$ converges a.s. to a finite r.v. W such that $P[W > 0] > 0$

they showed that, on the set $\{Z_n \to \infty\}$, the following results hold (González et. al. (2003a)):

- $\lim_{n\to\infty} Z_n^{-1} Z_{n+1} = \tau m \qquad a.s.$
- $\lim_{n\to\infty} Z_n^{-1}\phi(Z_n) = \tau \qquad a.s.$

They proposed the following method of moment estimator for m defined by

$$\bar{m}_n = \begin{cases} \dfrac{Z_n}{\phi(Z_{n-1})} & \text{if} \quad Z_{n-1} > 0 \\ 0 & \text{if} \quad Z_{n-1} = 0, \end{cases}$$

where the value of the estimator is set to 0 for $Z_{n-1} = 0$, by convention. González et. al. (2003a) showed that, under the above assumptions,

(i) $\lim_{n\to\infty} P[\sigma^{-1}\phi(Z_{n-1})^{1/2}(\bar{m}_n - m) \le x] = \Phi(x)$

(ii) $\lim_{n\to\infty} P[\sigma^{-1}(\tau m)^{(n-1)/2}(\bar{m}_n - m) \le x] = \int_0^\infty \Phi(x\sqrt{\tau w})dF(w)$

where $\Phi(.)$ is the standard normal distribution function. The estimator $\bar{m}_n$ only depends on the generation sizes at $n-1$ and n, and does not take into account any of the the previous history of the process. If we do observe $Z_0, \ldots, Z_n$ for $n \ge 1$, then it seems reasonable to find the best linear convex combination of estimators $\bar{m}_k, k = 1, 2, \ldots, n$. Consequently, one may consider an estimator of the form:

$$\bar{\bar{m}}_n = \sum_{k=1}^{n} v_k \bar{m}_k \qquad \text{where} \qquad \sum_{k=1}^{n} v_k = 1.$$

Observing that $Var(Z_n|Z_{n-1}) = \sigma^2\phi(Z_{n-1})$, González et. al. (2003a) set $v_k \propto \phi(Z_{k-1})$, $k = 1, 2, \ldots$, and proposed an estimator of m of the form

$$\bar{\bar{m}}_n = \frac{\sum_{k=0}^{n} Z_k - Z_0}{\Delta_{n-1}} \tag{2.7}$$

for $n = 0, 1, \ldots$, where $Y_n = \sum_{k=0}^{n} Z_k$ and $\Delta_n = \sum_{k=0}^{n} \phi(Z_k)$, $n = 0, 1, \ldots$ This estimator is always well-defined even if $Z_j = 0$ for some j between 1 and n, since they assume $Z_0 = N > 0$. They showed that the above estimator is strongly consistent for m on $\{Z_n \to \infty\}$. They also proved the following asymptotic normality of the above estimator:

(i) $\lim_{n\to\infty} P\left[\sigma^{-1}\Delta_{n-1}^{1/2}(\bar{\bar{m}}_n - m) \le x\right] = \Phi(x)$

(ii) $\lim_{n\to\infty} P\left[\sigma^{-1}\left(\frac{(\tau m)^n - 1}{\tau m - 1}\right)^{1/2}(\bar{\bar{m}}_n - m) \le x\right] = \int_0^\infty \Phi(x\sqrt{\tau w})dF(w)$

In fact, using the maximum likelihood estimator (MLE) of the offspring distribution proposed in González, Martínez and del Puerto (2004), it is easy to show that $\bar{\bar{m}}_n$ is the MLE based

on the entire family tree up to the n-th generation. Moreover, from Lemma 2.13.2 of Jagers (1975), it can also be shown that $\bar{\bar{m}}_n$ is the MLE based on $\{Z_0, \ldots, Z_n\}$.

Dion and Essebbar (1995) considered the CBP in (2.1) with a random control function $\phi_n(Z_n) = \alpha_n \phi(Z_n)$, where $\{\alpha_n\}_{n\geq 0}$ is a sequence of i.i.d. random variables with values in $\mathbb{N}^+$, and $E[\alpha_0] = \alpha$ and $Var[\alpha_0] = d^2$. They referred to this process as a multiplicative CBP. Assuming that $\{X_{n,j}\}$ and $\{\alpha_n\}$ are independent of each other, and that $\{Z_n\}$ is supercritical, they provided an estimator of $\theta = m\alpha$ given by

$$\hat{\theta}_n = \frac{1}{n} \sum_{k=1}^{n} \frac{Z_k}{\phi(Z_{k-1})}$$

and proved that as $n \to \infty$,

(i) $\hat{\theta}_n \to \theta \qquad a.s.$

(ii) $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, d^2)$

As a matter of fact, they noted in their work that $\hat{\theta}_n$ is an asymptotic quasi-likelihood estimator of θ. They also proposed another estimator of $\rho = E[\log(m\alpha_0)]$ defined by

$$\hat{\rho}_n = \frac{1}{n} \sum_{k=1}^{n} \log\left(\frac{Z_k}{\phi(Z_{k-1})}\right)$$

and established that

(i) $\hat{\rho}_n \to \rho \qquad a.s.$

(ii) $\sqrt{n}(\hat{\rho}_n - \rho) \xrightarrow{d} N(0, \eta^2)$, where $\eta^2 = Var(log(\alpha_0))$

Note that the available estimation results have been developed under the assumption that the CBP is supercritical. As in the case of BPI, it would be useful to develop a unified estimation theory which does not require any assumption on the growth behavior of the CBP.

To this end, note that we can use (2.3) to represent Z_n as

$$Z_n = m\varepsilon(Z_{n-1}) + \tilde{\delta}_n, \qquad n = 1, 2, \ldots, \tag{2.8}$$

where the error term $\tilde{\delta}_n$ has $E\left[\tilde{\delta}_n|\mathcal{F}_{n-1}\right] = 0$. However, as in Wei and Winnicki (1990), the fact that the conditional variance $Var[\tilde{\delta}_n|\mathcal{F}_{n-1}] = \sigma^2\varepsilon(Z_{n-1}) + m^2\tilde{\sigma}^2(Z_{n-1})$ can be used to show that the resulting conditional least squares estimator is not efficient.

To overcome this, we divide both sides of (2.8) by $(\varepsilon(Z_{n-1})+1)^{1/2}$ and re-write the model as

$$\frac{Z_n}{(\varepsilon(Z_{n-1})+1)^{1/2}} = \frac{m\varepsilon(Z_{n-1})}{(\varepsilon(Z_{n-1})+1)^{1/2}} + \delta_n, \qquad n = 1, 2, \ldots,$$

with $\delta_n = \tilde{\delta}_n/\left(\varepsilon(Z_{n-1})+1\right)^{1/2}$ where

$$E\left[\delta_n|\mathcal{F}_{n-1}\right] = 0, \quad \text{and} \quad Var[\delta_n|\mathcal{F}_{n-1}] = \frac{\sigma^2\varepsilon(Z_{n-1}) + m^2\tilde{\sigma}^2(Z_{n-1})}{\varepsilon(Z_{n-1})+1}.$$

In this article we will impose certain regularity conditions on $\{\varepsilon(k)\}$ and $\{\tilde{\sigma}^2(k)\}$ so that the conditional variance above is bounded. These considerations lead us to the following conditional weighted least squares estimator of the offspring mean

$$\hat{m}_n = \left(\sum_{i=1}^{n} \frac{Z_i\varepsilon(Z_{i-1})}{\varepsilon(Z_{i-1})+1}\right)\left(\sum_{i=1}^{n} \frac{\varepsilon^2(Z_{i-1})}{\varepsilon(Z_{i-1})+1}\right)^{-1}. \tag{2.9}$$

In the subsequent sections, we derive the asymptotic limit distribution of $\hat{m}_n$ in the subcritical, supercritical and the critical cases of CBP, which are determined by the limiting behavior (as $k \to \infty$) of the quantity

$$\tau_m(k) = E[Z_{n+1}Z_n^{-1}|Z_n = k] = m\varepsilon(k)k^{-1}. \tag{2.10}$$

2.3 Subcritical case

We are interested in obtaining the asymptotic limit distribution of $\hat{m}_n$, defined in (2.9), for the subcritical case. We say CBP is subcritical in the case when $\limsup_{k\to\infty}\tau_m(k) < 1$. González, Molina and del Puerto (2004) established that a CBP $\{Z_n\}$ defined by (1) with $P(X_{0,1} = 0) > 0$, $P(X_{0,1} \leq 1) < 1$ and $P(\phi_0(i) > i) > 0$, $i = 0, 1, \ldots$ converges in distribution to a positive, finite and non-degenerate random variable Z. We make use of this result to obtain the asymptotic distribution of $\hat{m}_n$. To this end, we will assume the following regularity conditions:

A1: $\limsup_{k\to\infty}\tau_m(k) < 1$.

A2: $P(X_{0,1}=0)>0,\ P(X_{0,1}\leq 1)<1.$

A3: $P(\phi_0(i)>i)>0, \quad i=0,1,\ldots.$

A4: $E[\mu_{2+\delta}(Z)]<\infty$, where $\mu_k(z)=E[|\phi_0(z)-\varepsilon(z)|^k]$ for $k\geq 1$.

Note that it is customary to impose condition **A2** on the offspring sequence in order to avoid trivial situations. However, condition **A3** is crucial because it ensures that Z_n will not become extinct eventually (see González, Molina and del Puerto (2004) for more details). Finally, the moment condition on random control function ϕ_0 stated in **A4** is required in our proof of Theorem 1. We make use of the following results in proving Theorem 1.

Result 1(Theorem 3.2 in Hall and Heyde (1980)): Let $\{S_{ni},\mathcal{F}_{ni},1\leq i\leq k_n,n\geq 1\}$ be a zero-mean square-integrable martingale array with differences X_{ni}, and let η^2 be an a.s. finite random variable. Suppose that

$$\max_i |X_{ni}| \xrightarrow{P} 0, \tag{2.11}$$

$$\sum_i X_{ni}^2 \xrightarrow{P} \eta^2, \tag{2.12}$$

$$E\left(\max_i X_{ni}^2\right) \quad \text{is bounded in } n, \tag{2.13}$$

and

$$\mathcal{F}_{n,i}\subseteq\mathcal{F}_{n+1,i} \quad \text{for} \quad 1\leq i\leq k_n, n\geq 1. \tag{2.14}$$

Then $S_{nk}=\sum_i X_{ni}\xrightarrow{D} Z(\text{stably})$, where the random variable Z has characteristic function $E\exp(-\eta^2t^2/2)$.

Corollary 1.1(Corollary 3.1 in Hall and Heyde (1980)): If (2.11) and (2.13) are replaced by the conditional Lindeberg condition: for all $\epsilon>0$,

$$\sum_i E\left[X_{ni}^2 I(|X_{ni}|>\epsilon)|\mathcal{F}_{n,i-1}\right]\xrightarrow{P}0,$$

if (2.12) is replaced by an analogous condition on the conditional variance

$$V_{nk_n}^2=\sum E\left(X_{ni}^2|\mathcal{F}_{n,i-1}\right)\xrightarrow{P}\eta^2,$$

and if (2.14) holds, then the conclusion of the Result 1 remains true.

Theorem 1: Let $\{Z_n\}$ be a CBP defined by (2.1) satisfying **A1-A4**. Then,

$$\left(\sum_{i=1}^{n}\frac{\varepsilon^2(Z_{i-1})}{\varepsilon(Z_{i-1})+1}\right)^{1/2}(\hat{m}_n-m)\xrightarrow{d}N(0,V)\quad\text{as}\quad n\to\infty,$$

where

$$V=\frac{m^2E\left[\left(\frac{\varepsilon(Z)}{\varepsilon(Z)+1}\right)^2\tilde{\sigma}^2(Z)\right]+\sigma^2E\left[\frac{\varepsilon^3(Z)}{(\varepsilon(Z)+1)^2}\right]}{E\left[\frac{\varepsilon^2(Z)}{\varepsilon(Z)+1}\right]}.$$

Proof: Write

$$\left(\sum_{i=1}^{n}\frac{\varepsilon^2(Z_{i-1})}{\varepsilon(Z_{i-1})+1}\right)^{1/2}(\hat{m}_n-m)=\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\varepsilon^2(Z_{i-1})}{\varepsilon(Z_{i-1})+1}\right)^{-1/2}\sum_{i=1}^{n}Y_{ni},$$

where

$$Y_{ni}=n^{-1/2}(Z_i-m\varepsilon(Z_{i-1}))\left(\frac{\varepsilon(Z_{i-1})}{\varepsilon(Z_{i-1})+1}\right).\tag{2.15}$$

Let us define

$$S_{nk}=\sum_{i=1}^{k}Y_{ni}\quad\text{and}\quad V_{nn}=\sum_{i=1}^{n}E\left[Y_{ni}^2|\mathcal{G}_{n,i-1}\right],$$

where $\mathcal{G}_{n,i-1}=\mathcal{F}_{i-1}$(defined in the introduction).

By Theorem 3.1 of González, Molina and del Puerto (2004) and the ergodic theorem we have that, as $n\to\infty$,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\varepsilon^2(Z_{i-1})}{\varepsilon(Z_{i-1})+1}\to E\left[\frac{\varepsilon^2(Z)}{\varepsilon(Z)+1}\right]\qquad a.s.$$

and

$$V_{nn}\to m^2E\left[\left(\frac{\varepsilon(Z)}{\varepsilon(Z)+1}\right)^2\tilde{\sigma}^2(Z)\right]+\sigma^2E\left[\frac{\varepsilon^3(Z)}{(\varepsilon(Z)+1)^2}\right]\quad a.s.,$$

where the former result gives the asymptotic behavior of the normalizing factor in Theorem 1. Since $\lim_{n\to\infty}V_{nn}$ is a constant, the required conclusion will follow from Result 1 and Corollary 1.1, once we show for Y_{ni} defined in (2.15) that for all $\epsilon>0$,

$$\sum_{i=1}^{n}E\left[Y_{ni}^2I_{\{|Y_{ni}|>\epsilon\}}|\mathcal{G}_{n,i-1}\right]\xrightarrow{P}0\quad\text{as}\quad n\to\infty.$$

For this, it is sufficient to show that for $\delta \in (0,2)$

$$\sum_{i=1}^{n} E\left[|Y_{ni}|^{2+\delta}\,|\mathcal{G}_{n,i-1}\right] \xrightarrow{P} 0 \quad \text{as} \quad n \to \infty.$$

Note that for some $K_\delta > 0$

$$\begin{aligned} E\left[|Y_{ni}|^{2+\delta}\,|\mathcal{G}_{n,i-1}\right] &\leq \frac{2^{\delta+1}}{n^{1+\delta/2}}\left(K_\delta E\left[\phi_0(Z_{i-1})^{1+\delta/2}\right]\right. \\ &\quad \left. + m^{2+\delta} E\left[|\phi_0(Z_{i-1}) - \varepsilon(Z_{i-1})|^{2+\delta}\right]\right) \\ &= \frac{2^{\delta+1}}{n^{1+\delta/2}}\left(K_\delta m_{1+\delta/2}(Z_{i-1}) + m^{2+\delta}\mu_{2+\delta}(Z_{i-1})\right) \end{aligned}$$

where $m_k(z) = E[\phi_0(z)^k]$ and $\mu_k(z) = E[|\phi_0(z) - \varepsilon(z)|^k]$ for $k \geq 1$, $z = 0, 1, 2, \ldots$. Note that $\varepsilon(z) = m_1(z)$ and $\tilde{\sigma}^2(z) = \mu_2(z)$. Since $E[\mu_{2+\delta}(Z)] < \infty$, we have by the ergodic theorem that

$$\sum_{i=1}^{n} E\left[|Y_{ni}|^{2+\delta}\,|\mathcal{G}_{n,i-1}\right] \xrightarrow{P} 0 \quad \text{as} \quad n \to \infty,$$

and therefore the result.

2.4 Supercritical case

We now focus our attention on the supercritical case, that is, when $\lim_{k\to\infty} \tau_m(k)$ $= m \lim_{k\to\infty} k^{-1}\varepsilon(k) = \tau m > 1$. Before we state and prove the main result of this section, we note that the following result holds for the supercritical CBP:

$$P(Z_n \to \infty) > 0 \quad \text{and} \quad \lim_{n\to\infty} L_n = L \quad \text{a.s.}, \tag{2.16}$$

with $L_n = (\tau m)^{-n} Z_n$ and $P(L > 0) > 0$.

Remarks: 1) Conditions that guarantee (2.16) can be found in the papers González, Martínez and Mota (2005a,b). In fact, these papers consider a class of homogeneous multi-type Markov chains, which includes the CBP as a special case.

2) Note that from (2.16), it follows that, on $\{Z_n \to \infty\}$,

$$\frac{\varepsilon(Z_n)}{(\tau m)^n} \to \tau L \qquad a.s., \quad \text{as } n \to \infty,$$

and therefore,

$$\frac{1}{(\tau m)^n}\sum_{i=1}^{n}\frac{\varepsilon^2(Z_{i-1})}{\varepsilon(Z_{i-1})+1}\to\frac{\tau L}{(\tau m-1)}\qquad a.s.\quad\text{as}\quad n\to\infty. \tag{2.17}$$

To establish the main result we will consider the following regular conditions and result:

B1: $\limsup_{k\to\infty}\tau_m(k)>1$ and (2.16) hold.

B2: $\lim_{k\to\infty}k^{-1}\tilde{\sigma}^2(k)=0$.

Result 2 (Doeblin-Anscombe Theorem, see Chow and Teicher (1997) Section 9.4, Thm. 1): Let $\{X_n, n\geq 1\}$ be independent random variables with $E(X_n)=0, E(X_n^2)=1$, and $\{t_n, n\geq 1\}$ positive integer-valued random variables with $t_n/b_n\xrightarrow{P}c$, where $\{c, b_n, n\geq 1\}$ are positive, finite constants such that $b_n\uparrow\infty$. If $n^{-1/2}\sum_{j=1}^{n}X_j\xrightarrow{d}N(0,1)$, then

$$\frac{1}{t_n^{1/2}}\sum_{j=1}^{t_n}X_j\xrightarrow{d}N(0,1).$$

Theorem 2: Let $\{Z_n\}$ be the CBP defined by (2.1) satisfying **B1**-**B2**. Then

$$\left(\sum_{i=1}^{n}\frac{\varepsilon^2(Z_{i-1})}{\varepsilon(Z_{i-1})+1}\right)^{1/2}(\hat{m}_n-m)\xrightarrow{d}N\left(0,\sigma^2\right)\quad\text{as}\quad n\to\infty.$$

Proof: For simplicity, throughout the proof we will consider that $P(Z_n\to\infty)=1$. We re-write

$$\begin{aligned}&\left(\sum_{i=1}^{n}\frac{\varepsilon^2(Z_{i-1})}{\varepsilon(Z_{i-1})+1}\right)^{1/2}(\hat{m}_n-m)\\&=\left(\sum_{i=1}^{n}\frac{\varepsilon^2(Z_{i-1})}{\varepsilon(Z_{i-1})+1}\right)^{-1/2}\left(\sum_{i=1}^{n}(Z_i-m\varepsilon(Z_{i-1}))\left(\frac{\varepsilon(Z_{i-1})}{\varepsilon(Z_{i-1})+1}\right)\right)\\&=\left(\sum_{i=1}^{n}\frac{\varepsilon^2(Z_{i-1})}{\varepsilon(Z_{i-1})+1}\right)^{-1/2}\left((I)+(II)\right),\end{aligned}$$

where

$$(I)=\sum_{i=1}^{n}\left(\frac{\varepsilon(Z_{i-1})}{\varepsilon(Z_{i-1})+1}\right)\frac{(Z_i-m\varepsilon(Z_{i-1}))}{(Z_{i-1}+1)^{1/2}}\left((Z_{i-1}+1)^{1/2}-L^{1/2}(\tau m)^{(i-1)/2}\right)$$

and

$$(II) = L^{1/2} \sum_{i=1}^{n} (\tau m)^{(i-1)/2} \left(\frac{\varepsilon(Z_{i-1})}{\varepsilon(Z_{i-1}) + 1} \right) \frac{(Z_i - m\varepsilon(Z_{i-1}))}{(Z_{i-1} + 1)^{1/2}}.$$

Consider term (II) above. Define $\Gamma_n = \{\gamma_{ni}\}$, where

$$\gamma_{ni} = \begin{cases} \left(\dfrac{\varepsilon(Z_{n-i})}{\varepsilon(Z_{n-i}) + 1} \right) \dfrac{Z_{n-i+1} - m\varepsilon(Z_{n-i})}{(Z_{n-i} + 1)^{1/2}} & \text{for } i = 1, 2, \ldots, n \\ 0 & \text{otherwise.} \end{cases}$$

It can be shown that

$$\gamma_{ni} \xrightarrow{d} \text{ iid } N(0, \tau\sigma^2) \quad \text{as} \quad n \to \infty. \tag{2.18}$$

It is known that on $(\mathcal{R}^\infty, \mathcal{B})$, $\Gamma_n \to \Gamma$ weakly is equivalent to the weak convergence of $\{\gamma_{nk}, 1 \le k \le r\}$ for all integers $r \ge 1$ (see Billingsley (1968), p. 19). For $r \ge 1$ and $(\theta_1, \ldots, \theta_r)$, consider

$$\begin{aligned} \varphi_n(\theta_1, \ldots, \theta_r) &= E\left[e^{i\sum_{k=1}^{r} \theta_k \gamma_{nk}}\right] \\ &= E\left(\prod_{k=1}^{r} A_{nk}\right), \end{aligned}$$

where $A_{nk} = e^{i\theta_k \gamma_{nk}}$. Next, define

$$R_r^{(n)} = \prod_{k=1}^{r} A_{nk} - \prod_{k=1}^{r} e^{-\frac{\theta_k^2}{2\tau\sigma^2}}.$$

We propose to use the arguments very similar to those in proof of Theorem 1 of Heyde and Brown (1971) and the assumption that $\lim_{n\to\infty} \frac{Z_n}{(\tau m)^n} = L$ a.s. where $P[L > 0] > 0$, to conclude that $E\left[R_r^{(n)}\right] \to 0$ as $n \to \infty$ and

$$\varphi_n(\theta_1, \ldots, \theta_r) \to \exp[-\sum_{k=1}^{r} \theta_k^2 / 2\tau\sigma^2].$$

Indeed, to prove (2.18), following the proof of Theorem 1 in Heyde and Brown (1971) (or Shete and Sriram (2003)), it is enough to check that, for all sequences $z_n \to \infty$, as $n \to \infty$,

$$\frac{1}{(z_n + 1)^{1/2}} \left(\sum_{j=1}^{\phi_n(z_n)} (X_{nj} - m) \right) + \frac{1}{(z_n + 1)^{1/2}} \left(m\phi_n(z_n) - m\varepsilon(z_n) \right) \xrightarrow{d} N(0, \tau\sigma^2) \tag{2.19}$$

Using the Central Limit Theorem for randomly stopped sums (Result 2), the first summand in (2.19) converges in distribution to $N(0, \tau\sigma^2)$, as $n \to \infty$. Moreover, it is easy to

check, using Markov's inequality and **B2**, that the second summand in (2.19) converges in probability to 0, as $n \to \infty$. Hence (2.19) holds.

From (2.18) yields

$$\sum_{i=1}^{n} (\tau m)^{-i/2} \gamma_{ni} \xrightarrow{d} N\left(0, \tau\sigma^2(\tau m - 1)^{-1}\right) \quad \text{as} \quad n \to \infty.$$

Therefore, by (2.17) we have

$$\left(\sum_{i=1}^{n} \frac{\varepsilon^2(Z_{i-1})}{\varepsilon(Z_{i-1})+1}\right)^{-1/2} (II) \xrightarrow{d} N\left(0, \sigma^2\right) \quad \text{as} \quad n \to \infty.$$

Now, to find the limiting behavior of term (I), we write

$$|(I)| \leq A_n^{1/2} B_n^{1/2},$$

using Cauchy-Schwarz inequality, where

$$A_n = \sum_{i=1}^{n} \left((\tau m)^{(i-1)/2} \left(\frac{\varepsilon(Z_{i-1})}{\varepsilon(Z_{i-1})+1} \right)^2 \frac{((Z_i - m\varepsilon(Z_{i-1}))^2}{(Z_{i-1}+1)} \right),$$

and

$$B_n = \sum_{i=1}^{n} \left((\tau m)^{(i-1)/2} \left(\frac{(Z_{i-1}+1)^{1/2}}{(\tau m)^{(i-1)/2}} - L^{1/2} \right)^2 \right).$$

Since $Z_{i-1}/(\tau m)^{(i-1)} \to L$ a.s. as $i \to \infty$ by (2.16), we have that

$$B_n = o\left(\sum_{i=1}^{n} (\tau m)^{(i-1)/2}\right) = o\left((\tau m)^{n/2}\right) \quad a.s. \quad \text{as} \quad n \to \infty.$$

Also, by conditioning on $\mathcal{F}_{i-1}$,

$$E\left[A_n\right] \leq \tau\sigma^2 \sum_{i=1}^{n} (\tau m)^{(i-1)/2} = O\left((\tau m)^{n/2}\right),$$

implying that

$$|A_n| = O_p\left((\tau m)^{n/2}\right) \quad \text{as} \quad n \to \infty.$$

Hence

$$|(I)| = o_p\left((\tau m)^{n/2}\right) \quad \text{as} \quad n \to \infty. \tag{2.20}$$

Combining (2.20) and using (2.17), we get

$$\left(\sum_{i=1}^{n} \frac{\varepsilon^2(Z_{i-1})}{\varepsilon(Z_{i-1})+1}\right)^{-1/2} (I) \to 0 \ \text{ a.s., as } n \to \infty$$

and consequently the proof of the theorem.

2.5 Critical case

In this section, we provide the asymptotic limit distribution of the estimator defined by (2.9) under the critical case, that is, when $\liminf_{k\to\infty} \tau_m(k) \leq 1 \leq \limsup_{k\to\infty} \tau_m(k)$ for $\tau_m(k)$ defined in (2.10). We consider the critical CBP satisfying $P(Z_n \to \infty) > 0$ and the following conditions:

C1: $\tau_m(k) = 1 + k^{-1}\alpha + o(k^{-1})$ as $k \to \infty$, where α is a real number.

C2: $\lim_{k\to\infty} k^{-1}\tilde{\sigma}^2(k) = 0$.

Remark: González, Molina and del Puerto (2005) studied the behavior of the critical CBP, where they provided sufficient conditions for $P(Z_n \to \infty) > 0$. More specifically, under **C1** and **C2**, they showed that if $\alpha > \sigma^2/(2m)$ and an assumption on conditional moments holds, then $P(Z_n \to \infty) > 0$ (see Theorem 2 of González, Molina and del Puerto (2005)).

To prove the asymptotic behavior of our estimator we will need some preliminary results, which are of independent interest. These will be proved in the Appendix. The first result concerns a Feller diffusion approximation for the CBP. We introduce for each $n \geq 1$, a stochastic process $W_n(t) = n^{-1}Z_{[nt]}$, for $t \in \mathbb{R}_+$, $[\cdot]$ denoting the integer part. It is easy to see that $\{W_n\}$ is a sequence of random elements that take values in $D_{[0,\infty)}[0,\infty)$, which is the space of non-negative functions on $[0,\infty)$ that are right continuous and have left limits. We also denote by $C_c^\infty[0,\infty)$ the space of infinitely differentiable functions on $[0,\infty)$ which have a compact support. First, we establish the following result of independent interest.

Lemma 1: Let $\{Z_n\}$ be a CBP defined by (2.1) satisfying **C1**-**C2**. Then, the random process $\{W_n\}$ converges weakly, as $n \to \infty$, to a non-negative diffusion process W, with generator $Tf(x) = \alpha f'(x) + \frac{1}{2}x\sigma^2 m^{-1} f''(x)$, for $f \in C_c^\infty[0,\infty)$. The process W is the (unique) solution of the stochastic differential equation

$$\mathrm{d}W(t) = \alpha \mathrm{d}t + \left(\sigma^2 m^{-1} W(t)\right)^{1/2} \mathrm{d}\mathcal{W}(t), \qquad t \in [0,\infty),$$

where $\mathcal{W}$ is a standard Wiener process.

As an application of a general version of the continuous mapping theorem, we have the following result.

Lemma 2: Let $\{Z_n\}$ be the CBP defined by (2.1) satisfying **C1-C2**. Then,

$$\frac{W_n(1) - W_n(0) - \alpha}{\left(\frac{1}{m}\int_0^1 \tau_m(Z_{[nt]})W_n(t)dt\right)^{1/2}} \xrightarrow{d} \frac{W(1) - W(0) - \alpha}{\left(\frac{1}{m}\int_0^1 W(t)dt\right)^{1/2}} \quad \text{as} \quad n \to \infty.$$

The next result is similar to Theorem 2.12 of Wei and Winnicki (1989).

Lemma 3: Let $\{Z_n\}$ be a CBP defined by (2.1) satisfying **C1-C2**. Define $\{U_n\}$ as

$$U_n = \sum_{j=1}^{n} \frac{1}{1 + \varepsilon(Z_{j-1})}.$$

Then, $U_n \to \infty$ a.s. on $\{Z_n \to \infty\}$ as $n \to \infty$.

Result 3(Theorem 2.18 in Hall and Heyde (1980)): Let $\{S_n = \sum_{i=1}^{n} X_i, \mathcal{F}_n n \geq 1\}$ be a martingale and $\{U_n, n \geq 1\}$ a nondecreasing sequence of positive random variable such that U_n is $\mathcal{F}_{n-1}$-measurable for each n. If $1 \leq p \leq 2$ then

$$\sum_{1}^{\infty} U_i^{-1} X_i \quad \text{converges} \quad a.s. \tag{2.21}$$

on the set $\{\sum_{i=1}^{\infty} U_i^{-p} E(|X_i|^p|\mathcal{F}_{i-1}) < \infty\}$

$$\lim_{n\to\infty} U_n^{-1} S_n = 0 \quad a.s. \tag{2.22}$$

on the set $\{\lim_{n\to\infty} U_n = \infty, \sum_{i=1}^{\infty} U_i^{-p} E(|X_i|^p|\mathcal{F}_{i-1}) < \infty\}$. If $2 < p < \infty$, then (2.21) and (2.22) both hold on the set

$$\left\{\sum_{i=1}^{\infty} U_i^{-1} < \infty, \sum_{i=1}^{\infty} U_i^{-1-p/2} E(|X_i|^p|\mathcal{F}_{i-1}) < \infty\right\}$$

The main result of this section is given next.

Theorem 3: Let $\{Z_n\}$ be the CBP defined by (2.1) satisfying **C1-C2**. Then

$$\left(\sum_{i=1}^{n}\frac{\varepsilon^2(Z_{i-1})}{\varepsilon(Z_{i-1})+1}\right)^{1/2}(\hat{m}_n-m)\xrightarrow{d}\frac{W(1)-W(0)-\alpha}{\left(\frac{1}{m}\int_0^1 W(t)dt\right)^{1/2}}\quad\text{as}\quad n\to\infty.$$

Proof: Again for simplicity, we assume throughout the proof that $P(Z_n\to\infty)=1$. Write

$$\left(\sum_{i=1}^{n}\frac{\varepsilon^2(Z_{i-1})}{\varepsilon(Z_{i-1})+1}\right)^{1/2}(\hat{m}_n-m)$$

$$=\frac{\sum_{i=1}^{n}(Z_i-Z_{i-1})+\sum_{i=1}^{n}(Z_{i-1}-m\varepsilon(Z_{i-1}))-\sum_{i=1}^{n}(Z_i-m\varepsilon(Z_{i-1}))/(1+\varepsilon(Z_{i-1}))}{\left(\sum_{i=1}^{n}\varepsilon(Z_{i-1})\right)^{1/2}\left(1-\dfrac{\sum_{i=1}^{n}\varepsilon(Z_{i-1})/(1+\varepsilon(Z_{i-1}))}{\sum_{i=1}^{n}\varepsilon(Z_{i-1})}\right)^{1/2}}$$

$$=\frac{A_n+B_n-C_n}{(1-D_n)^{1/2}},$$

where

$$A_n=\frac{\sum_{i=1}^{n}(Z_i-Z_{i-1})}{\left(\sum_{i=1}^{n}\varepsilon(Z_{i-1})\right)^{1/2}},\qquad B_n=\frac{\sum_{i=1}^{n}(Z_{i-1}-m\varepsilon(Z_{i-1}))}{\left(\sum_{i=1}^{n}\varepsilon(Z_{i-1})\right)^{1/2}}$$

$$C_n=\frac{\sum_{i=1}^{n}(Z_i-m\varepsilon(Z_{i-1}))/(1+\varepsilon(Z_{i-1}))}{\left(\sum_{i=1}^{n}\varepsilon(Z_{i-1})\right)^{1/2}},\qquad D_n=\frac{\sum_{i=1}^{n}\varepsilon(Z_{i-1})/(1+\varepsilon(Z_{i-1}))}{\sum_{i=1}^{n}\varepsilon(Z_{i-1})}.$$

Now, write

$$\begin{aligned} A_n + B_n &= \frac{\frac{1}{n}\sum_{i=1}^{n}(Z_i - Z_{i-1}) + \frac{1}{n}\sum_{i=1}^{n}(Z_{i-1} - m\varepsilon(Z_{i-1}))}{\left(\frac{1}{n^2}\sum_{i=1}^{n}\varepsilon(Z_{i-1})\right)^{1/2}} \\ &= \frac{\left(\frac{Z_n - Z_0}{n}\right) + \frac{1}{n}\sum_{i=1}^{n}(Z_{i-1} - m\varepsilon(Z_{i-1}))}{\left(\frac{1}{n^2}\sum_{i=1}^{n}\varepsilon(Z_{i-1})\right)^{1/2}} \\ &= \frac{W_n(1) - W_n(0) - \sum_{i=1}^{n}\int_{(i-1)/n}^{i/n} Z_{[nt]}\left(\frac{m\varepsilon(Z_{[nt]})}{Z_{[nt]}} - 1\right)dt}{\left(\frac{1}{n}\sum_{i=1}^{n}\int_{(i-1)/n}^{i/n}\varepsilon(Z_{[nt]})dt\right)^{1/2}} \\ &= \frac{W_n(1) - W_n(0) - \int_0^1 Z_{[nt]}\left(\tau_m(Z_{[nt]}) - 1 - \frac{\alpha}{Z_{[nt]}}\right)dt - \alpha}{\left(\frac{1}{m}\int_0^1 \frac{m\varepsilon(Z_{[nt]})}{Z_{[nt]}}W_n(t)dt\right)^{1/2}}. \end{aligned}$$

The proof of the result is obtained proving the following:

a) $D_n \xrightarrow{P} 0$ as $n \to \infty$.

b) $C_n \to 0$ a.s. as $n \to \infty$.

c)

$$\int_0^1 Z_{[nt]}\left(\tau_m(Z_{[nt]}) - 1 - \frac{\alpha}{Z_{[nt]}}\right)dt \to 0, \quad \text{a.s. as } n \to \infty.$$

d)

$$A_n + B_n \xrightarrow{d} \frac{W(1) - W(0) - \alpha}{\left(\frac{1}{m}\int_0^1 W(t)dt\right)^{1/2}}, \text{ as } n \to \infty.$$

For a), note that

$$|D_n| \leq \frac{n}{\sum_{i=1}^{n}\varepsilon(Z_{i-1})} = \frac{1/n}{\frac{1}{n^2}\sum_{i=1}^{n}\varepsilon(Z_{i-1})}. \tag{2.23}$$

Also it can be verified using arguments in the proof of Lemma 2 that, as $n \to \infty$,

$$\frac{1}{n^2}\sum_{i=1}^{n}\varepsilon(Z_{i-1}) \xrightarrow{d} \frac{1}{m}\int_0^1 W(t)dt \tag{2.24}$$

and

$$\frac{1}{n^2}\sum_{i=1}^{n}\frac{\varepsilon^2(Z_{i-1})}{\varepsilon(Z_{i-1})+1} \xrightarrow{d} \frac{1}{m}\int_0^1 W(t)dt.$$

Combining (2.23) and (2.24), we have that $D_n \xrightarrow{P} 0$, as $n \to \infty$.

Now consider b). We write

$$C_n = \frac{\displaystyle\sum_{i=1}^{n}\frac{1}{1+\varepsilon(Z_{i-1})}}{\left(\displaystyle\sum_{i=1}^{n}\varepsilon(Z_{i-1})\right)^{1/2}} \frac{\displaystyle\sum_{i=1}^{n}(Z_i - m\varepsilon(Z_{i-1}))/(1+\varepsilon(Z_{i-1}))}{\displaystyle\sum_{i=1}^{n}\frac{1}{1+\varepsilon(Z_{i-1})}},$$

and set

$$U_n = \sum_{j=1}^{n}\frac{1}{1+\varepsilon(Z_{j-1})}.$$

By Lemma 3, $U_n \to \infty$ a.s. Now clearly,

$$\frac{\displaystyle\sum_{i=1}^{n}\frac{1}{1+\varepsilon(Z_{i-1})}}{\left(\displaystyle\sum_{i=1}^{n}\varepsilon(Z_{i-1})\right)^{1/2}} \le \frac{1}{\left(\displaystyle\frac{1}{n^2}\sum_{i=1}^{n}\varepsilon(Z_{i-1})\right)^{1/2}}$$

and moreover by (2.24), as $n \to \infty$,

$$\frac{1}{\left(\displaystyle\frac{1}{n^2}\sum_{i=1}^{n}\varepsilon(Z_{i-1})\right)^{1/2}} \xrightarrow{d} \frac{1}{\left(\displaystyle\frac{1}{m}\int_0^1 W(t)dt\right)^{1/2}}. \tag{2.25}$$

It remains to consider

$$\frac{\displaystyle\sum_{i=1}^{n}(Z_i - m\varepsilon(Z_{i-1}))/(1+\varepsilon(Z_{i-1}))}{U_n}.$$

It can be verified that $\displaystyle\sum_{i=1}^{n}(Z_i - m\varepsilon(Z_{i-1}))/(1+\varepsilon(Z_{i-1}))$ is a martingale. Therefore, by Result 3, if we show that

$$\sum_{i=1}^{\infty}\frac{E\left[(Z_i - m\varepsilon(Z_{i-1}))^2/(1+\varepsilon(Z_{i-1}))^2\,|\mathcal{F}_{i-1}\right]}{U_i^2} < \infty, \quad \text{as } n \to \infty$$

then

$$\frac{\sum_{i=1}^{n}(Z_i - m\varepsilon(Z_{i-1}))/(1+\varepsilon(Z_{i-1}))}{U_n} \to 0 \quad a.s. \tag{2.26}$$

Now, since

$$\begin{aligned} E\left[(Z_i - m\varepsilon(Z_{i-1}))^2/(1+\varepsilon(Z_{i-1}))^2 \,|\mathcal{F}_{i-1}\right] &= \frac{m^2\tilde{\sigma}^2(Z_{i-1}) + \tau^2\varepsilon(Z_{i-1})}{(1+\varepsilon(Z_{i-1}))^2} \\ &\leq \frac{m^2\tilde{\sigma}^2(Z_{i-1})}{(1+\varepsilon(Z_{i-1}))^2} + \frac{\tau^2}{(1+\varepsilon(Z_{i-1}))} \\ &\leq \frac{C_0}{(1+\varepsilon(Z_{i-1}))} \quad \text{by } \mathbf{C1} \text{ and } \mathbf{C2}, \end{aligned}$$

where C_0 is a generic constant. Therefore,

$$\begin{aligned} \sum_{i=1}^{\infty} \frac{E\left[(Z_i - m\varepsilon(Z_{i-1}))^2/(1+\varepsilon(Z_{i-1}))^2 \,|\mathcal{F}_{i-1}\right]}{U_i^2} &\leq C_0 \sum_{i=1}^{\infty} \frac{1/(1+\varepsilon(Z_{i-1}))}{U_i^2} \\ &= C_0 \sum_{i=1}^{\infty} \frac{(U_i - U_{i-1})}{U_i^2} < \infty. \end{aligned}$$

Combining (2.25) and (2.26), we get that $C_n \to 0$ a.s., as $n \to \infty$.

As for c), for each fixed $\omega \in \{Z_n \to \infty\}$, define a sequence of random variables $\{g_n\}$, where $g_n : ([0,1], \mathcal{B}[0,1], \mu) \longrightarrow (\mathcal{R}, \mathcal{B}(\mathcal{R}))$ such that

$$g_n(t) = Z_{[nt]}(\omega)\left(\tau_m(Z_{[nt]}(\omega)) - 1 - \frac{\alpha}{Z_{[nt]}(\omega)}\right).$$

Here μ is the Lebesgue measure. Then, by condition **C1** we have that $g_n \to 0 \quad a.s. \quad [\mu]$. Therefore, by the Dominated Convergence Theorem,

$$\int_0^1 Z_{[nt]}(\omega)\left(\tau_m(Z_{[nt]}(\omega)) - 1 - \frac{\alpha}{Z_{[nt]}(\omega)}\right) dt \to 0 \quad \text{as} \quad n \to \infty.$$

Finally, using c) and Lemma 2, we obtain d) and consequently the result.

Remark: In the critical case, it can be shown using Lemma 3 and $\varepsilon(Z_n) \to \infty$ on $\{Z_n \to \infty\}$ that $\sum_{i=1}^{n} \varepsilon^2(Z_{i-1})/(\varepsilon(Z_{i-1})+1) \to \infty$ on $\{Z_n \to \infty\}$.

2.6 Concluding remarks

1) We have proposed a weighted least squares estimator to estimate the offspring mean of the control branching process with a random control function. Under certain regularity conditions, we have shown that a suitably normalized version of this estimator is asymptotically normal in the supercritical and the subcritical case, while the limit distribution of the estimator is a function of a diffusion process in the critical case. Moreover, we can also prove that the estimator $\hat{m}_n$ is strongly consistent in the three cases. Recall that

$$\hat{m}_n - m = \frac{\sum_{i=1}^{n}(Z_i - m\varepsilon(Z_{i-1}))\varepsilon(Z_{i-1})/(1+\varepsilon(Z_{i-1}))}{\sum_{i=1}^{n}\varepsilon^2(Z_{i-1})/(1+\varepsilon(Z_{i-1}))}.$$

Note that $\sum_{i=1}^{n}(Z_i - m\varepsilon(Z_{i-1}))/(1+\varepsilon(Z_{i-1}))$ is a martingale and, in all the cases,

$$\sum_{i=1}^{n}\frac{\varepsilon^2(Z_{i-1})}{\varepsilon(Z_{i-1})+1} \to \infty \quad \text{a.s., as } n \to \infty.$$

Hence, using Theorem 2.18 of Hall and Heyde (1980) (or as in the proof of part (b) in Theorem 3), we obtain that $\hat{m}_n \to m$ a.s. as $n \to \infty$.

2) For supercritical CBP with fixed control function $\phi(\cdot)$, González, Martínez and del Puerto (2004) showed that

$$\left(\sum_{i=0}^{n-1}\phi(Z_i)\right)^{1/2}(\bar{\bar{m}}_n - m) \xrightarrow{d} N(0,\sigma^2), \text{ as } n \to \infty,$$

for $\bar{\bar{m}}_n$ defined in (2.7). Note from Theorem 2 that the limit distribution of $\hat{m}_n$ defined in (2.9) is the same as that of $\bar{\bar{m}}_n$, which was shown to be the maximum likelihood estimator based on $\{Z_0, \ldots, Z_n\}$. Therefore, $\hat{m}_n$ is also asymptotically efficient. It should be also noted that the limit distribution of $\bar{\bar{m}}_n$ is not studied in González, Martínez and del Puerto (2004) for the subcritical and critical cases, whereas we study these for all three cases.

3) Note that in the construction of our estimator of the offspring mean it is implicitly assumed

that the mean function $\varepsilon(k)$ defined in (2) is known. It is possible that this assumption is not valid, for instance, in the special case of BPI, defined below (2.2) with unknown α. In the case of unknown $\varepsilon(k)$, we can replace $\varepsilon(Z_{i-1})$ in (2.9) by $\phi_{i-1}(Z_{i-1})$ and define a modified estimator of the offspring mean by

$$\hat{\hat{m}}_n = \left(\sum_{i=1}^{n} \frac{Z_i \phi_{i-1}(Z_{i-1})}{\phi_{i-1}(Z_{i-1}) + 1} \right) \left(\sum_{i=1}^{n} \frac{\phi_{i-1}^2(Z_{i-1})}{\phi_{i-1}(Z_{i-1}) + 1} \right)^{-1}.$$

With minor modifications in the proof of Theorem 2, it is possible to show that the limit distribution of $\hat{\hat{m}}_n$ is also $N(0, \sigma^2)$ in the supercritical case. The limit distribution of $\hat{\hat{m}}_n$ in the subcritical and the critical cases are not known yet. Finally, note that our results extend those of Wei and Winnicki (1990) for the estimation of the offspring mean in the case of known α.

2.7 Appendix

Proof of Lemma 1: The arguments used to prove Lemma 1 are similar to those given in Ethier and Kurtz (1986). Therefore, we will only indicate the necessary steps. Note that $\{Z_k/n\}_{k\geq 0}$ is a Markov process with values in $E_n = \{l/n : l = 0, 1, \ldots\}$, for $n = 1, 2, \ldots$. For each $f \in C_c^\infty[0, \infty)$, define

$$T_n f(x) = E\left[f\left(\frac{Z_{n+1}}{n} \right) \middle| \frac{Z_n}{n} = x \right] = E\left[f\left(\frac{1}{n} \sum_{j=1}^{\phi_n(nx)} X_{nj} \right) \right], x \in E_n.$$

By Theorem 6.5 and Corollary 8.9 in Ethier and Kurtz (1986), it is sufficient to show that

$$\lim_{n\to\infty} \sup_{x\in E_n} |\delta_n^f(x)| = 0 \quad \text{for all } f \in C_c^\infty[0, \infty)$$

with $\delta_n^f(x) = n(T_n f(x) - f(x)) - Tf(x),\ x \in E_n,\ f \in C_c^\infty[0, \infty)$ and $Tf(x) = \alpha f'(x) + \frac{1}{2} x\sigma^2 m^{-1} f''(x)$.

Let

$$S_k = \sum_{j=1}^{\phi_n(k)} X_{nj} - k \left(= \sum_{j=1}^{\phi_n(k)} (X_{nj} - 1) + (\phi_n(k) - k) \right), \quad k = 0, 1, \ldots,$$

then we have

$$T_n f(x) = E\left[f\left(x + \frac{1}{n} S_{nx} \right) \right].$$

Now, by Taylor's formula,

$$T_n f(x) - f(x) = f'(x)\frac{1}{n}E\left[S_{nx}\right] + \frac{1}{n^2}E\left[S_{nx}^2 \int_0^1 (1-t) f''(x + \frac{t}{n}S_{nx})dt\right],$$

Then, we rewrite $\delta_n^f(x) = \delta_{n1}^f(x) + \delta_{n2}^f(x) + \delta_{n3}^f(x)$, where

$$\begin{aligned}
\delta_{n1}^f(x) &= f'(x)(nx\left(\tau_m(nx) - 1\right) - \alpha) \\
\delta_{n2}^f(x) &= \tfrac{1}{n}E\left[S_{nx}^2 \textstyle\int_0^1 (1-t)\left(f''(x + \frac{t}{n}S_{nx}) - f''(x)\right)dt\right] \\
\delta_{n2}^f(x) &= \tfrac{1}{2}f''(x)\left(\tfrac{1}{n}\left(m^2\tilde{\sigma}^2(nx) + \sigma^2\varepsilon(nx) + (nx)^2(\tau(nx)-1)^2\right) - x\sigma^2 m^{-1}\right).
\end{aligned}$$

To prove

$$\lim_{n\to\infty} \sup_{x\in E_n} \left|\delta_n^f(x)\right| = 0$$

it is sufficient to show that

$$\lim_{n\to\infty} \delta_n^f(x_n) = 0$$

for every sequence $\{x_n\}_{n\geq 1}$ with $x_n \in E_n$ and $n \in \mathbb{N}$ such that $x_n \to x \in [0,\infty]$. Note that

$$\begin{aligned}
E[S_{nx}] &= m\varepsilon(nx) - nx = nx\left(m\frac{\varepsilon(nx)}{nx} - 1\right) \\
E[S_{nx}^2] &= Var[S_{nx}] + E[S_{nx}]^2 = m^2\tilde{\sigma}^2(nx) + \sigma^2\varepsilon(nx) + (nx)^2\left(m\frac{\varepsilon(nx)}{nx} - 1\right)^2.
\end{aligned}$$

Then, by conditions **C1** and **C2**, we have

$$\lim_{n\to\infty} E[S_{nx}] = \alpha \quad \text{and} \quad \lim_{n\to\infty} \frac{1}{n}E[S_{nx}^2] = x\sigma^2 m^{-1}.$$

Hence, it is clear that $\lim_{n\to\infty} \delta_{ni}^f(x_n) = 0$ for $i = 1, 3$ and for all sequences $x_n \to x \in [0,\infty]$. Now, we consider $\delta_{n2}^f(x_n)$. Arguing as in page 389 of Ethier and Kurtz (1986), it remains to show that for any sequence $\{x_n\}$ such that $x_n \to x$, $0 < x < \infty$, $\{(nx_n)^{-1/2}S_{nx_n}\}$ converges in distribution. Rewrite

$$\frac{1}{(nx_n)^{1/2}}S_{nx_n} = \frac{1}{(nx_n)^{1/2}}\sum_{i=1}^{\phi_n(nx_n)} (X_{ni} - m) + (nx_n)^{1/2}\left(m\frac{\phi_n(nx_n)}{nx_n} - 1\right). \tag{2.27}$$

As for the first term on the right side of (2.27), it would follow from the Central Limit Theorem for randomly stopped sums (Doeblin-Anscombe's Theorem, see Result 2) that, as $n \to \infty$,

$$\frac{1}{(nx_n)^{1/2}}\sum_{i=1}^{\phi_n(nx_n)} (X_{ni} - m) \xrightarrow{d} N(0, \sigma^2 m^{-1}),$$

provided we show that

$$\frac{\phi_n(nx_n)}{nx_n} \xrightarrow{P} m^{-1}, \text{ as } n \to \infty.$$

Indeed, for large enough n, by **C1** and Markov's inequality,

$$\begin{aligned} P\left(\left|\frac{\phi_n(nx_n)}{nx_n} - m^{-1}\right| \geq \epsilon\right) = &\leq P\left(\left|\frac{\phi_n(nx_n)}{nx_n} - \frac{\varepsilon(nx_n)}{nx_n}\right| \geq \epsilon/2\right) \\ &\leq \frac{Var[\phi_n(nx_n)/nx_n]}{(\epsilon/2)^2} \\ &= 4\frac{\tilde{\sigma}^2(nx_n)}{(nx_n)^2\epsilon^2} \to 0, \end{aligned}$$

as $n \to \infty$, by condition **C2**. As for the second term in (2.27), it is easy to see using conditions **C1** and **C2** that

$$E\left[(nx_n)^{1/2}\left(m\frac{\phi_n(nx_n)}{nx_n} - 1\right)\right] = \frac{\alpha}{(nx_n)^{1/2}} + o\left(\frac{1}{(nx_n)^{1/2}}\right) \to 0,$$

$$Var\left[(nx_n)^{1/2}\left(m\frac{\phi_n(nx_n)}{nx_n} - 1\right)\right] = \frac{m^2}{nx_n}\tilde{\sigma}^2(nx_n) \to 0,$$

as $n \to \infty$. Therefore, as $n \to \infty$

$$(nx_n)^{1/2}\left(m\frac{\phi_n(nx_n)}{nx_n} - 1\right) \xrightarrow{P} 0.$$

Hence, the Lemma.

Remark: As note below (2.2), BPI is a special case of CBP. In the case of BPI with $m = 1$ and $\alpha = E(Y_n)$, we obtain the weak convergence result in Lemma 1 with generator $Tf(x) = \alpha f'(x) + \frac{1}{2}x\sigma^2 f''(x)$. This is the same conclusion as in Wei and Winnicki (1989).

Proof of Lemma 2: In view of Lemma 1, the proof depends on the following general version of the continuous mapping theorem:

> *Let S and T be two metric spaces, and $X, X_1, X_2, \ldots$ be random elements in S with $X_n \xrightarrow{d} X$. Consider some measure mappings $h, h_1, h_2, \ldots : S \to T$ and a*

measure set $C \subset S$ with $X \in C$ a.s. such that $h_n(s_n) \to h(s)$ as $s_n \to s \in C$. Then $h_n(X_n) \xrightarrow{d} h(X)$ (see Kallenberg (1997) for a proof).

In our case, $S = T = D_{[0,\infty)}[0,\infty)$. Let Λ be the set of strictly increasing continuous mappings of $[0,\infty)$ onto itself. It is known that a sequence $\{g_n\}$ of elements in $D_{[0,\infty)}[0,\infty)$ converges to a limit $g \in D_{[0,\infty)}[0,\infty)$ in the Skorohod topology if and only if there exist functions $\lambda_n \in \Lambda$ such that

$$\lim_{n\to\infty} \sup_{0\le t\le T} |\lambda_n(t) - t| = 0 \quad \text{for all} \quad T > 0 \tag{2.28}$$

and

$$\lim_{n\to\infty} \sup_{0\le t\le T} |g_n(t) - g(\lambda_n(t))| = 0 \quad \text{for all} \quad T > 0. \tag{2.29}$$

Note that, to prove Lemma 2, we only need to prove that as $n \to \infty$

$$\int_0^1 \tau_m(Z_{[nt]}) W_n(t)dt \xrightarrow{d} \int_0^1 W(t)dt,$$

and then the rest would follow from well-known results on the convergence in distribution.

By condition **C1**

$$\varepsilon(k) = m^{-1}(k + \alpha + c(k)), \tag{2.30}$$

where $c(k) \to 0$ as $k \to \infty$. Now, from continuous mapping theorem stated above and Lemma 1, it will follow that

$$\begin{aligned} \int_0^1 \tau_m\left(Z_{[nt]}\right) W_n(t)dt \quad &= \quad \int_0^1 \left(W_n(t) + \alpha n^{-1} + n^{-1}c(nW_n(t))\right) dt \\ &\xrightarrow{d} \quad \int_0^1 W(t)dt \quad \text{as} \quad n \to \infty, \end{aligned}$$

if we check that $h_n(g_n) \to h(g)$ as $g_n \to g$, $g \in C_c^\infty[0,\infty)$ with

$$h_n(g) = \int_0^1 \left(g(t) + \alpha n^{-1} + n^{-1}c([ng(t)])\right) dt$$

and

$$h(g) = \int_0^1 g(t)dt$$

Taking into account that both $h_n(g(t))$ and $h(g(t))$ do not depend on t, to check that $h_n(g_n) \to h(g)$, as $g_n \to g$, we have to prove that for n large enough $|h_n(g_n) - h(g)|$ is small enough.

From the fact that $g_n \to g$, we deduce that there exists a sequence $\{\lambda_n\} \subseteq \Lambda$ satisfying (2.28) and (2.29). Now,

$$\begin{aligned} |h_n(g_n) - h(g)| &\leq \int_0^1 |g_n(t) - g(t)|\, dt + \int_0^1 n^{-1} \left(\alpha + |c([ng(t)])|\right) dt \\ &\leq \int_0^1 |g_n(t) - g(\lambda_n(t))|\, dt + \int_0^1 |g(\lambda_n(t)) - g(t)|\, dt \qquad (2.31) \\ &\quad + \int_0^1 n^{-1} \left(\alpha + |c([ng(t)])|\right) dt. \end{aligned}$$

The first and second terms of the right hand side of (2.31) are small enough for n large enough due to (2.28) and (2.29). In fact, for second term we use that g is uniformly continuous in $[0,1]$. As for the third term, using that $c(k) \to 0$ (so $|c([ng(t)])| \leq L_0$, for some $L_0 > 0$ and for all $n \geq 1$ and $t \in [0,1]$), we deduce that it is less than $n^{-1}(\alpha + L_0)$. Hence the third term is also small enough for large n.

Proof of Lemma 3: From (2.30), for some constant $M > 0$,

$$\varepsilon(k) \leq m^{-1}(k + |\alpha| + |c(k)|) \leq m^{-1}(k + |\alpha| + M).$$

Let $\tilde{\alpha} = |\alpha| + M$ and $\tilde{\varepsilon}(k) = m^{-1}(k + \tilde{\alpha})$. Clearly, $\varepsilon(k) \leq \tilde{\varepsilon}(k)$ for all k. Define

$$M_n = \frac{1 + \tilde{\varepsilon}(Z_n)}{\prod_{i=0}^{n-1} \left(1 + \dfrac{m^{-1}\tilde{\alpha}}{1 + \tilde{\varepsilon}(Z_i)}\right)}, \ n = 1, 2, \ldots$$

Then, $\{M_n\}$ is a supermartingale with respect to $\{\mathcal{F}_n\}$, since

$$E[\tilde{\varepsilon}(Z_n)|\mathcal{F}_{n-1}] = m^{-1}(E[Z_n|\mathcal{F}_{n-1}] + \tilde{\alpha}) \leq \tilde{\varepsilon}(Z_{n-1}) + m^{-1}\tilde{\alpha}.$$

Hence,

$$E[M_n|\mathcal{F}_{n-1}] \leq \frac{1 + \tilde{\varepsilon}(Z_{n-1}) + m^{-1}\tilde{\alpha}}{\prod_{i=0}^{n-1} \left(1 + \dfrac{m^{-1}\tilde{\alpha}}{1 + \tilde{\varepsilon}(Z_i)}\right)} = M_{n-1}.$$

Since $\{M_n\}$ is a non-negative supermartingale, it converges almost surely to a non-negative and finite limit. This and the fact that $\tilde{\varepsilon}(Z_n) \to \infty$ a.s. on $\{Z_n \to \infty\}$ imply that, on $\{Z_n \to \infty\}$,

$$\prod_{n=0}^{\infty} \left(1 + \frac{m^{-1}\tilde{\alpha}}{1 + \tilde{\varepsilon}(Z_n)}\right) = +\infty \ \text{ a.s.},$$

and consequently,

$$\sum_{n=0}^{\infty} \frac{1}{1+\tilde{\varepsilon}(Z_n)} = +\infty \quad \text{a.s.}$$

Since, $\varepsilon(Z_n) \leq \tilde{\varepsilon}(Z_n)$ for all n, on $\{Z_n \to \infty\}$,

$$\sum_{n=0}^{\infty} \frac{1}{1+\varepsilon(Z_n)} = +\infty \quad \text{a.s.}$$

Hence the Lemma.

Acknowledgments

The authors would like to thank the associate editor and the referee for providing valuable comments, which significantly improved the presentation of the article.

2.8 References

[1] Bhat, B. R. and Adke, S. R. (1981). Maximum likelihood estimation for branching processes with immigration. *Adv. Appl. Prob.* vol 13. pp 498-509.

[2] Billingsley, P. (1968). *Probability and Measure*. John Wiley & Sons, New York.

[3] Chow, Y. S. and Teicher H. (1997). *Probability Theory: Independence, Interchangeability, Martingales*. Springer-Verlag, New York. 3rd ed.

[4] Dion, J. P. and Essebbar, B. (1995). On the statistics of controlled branching processes. *Lecture Notes in Statist.*, 99. pp 14-21.

[5] Ethier, S. N. and Kurtz, T. G. (1986). *Markov Processes: Characterization and Convergence.* Wiley, New York.

[6] González, M., Martínez, R. and Mota, M. (2005a). On the geometric growth in a class of homogeneous multitype Markov chain, *J. Appl. Prob.* vol 42. pp 1015-1030.

[7] González, M., Martínez, R. and Mota, M. (2005b). On the unlimited growth of a class of homogeneous multitype Markov chains. *Bernoulli*. vol. 11. pp 559-570.

[8] González, M., Martínez, R. and del Puerto, I. (2004). Nonparametric estimation of the offspring distribution and the mean for a controlled branching process. *Test.* vol 13. pp 465-479.

[9] González, M., Martínez, R. and del Puerto, I. (2005). Estimation of the variance for a controlled branching process. *Test.* vol 14. pp 199-213.

[10] González, M., Molina, M. and del Puerto, I. (2002). On the class of controlled branching processes with random control functions. *J. Appl. Prob.* vol 39. pp 804-815.

[11] González, M., Molina, M. and del Puerto, I. (2003). On the geometric growth in controlled branching processes with random control function. *J. Appl. Prob.* vol 40. pp 995-1006.

[12] González, M., Molina, M. and del Puerto, I. (2004). Limiting distribution for subcritical controlled branching processes with random control function. *Stat. & Prob. Lett.* vol 67. pp 277-284.

[13] González, M., Molina, M. and del Puerto, I. (2005). Asymptotic behavior of critical controlled branching processes with random control functions. *J. Appl. Prob.* vol 42. pp 463-477.

[14] Hall, P. and Heyde, C. C., (1980). *Martingale Limit Theory and Its Application.* Academic Press, San Diego.

[15] Heyde, C. C., (1970). Extension of a result of Seneta for the supercritical Galton-Watson process. *Ann. Math. Statist.* vol 41. pp 739-742.

[16] Heyde, C. C. and Seneta, E., (1971). Analogous of classical limit theorems for the supercritical Galton-Watson process with immigration. *Math. Biosci.* vol 11. pp 249-259.

[17] Heyde, C. C. and Seneta, E., (1972). Estimation theory for growth immigration rates in a multiplicative process. *J. Appl. Prob.* vol 9. pp 235-258.

[18] Heyde, C. C. and Seneta, E., (1974). Notes on 'Estimation theory for growth immigration rates in a multiplicative process.' *J. Appl. Prob.* vol 11. pp 572-577.

[19] Heyde, C. C. and Brown, B. M. (1971). An invariance principle and some convergence rate results for branching processes. *Z. Wahrsch. Verw. Geb.* vol 20. pp 271-278.

[20] Jagers, P. (1975). *Branching Processes with Biological Applications*. Wiley, London.

[21] Kallenberg, O. (1997). *Foundations of Modern Probability.* Springer-Verlag, New York.

[22] Klimko, L. A. and Nelson, P. I., (1978). On conditional least squares estimation for stochastic processes. *Ann. Statist.* vol 6. pp 629-642.

[23] Qi, Y. and Reeves, J. (2002). On sequential estimation for branching processes with immigration. *Stoch. Proc. Appl.* vol 100. pp 41-51.

[24] Sevast'yanov, B. A. and Zukov, A. (1974). Controlled Branching Processes. *Theor. Prob. Appl.* vol 19. Issue 1. pp 14-21.

[25] Shete, S and Sriram, T. N. (2003). A note on estimation in multitype supercritical branching processes with immigration. *Sankhyā.* vol 65. part 1. pp 107-121.

[26] Smoluchowski, M. (1916). Drei vortrage uber diffusion Brownsche bewegung und wagulation von kolloidtelchen. *Physik. Zeits.* vol 17. pp 557-585.

[27] Sriram, T. N., Basawa, I. V. and Huggins, R. (1991). Sequential estimation for branching processes with immigration. *Ann. Statist.* vol. 19. pp 2232-2243.

[28] Venkataraman, K. N. (1982). A time series approach to the study of the simple subcritical Galton-Watson process with immigration. *Adv. Appl. Prob.* vol 14. pp 1-20.

[29] Venkataraman, K. N. and Nanthi, K. (1982). A limit theorem on a subcritical Galton-Watson process with immigration. *Ann. Prob.* vol 10. pp 1069-1074.

[30] Wei, C. Z. and Winnicki, J. (1989). Some asymptotic results for the branching process with immigration. *Stoch. Proc. Appl.* vol 31. pp 261-282.

[31] Wei, C. Z. and Winnicki, J. (1990). Estimation of the mean in the branching process with immigration. *Ann. Statist.* vol 18. pp 1757-1773.

[32] Yanev, N. M. (1975). Conditions for degeneracy of ϕ-branching processes with random ϕ. *Theor. Prob. Appl.* vol 28. pp 481-491

Chapter 3

A Bayesian Test for Excess Zeros in a Zero-inflated Power Series Distribution[1]

[1] A. Bhattacharya, Bertrand S. Clarke and Gauri S. Datta. Submitted to *IMS Lecture Notes - Monograph Series: Festschrift in honor of Professor P. K. Sen,* 11/30/2006.

3.1 Introduction and Literature Review

Standard models for count data often fail to fit in practice because of the presence of more zeros in the data than is explained by the model. This situation is often called zero inflation because the number of zeros is inflated from the baseline number of zeros that would be expected in, say, a Poisson distribution. Poisson family is a member of one-parameter discrete exponential families. Zero inflation is a special case of overdispersion that contradicts the relationship between the mean and variance in a one-parameter exponential family. One way to address this is to use a two-parameter distribution so that the extra parameter can be used to give a larger variance to the resulting distribution. Efron (1986) developed the notion of double exponential family, a two-parameter modification of a standard one-parameter exponential family, that allows a higher variance than permitted by the one-parameter exponential. While reasonable in some examples, typical count data distributions, such as Poisson, cannot be used to model data containing extra zeros.

Johnson, Kotz and Kemp (1992, 312-318) discuss a simple modification of a power series (PS) distribution $f(\cdot|\theta)$ to handle extra zeros. An extra proportion of zeros, p is added to the proportion of zeros from the original discrete distribution, while decreasing the remaining proportions in an appropriate way. So the zero-inflated PS distribution is defined as

$$f^*(y|p,\theta) = \begin{cases} p + (1-p)f(0|\theta) & \text{if} \quad y = 0 \\ (1-p)f(y|\theta) & \text{if} \quad y > 0 \end{cases}, \tag{1}$$

where $\theta \in \Theta$, the parameter space and the mixing parameter p ranges over the interval $-f(0|\theta)/(1-f(0|\theta)) < p < 1$. This allows the distribution to be well defined for certain negative values of p, depending on θ. Although the mixing interpretation is lost when $p < 0$, these values have a natural interpretation in terms of zero deflation, relative to a PS model. Correspondingly, $p > 0$ can be regarded as zero inflation relative to a PS model. Note that the PS family contains all discrete one-parameter exponential families so appropriate choice of PS model in (1) permits any desired interpretation for the data corresponding to the second term. The first term allows an extra proportion p of zeros to be added to the discrete PS distribution; this data is effectively regarded as a sort of contamination. Note that, zero

inflation (zero deflation, respectively) does not imply that model (1) has larger (smaller, respectively) variance than the non-inflated version.

The first question to be asked is whether the degenerate distribution at zero is necessary. If it is not, then no zero inflation needs to be modeled and the model simplifies to $f(y|\theta)$. Clearly, this is a hypothesis testing problem. If p is not allowed to be negative, $p = 0$ is a boundary point and testing $\mathcal{H}_0 : p = 0$ vs. $\mathcal{H}_1 : p > 0$ is a notoriously difficult problem for both Bayesians and frequentists for which few results are available. (See Self and Liang (1987) and Silvapulle and Silvapulle (1995) for some asymptotics from a frequentist perspective.) Permitting negative values of p removes the boundary point problem so that the analytic challenges become manageable. The Bayes test obtained here compares favorably with standard frequentist methods in the real and simulated data cases we have examined.

Familiar cases in which testing $\mathcal{H}_0 : p = 0$ is useful include the zero-inflated Poisson, ZIP, distribution with parameters (p, θ) given by

$$f^*(y|p,\theta) = pI_{\{y=0\}} + (1-p)\frac{e^{-\theta}\theta^y}{y!}, \qquad y = 0, 1, 2, \cdots \tag{2}$$

in which $\theta > 0$, $\frac{-e^{-\theta}}{1-e^{-\theta}} < p < 1$ and $E(Y|p,\theta) = (1-p)\theta$ and the zero-inflated geometric distribution with parameters (p, θ):

$$f^*(y|p,\theta) = pI_{\{y=0\}} + (1-p)(1-\theta)\theta^y, \qquad y = 0, 1, 2, \cdots \tag{3}$$

in which $0 < \theta < 1$, $-\frac{1-\theta}{\theta} < p < 1$, and $E(Y|p,\theta) = (1-p)\theta/(1-\theta)$. The zero-inflated binomial is similar.

These models have been examined from a frequentist standpoint. The earliest results on zero inflation can be found in Cochran (1954) and Rao and Chakravarti (1956). In fitting a Poisson model to count data these authors checked whether lack of fit was due to the presence of extra zeros in the data by using an exact test and likelihood ratio test. Also in the context of a ZIP model, El-Shaarawi (1985) obtained the ML estimator and used its asymptotic distribution to construct a confidence interval for the mean parameter. A peculiarity of the MLE for p is that it can give negative values if there are no zeros in the data. Broek (1995) derived the score test for the zero inflation parameter p for testing $\mathcal{H}_0 : p = 0$ vs. $\mathcal{H}_1 : p \neq 0$.

This test developed for the two-sided alternative unfortunately gives up some power because the desired alternative is one-sided $\mathcal{H}_1 : p > 0$. A secondary problem is that the performance of this test deteriorates as the mean parameter increases. This may not be a serious problem because, as the mean increases, excess zeros will become more visually obvious since the Poisson model assigns ever less probability to zero.

More generally, Deng and Paul (2000) extended the score test to general one-parameter exponential families. Thus, they studied a regression model for the mean parameter of the exponential distribution. Later, Deng and Paul (2005) treated overdispersion and zero inflation simultaneously. In the ZIP context, motivated by industrial applications, Lambert (1992) fitted a logistic regression model for p and a log-linear model for θ, using an EM algorithm to obtain estimates. Hall (2000) extended this approach by adding random effects to the ZIP model and considered the case of a zero-inflated binomial model as well.

From the Bayesian standpoint, Ghosh et. al. (2006) estimated the parameters in a ZIP model in regression context as an alternative to traditionally used maximum likelihood based methods. Their simulation studies showed that the Bayesian method had better finite sample performance than the classical method, giving tighter interval estimates and higher coverage probabilities. Our work is in the spirit of their work for hypothesis testing.

In this paper, the main goal is to obtain a Bayes test for testing $\mathcal{H}_0 : p = 0$ vs. $\mathcal{H}_1 : p > 0$. To this end we consider the posterior probability

$$T(\mathbf{Y}) = P(p > 0|\mathbf{Y}) = \frac{\int_{\Theta}\int_{0}^{1} L(p,\theta)\pi(p|\theta)\pi(\theta)dpd\theta}{\int_{\Theta}\int_{\frac{-f(0|\theta)}{1-f(0|\theta)}}^{1} L(p,\theta)\pi(p|\theta)\pi(\theta)dpd\theta}, \tag{4}$$

in which $L(p,\theta)$ is the likelihood function from a zero-inflated PS model and $\mathbf{Y}$ is a vector of n data points. The corresponding rejection region is of the form $T(\mathbf{Y}) > c$ for suitable c. Using (4) necessitates careful consideration of prior selection so that neither the null nor the alternative hypothesis will be unduly favored. This is accomplished here by using Jeffreys' prior.

Treating (4) as a frequentist test statistic, we derive some of its properties. In particular, we obtain second order corrections for its asymptotic behavior. Then, we verify computationally that the power, a frequentist property, of the Bayes test for the ZIP family is roughly the same or a little higher than the power of the score test and the likelihood ratio test, for the hardest and most important ranges of n, p and θ, i.e., small to moderate p, small-ish θ, and small to moderate n. This is striking because the Jeffreys' priors used favor small θ's and p's near 0 and 1 and so are relatively unfavorable to the null. From the estimation standpoint, we verify that the posterior density is well behaved and gives reasonable credible intervals. As a final verification, we apply our techniques to three real data sets computing Bayes factors and score tests for the presence of zero inflation and obtaining estimates for the zero inflation as appropriate.

The structure of the paper is as follows. In Section 3.2 we provide some background on the properties of zero-inflated models from a Bayesian standpoint. In Section 3.3 we present the Bayesian test and give some of its properties. In Section 3.4 we develop Bayes estimation. In Section 3.5 we give our comparisons and in Section 3.6 we use our method to analyze three data sets.

3.2 Specifying the Bayes Model

Given a PS distribution it is easy to write down the zero-inflated model (1). Specification of a Bayes model also requires a prior distribution. In this section, we present some forms and properties of (1) along with the Fisher information matrix that will be required for finding objective priors. We start with the PS distribution case and then specialize.

3.2.1 *Zero-inflated Power Series Distributions*

Let $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$ be random sample of size n, from $f^*(y|p,\theta)$ defined in (1), where $f(y|\theta)$ is given by

$$f(y|\theta) = \frac{a_y \theta^y}{g(\theta)}, \qquad y = 0, 1, 2, \cdots, \tag{5}$$

in which $g(\theta) = \sum_{y=0}^{\infty} a_y \theta^y$ is the normalizing constant. It is easy to verify that $E(Y_1|p,\theta) = (1-p)\theta g'(\theta)/g(\theta)$. Writing

$$n_0 = \sum_{i=1}^{n} I_{[Y_i=0]}, \qquad S = \sum_{i=1}^{n} Y_i \quad \text{and} \quad \bar{Y} = S/n,$$

the likelihood function based on $\mathbf{Y}$ is

$$L(p,\theta) = \{p + (1-p)f(0|\theta)\}^{n_0} \left(\frac{1-p}{g(\theta)}\right)^{n-n_0} \theta^S. \tag{6}$$

Using (6), it is an exercise to derive ML estimates for (p,θ). From (5) it is easy to derive that the per unit Fisher information matrix $I(p,\theta) = ((I_{ij}(p,\theta)))$ for $i,j = 1,2$ is given by

$$I(p,\theta) = \begin{pmatrix} \frac{1-f(0|\theta)}{(1-p)\{p+(1-p)f(0|\theta)\}} & -\frac{\frac{g'(\theta)}{g(\theta)} f(0|\theta)}{p+(1-p)f(0|\theta)} \\ -\frac{\frac{g'(\theta)}{g(\theta)} f(0|\theta)}{p+(1-p)f(0|\theta)} & (1-p)\left[-\frac{\left\{\frac{g'(\theta)}{g(\theta)}\right\}^2 (p+f(0|\theta))}{p+(1-p)f(0|\theta)} + \frac{1}{g(\theta)}\left(g''(\theta) + \frac{g'(\theta)}{\theta}\right)\right] \end{pmatrix}.$$

It is seen that the off-diagonal terms are nonzero. (In general, however, the off-diagonal terms are zero under the reparametrization $p^* = p + (1-p)f(0|\theta)$.)

Two special cases of (5) occur regularly, the zero-inflated Poisson and geometric. The zero-inflated binomial is similar; we do not treat it explicitly here.

Zero-inflated Poisson

The ZIP distribution with parameters (p,θ) results from (1) by using the Poisson (θ) probability mass function in place of $f(y|\theta)$ as indicated in (2). Parallel to (6), the likelihood function based on a sample of size n is given by

$$L(p,\theta) = \left\{p + (1-p)e^{-\theta}\right\}^{n_0} \left\{(1-p)e^{-\theta}\right\}^{(n-n_0)} \theta^s. \tag{9}$$

Using (9), the MLE for (p,θ) can be derived, see El-Shaarawi (1985), as

$$\hat{\theta}_1 = \frac{S(1-e^{-\hat{\theta}_1})}{n-n_0} \quad \text{and} \quad \hat{p}_1 = \frac{\frac{n_0}{n} - e^{-\hat{\theta}_1}}{1-e^{-\hat{\theta}_1}} \tag{10}$$

provided $0 < n_0 < n$. Likewise, the test statistic for the score test for $\mathcal{H}_0 : p = 0$ can be derived as

$$T_s(\mathbf{Y}) = \frac{\left(\frac{n_0}{e^{-\hat{\theta}_0}} - n\right)^2}{n\left[\frac{1 - e^{-\hat{\theta}_0}}{e^{-\hat{\theta}_0}} - \hat{\theta}_0\right]}, \tag{11a}$$

where $\hat{\theta}_0 = \bar{Y}$ is MLE under $\mathcal{H}_0$, see Broek (1995). It can be shown that $\text{sgn}(\hat{p})\sqrt{T_s(\mathbf{Y})}$ asymptotically follows a standard normal distribution under $\mathcal{H}_0 : p = 0$ and a level α rejection region for testing $\mathcal{H}_0 : p = 0$ vs. $\mathcal{H}_1 : p > 0$ is given as

$$\text{sgn}(\hat{p})\sqrt{T_s(\mathbf{Y})} > z_\alpha, \tag{11b}$$

where z_α is the upper α cut-off point on standard normal distribution.

Similarly, the likelihood ratio test can be derived. We omit the details since unlike the score test statistic, the likelihood ratio statistic does not have an explicit expression. If we denote the likelihood ratio test statistic by $T_l(\mathbf{Y})$, then $[\text{sgn}(\hat{p})\sqrt{T_l(\mathbf{Y})}]$ asymptotically follows $N(0, 1)$ under $\mathcal{H}_0 : p = 0$ and a level α rejection region for testing $\mathcal{H}_0 : p = 0$ vs. $\mathcal{H}_1 : p > 0$ is given as

$$\text{sgn}(\hat{p})\sqrt{T_l(\mathbf{Y})} > z_\alpha. \tag{12}$$

For the ZIP family, the Fisher information matrix has entries

$$I(p, \theta) = \begin{pmatrix} \dfrac{1 - e^{-\theta}}{(1-p)\{p + (1-p)e^{-\theta}\}} & -\dfrac{e^{-\theta}}{p + (1-p)e^{-\theta}} \\ \\ -\dfrac{e^{-\theta}}{p + (1-p)e^{-\theta}} & \dfrac{1-p}{\theta} - \dfrac{p(1-p)e^{-\theta}}{p + (1-p)e^{-\theta}} \end{pmatrix}.$$

Zero-inflated Geometric

The zero-inflated geometric distribution with parameters (p, θ) results from (1) by using the geometric (θ) probability mass function in place of $f(y|\theta)$ as indicated in (3). Parallel to (6), the likelihood function based on a sample of size n is given by

$$L(p, \theta) = \{p + (1-p)(1-\theta)\}^{n_0} \{(1-\theta)(1-p)\}^{n-n_0} \theta^s. \tag{13}$$

Using (13) the MLE's for (p, θ) can be derived; the test statistic for the score test for $\mathcal{H}_0$ is

$$T_s(Y) = \frac{n(1+\bar{Y})}{\bar{Y}^2}\left[\frac{n_0}{n}(1+\bar{Y}) - 1\right]^2.$$

In this case, the Fisher information matrix has entries

$$I(p,\theta) = \begin{pmatrix} \dfrac{\theta}{(1-p)\{p+(1-p)(1-\theta)\}} & \dfrac{1}{p+(1-p)(1-\theta)} \\ \\ \dfrac{1}{p+(1-p)(1-\theta)} & (1-p)\left\{\dfrac{\theta+(1-\theta)^2}{(1-\theta)^2\theta} - \dfrac{1-p}{p+(1-p)(1-\theta)}\right\} \end{pmatrix}.$$

For testing $\mathcal{H}_0 : p = 0$ vs. $\mathcal{H}_1 : p > 0$, the score test and likehood ratio test can be expressed in terms of rejection regions similar to those given in (11b) and (12).

3.2.2 *Prior Specification*

It is well known that Jeffreys' prior is the reference prior in the absence of nuisance parameters, see Bernardo (1979) and Clarke and Barron (1994). That is, Jeffreys' prior is objective in the sense that using Jeffreys' prior gives a posterior that updates the prior as much as possible on average in relative entropy. Informally, it permits maximal information gain in a data transmission sense. For a small number of parameters, here 2, this is a reasonable optimality criterion.

By definition, Jeffreys' prior on (p, θ) is proportional to the square root of the determinant of the Fisher information matrix,

$$\pi_J(p,\theta) \propto \left(\det(I(p,\theta))\right)^{1/2}. \tag{14}$$

For a zero-inflated power series distribution, there is no convenient expression in general for $\det(I(p,\theta))$. However, for the ZIP model, Jeffreys' prior is

$$\pi_J(p,\theta) \propto \frac{(1-e^{-\theta}-\theta e^{-\theta})^{1/2}}{[\theta\{p+(1-p)e^{-\theta}\}]^{1/2}}. \tag{15a}$$

As is typical for reference priors, (15a) is improper: The integral over $\theta \in [0,\infty)$ diverges. Likewise, in a zero-inflated geometric, Jeffreys' is

$$\pi_J(p,\theta) \propto \frac{\theta^{1/2}}{(1-\theta)\{p+(1-p)(1-\theta)\}^{1/2}}. \tag{15b}$$

Again, this is improper because the integral over θ diverges.

Jeffreys' prior given by (14) would be appropriate if both p and θ are of equal interest. Here, we are mainly interested in p. So, we used the Jeffreys' prior for p for given θ, that is

$$\pi_J^c(p|\theta) \propto [I_{11}(p,\theta)]^{1/2}, \tag{16}$$

and used the Jeffreys' prior for θ obtained from the regular model derived from $f(y|\theta)$.

For a zero-inflated PS model, (16) gives

$$\pi_J^c(p|\theta) = \frac{(1-f(0|\theta))^{1/2}}{\pi(1-p)^{1/2}\left[p+(1-p)f(0|\theta)\right]^{1/2}}. \tag{17}$$

If $g(\theta)$ corresponds to a zero-inflated Poisson model, (17) gives

$$\pi_J^c(p|\theta) = \frac{1}{\pi}\cdot\left[\frac{1-e^{-\theta}}{(1-p)\{p+(1-p)e^{-\theta}\}}\right]^{1/2}, \tag{18a}$$

and if $g(\theta)$ corresponds to a zero-inflated geometric distribution, (17) gives

$$\pi_J^c(p|\theta) = \frac{1}{\pi}\cdot\left[\frac{\theta}{(1-p)\{p+(1-p)(1-\theta)\}}\right]^{1/2}. \tag{18b}$$

Note that in both (18a,b) the range of p includes a range of negative values, depending on θ, for which the prior density is well defined.

Parallel to (16), the Jeffreys' prior for θ in the PS model is

$$\pi_J(\theta) \propto \left[\frac{g''(\theta)}{g(\theta)} + \frac{g'(\theta)}{\theta g(\theta)} - \left\{\frac{g'(\theta)}{g(\theta)}\right\}^2\right]^{1/2}. \tag{19}$$

Expression (19) gives $1/\sqrt{\theta}$ and $1/[(1-\theta)\sqrt{\theta}]$ for the Poisson and geometric model, respectively. These are improper, however, the posterior turns out to be proper because a finite number of data points suffice to make it so. If a proper joint objective prior is desired, Rissanen's prior, see Rissanen (1983) can be used and gives similar results but is computationally more demanding.

3.3 Test Criterion Based on Posterior Probability

For the general case of a zero-inflated PS model, the posterior is formed by using (6) and (14). In the ZIP model, these expressions become (9) and (15a). Another reasonable choice would be (9) with (17), which becomes (18a), and $\pi(\theta) = 1/\sqrt{\theta}$.

Given these choices, the Bayes test for zero inflation is based on the posterior probability that $p > 0$. Thus, consider the statistic

$$T(\mathbf{Y}) = P(p > 0|\mathbf{Y}) = \frac{\int_{\Theta}\int_{0}^{1} L(u,\theta)\pi(u|\theta)\pi(\theta)dud\theta}{\int_{\Theta}\int_{\frac{-f(0|\theta)}{1-f(0|\theta)}}^{1} L(u,\theta)\pi(u|\theta)\pi(\theta)dud\theta}. \tag{20}$$

The main point of this section is to derive an asymptotic test based on T by finding the asymptotic distribution of $T(\mathbf{Y})$ under $\mathcal{H}_0 : p = 0$. It is reasonable to conclude there is zero inflation when $P(p > 0|\mathbf{Y})$ is close to one. Consequently, from a frequentist standpoint, the rejection region is given by $T(\mathbf{Y}) > c$ where c is chosen based on the given level of significance.

3.3.1 *Finite Sample Properties of the Test Statistic*

Note that (20) exploits the extended parameter space for p, namely, $-f(0|\theta)/(1-f(0|\theta)) < p < 1$ so that as θ increases, the lower bound approaches 0 from the left.

Let $P_{(p,\theta)}(\cdot)$ be the probability measure for a zero-inflated PS family. It can be verified that for large sample size, $P_{(p,\theta)}(T(\mathbf{Y}) > c)$ is increasing in p for fixed θ and increasing in θ for fixed p. That is, as zero inflation increases the probability that T is large (close to one) increases and that as the probability of large outcomes of Y increases the probability of zero inflation also increases. This latter fact just means that even a single occurrence of zero can appear to suggest zero inflation if θ is large enough.

One feature which makes T easy to use is that the joint posterior distribution for (p, θ) and the marginal posterior for p are typically unimodal. Indeed, $\pi(p|\mathbf{y})$ is typically unimodal, even for small sample sizes. The posterior densities from the simulations reported in Sec. 3.5 and the data analysis in Sec. 3.6 are all unimodal.

3.3.2 *Asymptotic Distribution of the Test Statistic*

For ease of exposition, let $\eta = (\eta_1, \eta_2) = (p, \theta)$ so that $Y_1, ..., Y_n$ are IID with density $f^*(\cdot|\eta)$ and $T(\mathbf{Y}) = P_\eta(\eta_1 > \eta_{10}|\mathbf{Y})$ where η_{10} is a fixed value.

First, we sketch a proof that under $\eta = (\eta_{10}, \eta_{20}) = \eta_0$, the frequentist distribution of T is asymptotically Uniform$[0, 1]$, i.e., $T = U + o_p(1)$ as $n \to \infty$ where U has a uniform$[0, 1]$ distribution. Then we derive an expression for the asymptotic behavior of the first two moments of T. Although these arguments are presented in the zero-inflated PS context, they appear to be more general.

Start by writing

$$\ell(\eta) = (1/n) \sum_{i=1}^{n} \log f^*(y_i|\eta) \quad \text{and} \quad \hat{\eta} = (\hat{\eta}_1, \hat{\eta}_2) = \arg\max_{\eta} \ell(\eta).$$

Letting D_j denote $\partial/\partial\eta_j$ for $j = 1, 2$ define

$$a_{ij} = D_i D_j \ell(\hat{\eta}) \quad \text{and} \quad a_{ijk} = D_i D_j D_k \ell(\hat{\eta}).$$

Under consistency conditions for the MLE and expected local supremum conditions on $f^*(\cdot|\eta)$ on a neighborhood around a fixed value η_0,

$$a_{ijk} \to E_{\boldsymbol{\eta_0}} D_i D_j D_k \log f^*_\eta(Y_1|\boldsymbol{\eta_0}), \quad a.e., P_{\boldsymbol{\eta_0}}.$$

The empirical Fisher information is $I(\hat{\eta}) = (I_{ij}(\hat{\eta}))$ and it is seen that $I_{ij}(\hat{\eta}) = -a_{ij}(\hat{\eta})$. To ensure $I(\hat{\eta})$ is well defined, assume that it is positive definite on a set S^* with $P_\eta(S^*) = 1 + o(1/\sqrt{n})$. Now the inverse is $I^{-1}(\hat{\eta}) = (I^{ij}(\hat{\eta}))$; it is needed to define the quantities that will appear in the asymptotic expansion for T. Set

$$m_i(\hat{\eta}) = \frac{I^{i1}(\hat{\eta})}{I^{11}(\hat{\eta})} \quad \text{and} \quad K^{ij}(\hat{\eta}) = I^{ij}(\hat{\eta}) - \frac{I^{i1}(\hat{\eta}) I^{j1}(\hat{\eta})}{I^{11}(\hat{\eta})}$$

and denote $\hat{\pi}_j(\hat{\eta}) = D_j \pi(\hat{\eta})$. Finally, the quantities that appear in the asymptotic expansion to order $O(1/\sqrt{n})$ are the second Hermite polynomial $J_2(t) = t^2 - 1$, and two correction terms

$$G_3(\hat{\eta}) = \frac{1}{6} a_{ijk} m_i m_j m_k (I^{11}(\hat{\eta}))^{3/2}$$

and

$$G_1(\pi,\hat{\eta}) = \frac{\hat{\pi}_j m_j}{\pi(\hat{\eta})}\sqrt{I^{11}} + \frac{1}{2}a_{ijk}K^{ij}m_k\sqrt{I^{11}} + \frac{1}{2}a_{ijk}m_i m_j m_k (I^{11}(\hat{\eta}))^{3/2},$$

using the convention that repeated indices indicate summation. To get the form of the result, let

$$W = \sqrt{\frac{n}{I^{11}(\hat{\eta})}}(\eta_{10} - \hat{\eta}_1) \text{ and } V = \sqrt{\frac{n}{I^{11}(\hat{\eta})}}(\eta_1 - \hat{\eta}_1).$$

Note that under $\eta_1 = \eta_{10}$, V is the same as W.

At last, from (2.3.19) in Datta and Mukerjee (2004), taking $\beta_1 = \beta_2 = 0$ we get

$$P(\eta_1 \leq \eta_{10}|\mathbf{Y}) = P(V \leq w|\mathbf{Y}) = \Phi(w) + n^{-1/2}\phi(w)\left\{G_1(\pi,\hat{\eta}) + G_3(\hat{\eta})J_2(w)\right\} + o_p(n^{-1/2}), \tag{21}$$

where w is the observed value of W and $\phi(\cdot)$ and $\Phi(\cdot)$ are standard normal pdf and cdf respectively. However, when η_{10} is true W is asymptotically $N(0,1)$. So, by the inverse probability integral transform $\Phi(W)$ is Uniform$[0,1]$ and the $1/\sqrt{n}$ terms ensure the required rate. Thus, since T is of the form $1 - P(\eta_1 \leq \eta_{10}|\mathbf{Y})$, T is asymptotically Uniform$[0,1]$ as well.

To derive expressions for the moments of T, write $G_1(\pi,\hat{\eta}) = \Gamma_1(\eta_0) + o_p(1)$, and $G_3(\hat{\eta}) = \Gamma(\eta_0) + o(1)$ and recognizing that J_2 is just a polynomial, it can be seen that there is an $H(\eta_0)$ so that (21) can be written as

$$P(\eta_1 \leq \eta_{10}|\mathbf{Y}) = P(V \leq w|\mathbf{Y}) = \Phi(w) + n^{-1/2}\phi(w)H(\eta_0) + o_p(n^{-1/2}), \tag{22}$$

and the expectation with respect to P_η can be taken on both sides. Using the result from that and applying Step 3 from Datta and Mukerjee (2004, p. 19) gives an expression for the frequentist probability from the middle term in (22):

$$P_{\eta_0}(V \leq w) = \Phi(w) + n^{-1/2}\phi(w)H^*(\eta_0) + o_p(n^{-1/2}), \tag{23}$$

where H^* is derived from H and the η_1 in the probability is $\eta_{1,0}$ the true value. Differentiating (23) it is possible to derive an approximation for the density of V, $f_V(w|\eta_0)$.

Finally, by using (21) and $f_W(w|\eta_0)$, it is possible to derive expressions for the first 2 moments of T, because they only depend on W. Doing so gives that they are $1/2$ and $1/12$, as

expected from the limiting uniform, However, given the $1/\sqrt{n}$ correction terms, it is possible to equate the expressions for the first 2 moments of $P(\eta_1 \leq \eta_{10}|\mathbf{Y})$ to the first two moments of a $Beta(\alpha, \beta)$ and thereby derive expressions for α and β. Obviously, the resulting $\hat{\alpha}$ and $\hat{\beta}$ must converge to 1, i.e., give the Uniform$[0, 1]$ in the limit of large n, but for finite n this provides a more refined approximation.

3.4 Credible Intervals

Although one can in principle find a Bayes estimate for p, under say squared error loss, and find its posterior variance, Bayes tests are based on posterior probabilities which in turn are based on the posterior density. These also lead to credible sets.

There are two main types of credible sets. The first is analogous to equitailed confidence intervals: $\alpha/2$ of the probability in the tails is clipped off and the upper and lower boundaries announced. The second is highest posterior density HPD, i.e., a set of the form $R(\pi_\alpha) = \{p : \pi(p|\mathbf{y}) \geq \pi_\alpha\}$, where π_α is the largest constant such that $P(p \in R(\pi_\alpha)|\mathbf{y}) \geq 1 - \alpha$. For symmetric unimodal densities the two types of interval are equal, and here HPD sets are obtained from a variant on the procedure used to get α-clipped credible sets. The basic idea is that if the credible interval or HPD set for p contains the value $p = 0$, then we may conclude that there is not enough evidence of zero inflation in the data.

Difficulties in the cases studied here arise because the posterior is not available in a convenient analytic form: The priors discussed in Sec. 3.2 do not yield tractable marginal posteriors for p by directly integrating θ out of the joint posterior density. Consequently, we find a Markov chain Monte Carlo (MCMC) estimate of the marginal posterior distribution and use it to find the $1 - \alpha$ credible and HPD sets. Thus, given a sample from the marginal posterior $\pi(p|\mathbf{Y} = \mathbf{y})$ it is easy to form a $1 - \alpha$ credible interval by choosing the $\alpha/2$ and $1 - \alpha/2$ sample quantiles. This can also be done using draws from the joint (p, θ) posterior density. The HPD set can be found by using the draws to estimate $\pi(p|\mathbf{Y} = \mathbf{y})$ by, say, $\hat{\pi}(p|\mathbf{Y} = \mathbf{y})$ and obtaining approximate HPD sets from $\hat{\pi}$.

Suppose that $\pi(p, \theta|\mathbf{y})$ and $\pi(p|\mathbf{y})$ are the joint and marginal posterior, respectively, so that

$$\pi(p|\mathbf{y}) = \int_0^\infty \pi(p, \theta|\mathbf{y}) d\theta.$$

In the case of ZIP model (9), with the conditional Jeffreys' prior (18a) for p, and Jeffreys' prior $\pi(\theta) = 1/\sqrt{\theta}$ for θ, the joint posterior is based on

$$\pi(p, \theta|\mathbf{y}) \propto \left\{p + (1-p)e^{-\theta}\right\}^{n_0 - 1/2} (1-p)^{n-n_0-1/2} e^{-\theta(n-n_0)} \theta^{s-1/2}. \tag{24}$$

Using (24), the goal is to estimate $\pi(p|\mathbf{y})$ from a joint sample of (p, θ) drawn from $\pi(p, \theta|\mathbf{y})$. Let $\{(p^{(i)}, \theta^{(i)}), i = 1, \ldots, B\}$ be an MCMC sample from $\pi(p, \theta|\mathbf{y})$ so that $\pi(p|\mathbf{y})$ can be estimated at $p = p^{(j)}$ by

$$\hat{\pi}(p^{(j)}|\mathbf{y}) = \frac{1}{B} \sum_{i=1}^{B} \pi(p^{(j)}, \theta^{(i)}|\mathbf{y}). \tag{25}$$

Since it is computationally difficult to draw samples $(p^{(i)}, \theta^{(i)})$ from (24), we use a reparametrized model by transforming $p^* = p + (1-p)e^{-\theta}$. Incidentally, note that the parameters p^* and θ result in an orthogonal reparameterization of the ZIP model. It can be checked that the Fisher information matrix is diagonal given by

$$I(p^*, \theta) = \text{diag}\left(\frac{1}{p^*(1-p^*)}, \frac{(1 - e^{-\theta} - \theta e^{-\theta})(1-p^*)}{\theta(1-e^{-\theta})^2}\right).$$

As a result of this reparameterization, the joint posterior for (p^*, θ) can be written as a product of their marginals. In fact this idea can be extended in general for all zero-inflated PS distributions.

Therefore, using the above fact the joint posterior distribution can be written as

$$\pi(p^*, \theta|\mathbf{y}) \propto (p^*)^{n_0 - 1/2} (1-p^*)^{n-n_0-1/2} \left(\frac{e^{-\theta}}{1-e^{-\theta}}\right)^{n-n_0} \theta^{s-1/2}. \tag{26}$$

From (26) it is seen that $(p^*|\mathbf{y})$ follows a Beta$(n_0 + 1/2, n - n_0 + 1/2)$, so it is easy to draw posterior samples of p^*. To draw samples of θ, we use rejection sampling with a suitably chosen gamma distribution as envelope. In this way it is possible to generate a representative sample $\{(p^{*(i)}, \theta^{(i)}), i = 1, \ldots, B\}$ from the joint posterior and using the relationship $p^* = p + (1-p)e^{-\theta}$, we get $\{(p^{(i)}, \theta^{(i)}), i = 1, \ldots, B\}$ where $p^{(i)} = (p^{*(i)} - e^{-\theta^{(i)}})/(1 - e^{-\theta^{(i)}})$.

Subsequently, using (25), we get an estimate of the marginal posterior density $\pi(p|\mathbf{y})$ at $p = p^{(j)}$.

A $100(1-\alpha)\%$ Bayesian credible interval for p is simply $(p_{(\alpha/2)}, p_{(1-\alpha/2)})$, where $p_{(k)}$ is the k-th quantile of $\{p^{(i)}, i = 1, \ldots, B\}$. To find the HPD interval, there are several methods and algorithms, see Chen and Shao (1999), and the references therein. Here, using the HPD set from the posterior sample $\{p^{(i)},\ i = 1, \ldots, B\}$, we find $\{\hat{\pi}(p^{(i)}|\mathbf{y}),\ i = 1, \ldots, B\}$ and set π_α to be the 100α-th percentile of $\hat{\pi}(p^{(i)}|\mathbf{y})$. This is adequate because the estimated posteriors are unimodal. Once we get π_α, we solve $\hat{\pi}(p|\mathbf{y}) = \pi_\alpha$ for cutoff values of p to find the lower and upper limits of the HPD interval.

3.5 Performance Comparison

In this section the performance of the test statistic (20) for $\mathcal{H}_0 : p = 0$ vs. $\mathcal{H}_1 : p > 0$ in the ZIP family is compared to the performance of the score test and the likelihood ratio test in simulations. Recall that the Bayes test is formed from (9), (18a), and $\pi(\theta) = 1/\sqrt{\theta}$. The score test is given by (11a) and we numerically obtain the likelihood ratio statistic. In the Table 1 below, we have computed the power of these three tests for several choices of n, p, θ for $\mathcal{H}_0$ vs $\mathcal{H}_1$ for level $\alpha = .05$; in Table 2 the power of the one-sided Bayes test is compared to the power of the two-sided score and LR tests as well, also at the $\alpha = .05$ level. Usually, two-sided tests have lower power than one-sided tests because the level forces mass to be assigned to both sides under the null rather than just to one side. We can see that the Bayes test performs somewhat better that the two-sided score test and the two-sided likelihood ratio test.

However, as in Sec. 3.5, the simulations for the Bayes test require MCMC because calculating $P(p > 0|\mathbf{Y})$ under the ZIP model is not straightforward. Indeed, in general, it is not possible to provide a procedure that will work for any zero-inflated PS distribution.

Nevertheless, for the ZIP model, (26) and (12) imply (20) can be written as

$$\begin{aligned} P(p>0|\mathbf{Y}) &= \frac{\int_0^\infty \int_0^1 I_{[p^*>e^{-\theta}]} \left(\frac{\theta}{1-e^{-\theta}}\right)^{n-n_0} p^{*n_0}(1-p^*)^{n-n_0} e^{-\theta(n-n_0)} \theta^{s-(n-n_0)} \pi^*(p^*,\theta) dp^* d\theta}{\int_0^\infty \int_0^1 \left(\frac{\theta}{1-e^{-\theta}}\right)^{n-n_0} p^{*n_0}(1-p^*)^{n-n_0} e^{-\theta(n-n_0)} \theta^{s-(n-n_0)} \pi^*(p^*,\theta) dp^* d\theta} \\ &= \frac{E^g\left[I_{[p^*>e^{-\theta}]}\left(\frac{\theta}{1-e^{-\theta}}\right)^{n-n_0} \pi^*(p^*,\theta)\right]}{E^g\left[\left(\frac{\theta}{1-e^{-\theta}}\right)^{n-n_0} \pi^*(p^*,\theta)\right]} \end{aligned} \tag{27}$$

where $\pi^*(p^*,\theta)$ is the joint prior of (p^*,θ) and

$$g(p^*,\theta) \propto p^{*n_0}(1-p^*)^{n-n_0} e^{-\theta(n-n_0)} \theta^{s-(n-n_0)}.$$

So, we draw a random sample $\{(p^{*(i)},\theta^{(i)}), i=1,\ldots,B\}$ where $p^{*(i)} \sim Beta(n_0+1, n-n_0+1)$ and $\theta^{(i)} \sim gamma(n-n_0, s-(n-n_0)+1))$ and calculate

$$P(p>0|\mathbf{Y}) = \frac{\sum_{i=1}^B I_{[p^{*(i)}>e^{-\theta^{(i)}}]} \pi^*(p^{*(i)},\theta^{(i)}) \left(\frac{\theta^{(i)}}{1-e^{-\theta^{(i)}}}\right)^{n-n_0}}{\sum_{i=1}^B \pi^*(p^{*(i)},\theta^{(i)}) \left(\frac{\theta^{(i)}}{1-e^{-\theta^{(i)}}}\right)^{n-n_0}}.$$

In Tables 1 and 2, the test statistic $T(\mathbf{Y})$ from (20) is compared to the cut-off point found from the asymptotic distribution of $T(\mathbf{Y})$ under H_0 i.e. we use upper α point on $uniform(0,1)$ as described in Sec. 3.3.2. All the simulations are based on 10,000 replications with $B = 10,000$ MCMC samples in each replication.

We comment that for the case of a zero-inflated geometric distribution, the procedure is a little easier: It is just a matter of drawing samples from two different Beta distributions with parameters based on the sample. So, it is easy to find an MCMC estimate of the test statistic.

Table 1 shows that for the one-sided test, all 3 tests have roughly the same level when $p = 0$. In fairness, the level for the score and LR tests is a little lower leading to lower power against alternatives. However, looking at how the power of all three tests indicated rises as p increases, it is clear that for mid-sized p and smallish θ the Bayes test has noticeably higher power, especially for small n. In fact, a good test is most important on this range because it

is hard to distinguish zero inflation from its absence when θ is small or moderate, p ranges from small to mid-sized values, and n is not large.

Table 2 is similar to Table 1 but the score and LR tests are two-sided. It shows that the same properties hold, but a little more strongly. This may be attributed to the fact that the Bayes test uses an extended parameter space, allowing some mass to represent zero deflation.

3.6 Data Analysis

To demonstrate the efficacy of our technique, we apply it to test for presence of zero inflation in three widely used datasets. We also give comparative values from other techniques. In general, the results from these techniques corroborate each other so the fact they are based on different principles lends credence to the conclusions.

The first dataset that we look at is the Urinary Tract Infection (UTI) data used in Broek (1995) who used a score test to detect zero inflation in a Poisson model. The data are collected from 98 HIV-infected men, attending the department of internal medicine at the Utrecht University hospital. The number of times they had an urinary tract infection was recorded as Y. The data are recorded in Table 3.3. Merely by looking at the data it is clear that zero inflation is present. Also, $P(p > 0|\mathbf{y}) = .999$ while the observed value of the score statistic is 15.34 giving a $p-$value 0.0001 and a Bayes factor for testing $\mathcal{H}_0 : p = 0$ vs. $\mathcal{H}_1 : p > 0$ of $B_{10} = 223.13$. The details of computation of Bayes factor will be reported elsewhere. This is again strong evidence in favor of the alternative, which is no surprise.

The next data set we consider is the Terrorism data from Conigliani, Castro and O'Hagan (2000). Table 3.4 shows the data concerning the number of incidents of international terrorism per month (Y) in the United States between 1968 and 1974. This data set is not obviously zero-inflated or deflated; it is an open question. Conigliani et al (2000) find a Fractional Bayes factor for this data set of .0089; we find a Bayes factor of $B_{10} = 0.28$. In fact, $P(p > 0|\mathbf{y}) = 0.507$, an indeterminate value. The observed value of the score statistic is 0.04 and a $p-$value 0.83. All three assessments agree that there is no evidence of zero inflation.

The third data set we analyzed is the Cholera data first analyzed by McKendrick (1926). Table 3.5 shows the number of patients per household suffering from cholera in a village in India in 1920's. Again, looking at the data strongly suggests zero inflation. Under our method, $P(p > 0|\mathbf{y}) = .9999$ and the Bayes factor is $B_{10} = 238090$, very strong evidence for zero inflation . The observed value of the score statistic is 30.56, effectively giving a $p-$value of 0.

Although tests are useful for quantifying degree of belief, there is no substitute for looking at the posterior distributions directly. Figure 1 shows plots of the marginal posteriors for p resulting from applying the ZIP model to each of the three data sets. All three posteriors are unimodal and appear roughly symmetric. The location of the mode, and the spread around it determine the most credible values of p. For the UTI and Cholera data the determination is clear: Substantial zero inflation is present. For the Terror data, looking from the graph the evidence for zero inflation is equivocal. The slight asymmetry throws off the eye making it impossible to tell whether $p = 0$ is reasonable. In fact, the test shows it is, but this would be open to question from merely looking at the diagram.

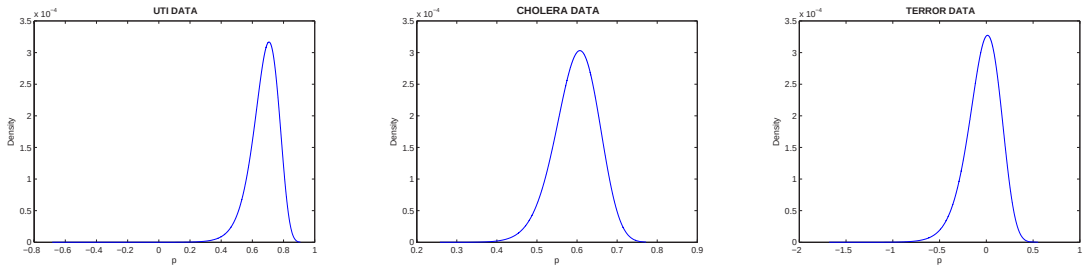

Figure 3.1: Estimated posterior densities of p

Table 3.6 gives 95% Bayesian credible and HPD intervals for the three data sets under consideration. Also the marginal posterior distributions of p are given in Figure 3.1. From the intervals and the figures as well, it is evident that there is noticeable amount of zero inflation in Cholera data and UTI data because the interval of concentration of the posterior distributions does not include zero whereas for Terror data the posterior distribution of p is centered around zero and the interval contains zero, signifying the absence of zero inflation in the data.

Finally, for the sake of completeness, Table 6 gives 95% credible intervals and HPD sets calculated from the posteriors. It is seen that for the Cholera and UTI data that 0 is not in the intervals. This is consistent with the presence of zero inflation. For the Terror data, 0 is in the interval. Even though the interval is wide, much of it includes negative values. So, it is not a surprise that zero inflation is not indicated by the test. Note that the equitailed credible and HPD sets are close for the cholera data indicating symmetry, but for the other two data sets the difference in the intervals suggests some left skewing, more for Terror than for Cholera.

3.7 Conclusions

Overall, this article gives a general Bayesian setup for testing for zero inflation in PS distributions that can be compared to existing likelihood based methods occurring in frequentist treatments. The basic idea is to extend the parameter space to include a small range of negative values for the weight on zero inflation. Thus, the null hypothesis $\mathcal{H}_0 : p = 0$ becomes an interior point of the parameter space and a standard Bayesian approach is feasible.

Our simulations suggest the Bayesian test has power as high as, or slightly higher than the standard frequentist tests, even when objective priors that are somewhat unfavorable to the hypothesis $\mathcal{H}_0 : p = 0$ are used to automate the procedure. Interval estimation for p proceeds similarly, using the extended parameter space.

The technique of extending the parameter space applies generally to Bayes, and potentially frequentist, testing for zero inflation with count data, but obviously can apply to many situations where two distributions are mixed and one wants to know whether one component can be set to zero. In fact, the asymptotics for this test require only generic regularity conditions; they do not rely on specific forms of the likelihood such as exponential families. A further test of the method, aside from applying it to more general mixtures, would be extending it to a class of regression problems by including covariates.

Acknowledgments

The authors would like to thank Prof. P. K. Sen and Prof. J. K. Ghosh for their valuable comments. The authors acknowledges Prof. Xiao-Li Meng for sharing the Cholera data. Part of this research was conducted while G. S. Datta was a SAMSI/Duke University visiting fellow and B. Clarke was a SAMSI visitor. This research was partially supported by NSF grant SES-0241651. A. Bhattacharya received student support from this grant.

3.8 References

[1] Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *J. R. Statist. Soc. B.* vol 41 pp 113-147.

[2] Broek, J. V. D. (1995). A score test for zero inflation on a Poisson distribution. *Biometrics.* vol 51. pp 738-743.

[3] Chen, M.-H. and Shao, Q.-M. (1999). Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics.* vol 8. pp 69-92.

[4] Clarke, B. and Barron, A. (1994) Jeffreys prior is asymptotically least favorable under entropy loss. *Journal of Statistical Planning and Inference.* vol. 41, pp. 37-60.

[5] Cochran, W. G. (1954). Some methods of strengthening χ^2 tests. *Biometrics.* vol 10. pp 417-451.

[6] Conigliani, C., Castro, J. I. and O'Hagan, A. (2000). Bayesian Assessment of Goodness of Fit against Nonparametric Alternatives. *Canadian Journal of Statistics.* vol 28. pp 327-342.

[7] Datta, G. S. and Mukerjee, R. (2004). Probability Matching Priors: Higher Order Asymptotics. *Springer-Verlag.*

[8] Deng, D and Paul, S. R. (2000). Score test for zero inflation in generalized linear models. *Canadian Journal of Statistics.* vol 28. pp 563-570.

[9] Deng, D and Paul, S. R. (2005). Score tests for zero-inflation and over-dispersion in generalized linear models. *Statistica Sinica.* vol 15. pp 257-276.

[10] Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association.* vol 81. pp 709-721.

[11] El-Shaarawi, A. H. (1985). Some goodness-of-fit methods for the Poisson plus added zeros distribution. *Applied and Environmental Microbiology.* vol 49. pp 1304-1306.

[12] Ghosh, S. K., Mukhopadhyay, P. and Lu, J. C. (2006). Bayesian analysis of zero-inflated regression models. *Journal of Statistical Planning and Inference.* vol 136. pp 1360-1375.

[13] Hall, D. B. (2000). Zero-inflated Poisson and binomial regression with random effects: A case study. *Biometrics.* vol 56. pp 1030-1039.

[14] Johnson, N. L., Kotz, S., and Kemp, A. W. (1992). *Univariate Discrete Distributions.* Second edition. John Wiley & Sons Inc.

[15] Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics.* vol 34. pp 1-14.

[16] McKendrick, A. G. (1926). Application of mathematics to medical problems. *Proc. Edin. Math. Soc.* vol 44. pp 98-130.

[17] Rissanen, J. (1983). A universal prior for integers and estimation by minimum description length. *Ann. Statist.* vol. 11 no. 2, pp 416-431.

[18] Rao, C. R. and Chakravarti, I. M. (1956). Some small sample tests of significance for a Poisson distribution. *Biometrics.* vol 12. pp 264-282.

[19] Self, S. G. and Liang, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association.* vol 82. pp 605-610.

[20] Silvapulle, M. S. and Silvapulle, P. (1995). A score test against one-sided alternatives. *Journal of the American Statistical Association.* vol 90. no 429. pp 342-349.

	p	0.00			0.10			0.30			0.40		
θ	n	Score	Bayes	LR	Score	Bayes	LR	Score	Bayes	LR	Score	Bayes	LR
0.5	20	0.0494	0.0449	0.0465	0.0652	0.0682	0.0636	0.1111	0.1048	0.1034	0.1436	0.1344	0.1179
	50	0.0458	0.0427	0.0421	0.0781	0.0758	0.0717	0.1800	0.1585	0.1544	0.2509	0.2124	0.2089
	100	0.0496	0.0467	0.0457	0.0960	0.0899	0.0807	0.2838	0.2616	0.2628	0.3755	0.3634	0.3450
1.0	20	0.0402	0.0487	0.0356	0.0829	0.0941*	0.0822	0.2316	0.2474*	0.2276	0.3181	0.3227*	0.3107
	50	0.0399	0.0485	0.0404	0.1228	0.1332*	0.1215	0.4334	0.4343*	0.4166	0.5850	0.5815	0.5659
	100	0.0447	0.0474	0.0482	0.1809	0.1821	0.1875	0.6698	0.6714	0.6804	0.8401	0.8410	0.8431
1.5	20	0.0421	0.0529	0.0402	0.1226	0.1430*	0.1161	0.3892	0.4199*	0.3867	0.5442	0.5644*	0.5371
	50	0.0397	0.0472	0.0426	0.2139	0.2251*	0.2116	0.7295	0.7468*	0.7394	0.8843	0.8948*	0.8882
	100	0.0445	0.0458	0.0458	0.3454	0.3107	0.3514	0.9509	0.9361	0.9528	0.9921	0.9906	0.9933
2.0	20	0.0463	0.0517	0.0350	0.1939	0.2129*	0.1750	0.6151	0.6490*	0.5997	0.7634	0.8013*	0.7580
	50	0.0531	0.0530	0.0453	0.3453	0.3629*	0.3464	0.9356	0.9354	0.9350	0.9876	0.9875	0.9860
	100	0.0438	0.0530	0.0415	0.5770	0.4842	0.5565	0.9977	0.9947	0.9979	1.0000	0.9999	1.0000

Table 3.1: The entries are the powers for the Bayesian, one-sided score, and one-sided LR tests, with 10,000 simulations. The asterisks indicate when the values where the Bayes test has highest power. They are clustered around small to moderate p, small θ and small to moderate n.

	p	0.00			0.10			0.30			0.40		
θ	n	Score	Bayes	LR	Score	Bayes	LR	Score	Bayes	LR	Score	Bayes	LR
0.5	20	0.0451	0.0449	0.0605	0.0427	0.0682	0.0518	0.0653	0.1048	0.0573	0.0870	0.1344	0.0658
	50	0.0456	0.0427	0.0503	0.0553	0.0758	0.0563	0.1218	0.1585	0.1057	0.1805	0.2124	0.1355
	100	0.0510	0.0467	0.0512	0.0656	0.0899	0.0583	0.1854	0.2616	0.1744	0.2766	0.3634	0.2484
1.0	20	0.0483	0.0487	0.0578	0.0568	0.0941	0.0619	0.1423	0.2474	0.1430	0.2025	0.3227	0.1983
	50	0.0491	0.0485	0.0513	0.0747	0.1332	0.0778	0.3025	0.4343	0.2960	0.4427	0.5815	0.4303
	100	0.0517	0.0474	0.0502	0.1174	0.1821	0.1148	0.5706	0.6714	0.5415	0.7666	0.8410	0.7390
1.5	20	0.0466	0.0529	0.0571	0.0810	0.1430	0.0738	0.2804	0.4199	0.2674	0.4107	0.5644	0.4092
	50	0.0509	0.0472	0.0505	0.1398	0.2251	0.1308	0.6184	0.7468	0.6120	0.8061	0.8948	0.8092
	100	0.0488	0.0458	0.0543	0.2436	0.3107	0.2355	0.9133	0.9361	0.9082	0.9827	0.9906	0.9819
2.0	20	0.0408	0.0517	0.0705	0.1280	0.2129	0.1131	0.5010	0.6490	0.4711	0.6700	0.8013	0.6443
	50	0.0491	0.0530	0.0566	0.2566	0.3629	0.2284	0.8896	0.9354	0.8804	0.9749	0.9875	0.9729
	100	0.0474	0.0530	0.0451	0.4508	0.4842	0.4401	0.9949	0.9947	0.9938	1.0000	0.9999	0.9997

Table 3.2: The entries are the powers for the Bayesian test and the two-sided score and two-sided LR tests, with 10,000 simulations. Putting asterisks in this table gives the same pattern as in Table 1, but stronger.

Y	0	1	2	3	Total
Frequency	81	9	7	1	98

Table 3.3: UTI Data

Y	0	1	2	3	4	Total
Frequency	38	26	8	2	1	75

Table 3.4: Terror Data

Y	0	1	2	3	4	Total
Frequency	168	32	16	6	1	223

Table 3.5: Cholera Data

Data	Credible Interval	HPD Interval
Terror	(-0.6735, 0.2945)	(-0.5560, 0.3654)
Cholera	(0.4619, 0.7095)	(0.4700, 0.7144)
UTI	(0.3433, 0.8240)	(0.4271, 0.8561)

Table 3.6: Bayesian credible and HPD Intervals

Chapter 4

Hairline Fracture Detection and Target Pattern Generation using MRF and Bayesian Image Restoration[1]

[1] A. S. Chowdhury, A. Bhattacharya, S. M. Bhandarkar, Gauri S. Datta and J. C. Yu. Submitted to *IEEE Transactions on Madical Imaging,* 01/11/2007.

4.1 Introduction

Craniofacial fractures, especially mandibular fractures, are frequently encountered in modern society with the major causes being gunshot wounds, motor vehicle accidents and sports-related injuries (see King et. al. (2004)). Mandibular fractures are observed to possess certain distinct patterns in X-ray or Computer Tomography (CT) images (see Ogundare et. al.(2004)). In some cases, the fractures are observed to be hairline or minor in nature. By the terms *hairline fracture* or *minor fracture* we respectively refer to those situations where the broken bone fragments are not visibly out of alignment or have incurred very little relative displacement. The presence of noise makes the detection and subsequent visualization of such types of fractures in X-ray or CT images a very challenging task. In case of a *major fracture*, i.e., fractures where the broken fragments are clearly displaced relative to each other, surgical intervention is almost mandatory. However, in the case of a *hairline/minor fracture*, the decision regarding surgical intervention is less clear as a surgeon can rely on natural bone healing. In this paper, we propose a Markov Random Field (MRF)- Maximum A Posteriori probability (MAP) based scheme for (a) mandibular hairline fracture detection in the presence of noise and (b) target pattern generation for hairline/minor fractures. The target pattern depicts how a jaw with a hairline fracture would appear if allowed to heal naturally without explicit surgical intervention. This *in silico* reconstructed pattern may have potential prognostic significance because surgeons can decide if open surgical reduction and fixation is necessary or the fractures can be managed by allowing them to heal spontaneously based on what the naturally healed mandible will look like. For a major single fracture or multiple fractures that are visibly out of alignment, the proposed technique can be used as smoothing technique at the sites where the opposable bone fragments have been brought into coarse registration. This would result in the generation of a smooth or continuous 3D model of the reconstructed jaw where the individual bone fragments are fused together during free-form solid fabrication. Note that a 3D model is essential for pre-adaptation of the prostheses needed to stabilize the fractures and for performing the necessary biomechanical analysis to

validate the virtual reconstruction of the fractured mandible or craniofacial skeleton along with the affixed prostheses.

From the standpoint of computer vision and pattern recognition research, the problems of hairline fracture detection and target pattern generation are of significant interest. Conventional corner detection techniques for determining points of surface discontinuity, such as the Harris detector, Harris and Stephens (1988), do not perform well because of the pronounced intensity inhomogeneity and noise present in X-ray or CT images. In the proposed scheme, hairline fracture detection and target pattern generation are simultaneously achieved by a process of implicit image restoration. The visual comparison of an X-ray or CT image of a mandible containing a hairline/minor fracture with that of an unbroken (intact) mandible reveals changes in pixel intensity only in the vicinity of the fracture site. This results in the formulation of an image restoration problem with a *partially unknown local* degradation. This is in sharp contrast to the more conventional image restoration problem with a *known global* degradation as outlined by Geman and Geman (1984) in their classical paper.

The proposed scheme takes as input a stack of 2D CT image slices of a human mandible with a hairline fracture. A two-step approach is taken where, in the first step, a hairline fracture is approximately localized within a block of pixels by exploiting the (approximate) bilateral symmetry of the human mandible and by using statistical correlation of the pixel intensities as a measure of intensity mismatch. An MRF modeling approach coupled with MAP estimation, is employed within the localized pixel blocks in the second step to detect the fracture and generate the target pattern with high accuracy.

4.2 Literature Review

In this section, we first mention some of the existing applications of MRF modeling and Bayesian restoration in the area of medical image analysis. Existing approaches for fracture detection in X-ray/CT images are described next. The theoretical and practical importance of the proposed scheme are discussed subsequently. The two classical applications of MRF-MAP paradigm for Bayesian image restoration can be found in Geman and Geman (1984)

and Besag (1986). MRF-based techniques have gained popularity in the past decade for solving various biomedical imaging problems. Some interesting applications include MRF-based image reconstruction for SPECT images by Lee et al. where they used an MRF for a spatial smoothness regularizer in Lee, Rangarajan and Gindi (1995); 3D MRF-based volumetric object reconstruction in Magnetic Resonance (MR) images by Choi et. al. (1997); multiresolution MRF for tumor detection in mammograms by Zheng and Chan (2001); contextual clustering for analysis of functional MR (fMR) images using an MRF prior by Salli et. al. (2001); segmentation of MR brain image data using MRF and a Gibbs prior by Chen and Metaxas (2004); matching of digital mammograms using Markovian (and variational) approaches by Richard (2005); penalized least square technique for Borehole Tomography using an MRF and a Gibbs prior by Popa and Zdunek (2005); contour detection of human kidneys using MRF and active contours by Martin-Fernandez and Alberola-Lopez (2005) and reconstruction of MR images from raw Fourier data where an MRF is used to enforce spatial smoothness by Raj et. al. (2006).

Existing published literature on fracture detection in X-ray/CT images describes use of various approaches such as exploitation of anatomical knowledge combined with a *divide-and-conquer* approach by Ozanian and Philips (2000); texture analysis by Yap et. al. (2004); an adaptive interface agent with neural networks by Syiam et al. (2004); an affine morphological scale space along with iterative peak detection and modified Hough transform by Donnelley and Knowles (2005); probabilistic combination of various classifiers by Lum et. al. (2005) and an active contour modeling coupled with shape constraints by Jia and Jiang (2006). An example of semi-automated fracture detection scheme for the class of well-displaced fractures can be found in Chowdhury et. al. (2006).

To the best of our knowledge, the classical MRF-MAP paradigm has not been previously applied to the problem of fracture detection in medical images (X-ray or CT) in general. From a biomedical imaging perspective, the proposed two-phase MRF-MAP based scheme addresses some important clinical problems such as hairline fracture detection and target pattern generation with accompanied prognosis of fracture healing. The proposed MRF-MAP paradigm is shown to handle input noise in an explicit and efficient manner. The

approximate localization of fractures within pixel blocks in the first phase is shown to result in significant computational savings in the second phase since the MRF modeling and MAP estimation using Gibbs sampling, is restricted only to those pixel blocks in the CT image stack which are known to contain potential fractures. From a theoretical standpoint, the proposed scheme extends the conventional MRF-MAP paradigm to deal with a *partially unknown local* degradation of image pixel intensities at the fracture site. This is in contrast to the conventional MRF-MAP paradigm which incorporates a *global* and *known* deformation model. The proposed scheme for hairline fracture detection is also designed to perform an implicit restoration of the broken mandible at the fracture sites, thus offering the surgeon a prognostic view of the bone healing process. A preliminary version of the present work appears in Chowdhury, Bhattacharya, Bhandarkar, Datta, Yu and Figueroa (2007).

The rest of the paper is organized as follows: in Section 4.3 we discuss the method used for fracture localization; in Section 4.4 we describe the mathematical foundations of the model; Section 4.5 contains the description and analysis of the experimental results; Section 4.6 concludes the paper and outlines the directions for future research.

4.3 Fracture Localization

The input to the proposed fracture detection scheme is a stack of 2D CT image slices of the human mandible with a hairline fracture. Each 2D CT image slice is assumed to be parallel to the xy plane whereas the z axis is assumed to be the axial direction along which the CT image slices are acquired. The CT image stack is divided into a number of pixel blocks. The fracture localization phase which constitutes the first phase of the proposed two-phase scheme, is decomposed into three sub-phases as follows:

4.3.1 *Localization of the Mandible*

Different anatomical structures within the human body are known to possess different types of symmetry Prima et. al. (2002). In the context of our problem, we exploit the (approximate) bilateral symmetry exhibited by the human mandible. In the case of a hairline/minor fracture, the bilateral symmetry of the mandible is still preserved to a great extent despite

the presence of the fracture. The general equation of a 3D plane (of bilateral symmetry) is given by:

$$Ax + By + Cz = D \tag{4.1}$$

For an axial CT scan of the human mandible we assumed B and C to be approximately equal to zero and the mandibular cross-section to be approximately centered within each CT image slice of width W. Thus, the equation of approximate plane of bilateral symmetry reduces to:

$$x = W/2 \tag{4.2}$$

Now, for every incident pixel g_i with coordinates (x, y, z) in the left half of the mandible with a hairline fracture, a bilaterally symmetric pixel g_i^R with coordinates (x^R, y^R, z^R) can be determined as:

$$x^R = W - x,\ y^R = y,\ z^R = z \tag{4.3}$$

Two heuristics are exploited to reduce the search space for coarse fracture localization. These heuristics along with their justifications (based on domain knowledge) are given below:

1. Since mandibles are essentially bone structures that typically exhibit higher intensity (i.e. Hounsfield unit) values in CT images, we seek pairs of pixel blocks with high average intensity. This helps to remove pixel blocks containing artifacts and/or large amounts of soft tissue from further consideration.

2. The mandible is typically larger in size compared to other bones in the CT images of the craniofacial skeleton. Since we are primarily interested in detecting mandibular fractures, we perform a second round of filtering by applying the connected component labeling algorithm, at the pixel block level rather than at the level of individual pixels, and eliminating components which span only a small number of pixel blocks.

By using the above heuristics, we basically retain only a few (say q) pixel blocks which are deemed to constitute solely the fractured mandible.

4.3.2 *Identification of the Appropriate Symmetric Block(s)*

Having localized the mandible in the CT image, the next goal is to localize the fracture within it. This is done by taking a block from the left half of the image, all of whose pixels have x coordinate values: $0 \leq x \leq W/2$, and a corresponding bilaterally symmetric block in the right half of the image, all of whose pixels have x coordinate values $W/2 \geq x \geq W$, and computing the statistical correlation between the two. The correlation coefficient between a typical incident block g, with individual pixels g_i, and its bilaterally symmetric counterpart g^R, with corresponding pixels g_i^R is given by:

$$r(g, g^R) = \frac{1}{(n-1)} \sum_{i=1}^{n} \frac{(g_i - \bar{g})}{(s_g)} \frac{(g_i^R - \bar{g}^R)}{(s_g^R)} \tag{4.4}$$

where $\bar{g}$ and $\bar{g}^R$ denote the mean, and s_g and s_g^R denote the standard deviation of the pixels within the blocks g and g^R respectively.

Having obtained a value of $r(g, g^R)$ for each pair of pixel blocks (g, g^R), the pixel block pairs are sorted in increasing order of their $r(g, g^R)$ values. Note that the pixel block within the intact (unbroken) half of the mandible will have more pixels with higher intensity values (due to the presence of more bone material) compared to its bilaterally symmetric counterpart which contains the hairline fracture (resulting in some loss of bone material). Thus, the underlying rationale is that pairs of pixel blocks which potentially contain fractures should exhibit a higher intensity mismatch and hence lower correlation values. The user can then choose the best k out of q pixel blocks as sites containing potential hairline fractures. The above technique for coarse fracture localization provides the following two advantages:

1. It achieves computational efficiency by effectively reducing the image size over which the proposed MRF-MAP scheme coupled with Gibbs sampling is to be applied. Thus, instead of applying the proposed MRF-MAP scheme over the entire CT image slice, we do so only over the select k pixel blocks in each CT image slice.

2. It renders the prior shape information in each CT image slice more relevant and more accurate. Instead of determining two quadratic polynomials to describe the inner contour and outer contour of the entire mandible, we now only need to determine the

quadratic polynomials that describe the inner and outer contours of the portion of the mandible that appears within the selected set of k pixel blocks.

4.3.3 *Identification of the Fracture Site*

After corresponding pairs of blocks are determined on both the halves of the approximate plane of symmetry in the CT scans, we identify the fracture site using the following three clinical observations:

1. In some CT scans, the fracture sites are marked by the presence of pixel regions with distinctly low intensity values (as compared to those of the bone fragments and soft tissue). When the soft tissue integrity is violated (due to the impact of the injury), air can enter the deeper tissue planes and form these low intensity regions, known as *emphysema*. The reason *emphysema* appears as dark (low-intensity) spots in the CT scans is due to its low electron density (see Giannoudis and Dinopoulos (2005) and Vodovotz et. al. (2006)).

2. The fracture sites in most of the CT scans are accompanied by *swelling of the soft tissue* in its vicinity. When tissue injury occurs, the resulting inflammation increases the capillary permeability in the region. This results in transudate and exudate formation, which is manifested in the form of soft tissue swelling (see Giannoudis and Dinopoulos (2005) and Vodovotz et. al. (2006)).

3. The hairline/minor fractures are essentially marked by loss of bone (i.e. high intensity) pixels. Thus, given the fact that the localized fracture blocks are (approximately) bilaterally symmetrical, the fractured half should typically contain a lower number of bone pixels within the localized area (from phase-I).

We first check for *emphysema* within the localized blocks in the CT slices. Obviously, the image half which contains the *emphysema* is designated as the one containing the hairline/minor fracture. In cases, where *emphysema* is absent, we look for *soft tissue swelling*. The image half in which the *soft tissue swelling* is present, is deemed to contain the fracture.

In cases, where *emphysema* is absent and *soft tissue swelling* is not particularly prominent (an extremely unlikely case from a clinical viewpoint; we have found only a single such case in our dataset), we count the bone pixels within the localized blocks in the two image halves and conclude that the image half containing the fewer number of bone pixels is the one containing the fracture.

4.4 Hierarchical Bayesian Restoration Framework

In this section, we first describe the statistical model adopted for explaining the pixel intensity distribution of the image within a Bayesian framework and then illustrate the manner in which we model the degradation matrix stochastically. We also prove that in the proposed Bayesian restoration approach the posterior energy remains Gibbsian (in the manner defined by Geman and Geman (1984)). Although we describe the model in terms of the entire CT image, however we apply our restoration procedure effectively to the selected localized block.

4.4.1 *Statistical Model*

Suppose we have an image consisting $m \times n$ pixels. Let $p = m \times n$. Based on the formulation detailed in Geman and Geman (1984), the pixel intensities in the image can be expressed as

$$g = \Phi(f) + \epsilon \tag{4.5}$$

where g, f and ϵ are $p \times 1$ vectors such that g represents the vector of all observed image intensities, f represents the vector of intensities corresponding to the true image and ϵ is zero-mean random Gaussian noise

$$\epsilon \sim \mathcal{N}(0, \sigma^2 I_p). \tag{4.6}$$

where I_p is the p-th order identity matrix. The function $\Phi(\cdot)$ in equation (4.5) denotes a known degradation (or perturbation) function. Furthermore, we assume the that true pixel intensity f has a known prior distribution. The conditional autoregressive model (CAR) is one of several typical prior distributions used extensively in the domain of image processing.

The CAR model also ensures the Markovian property of dependence of the intensity value of any pixel on those of the pixels in its neighborhood (Molina (1994)). Therefore,

$$\begin{aligned} p_{\sigma^2}(g|f) &\propto (\sigma^2)^{-p/2} \exp\left\{-\tfrac{1}{2\sigma^2}||g-\Phi(f)||^2\right\} \\ p(f|\tau^2) &\propto (\tau^2)^{-p/2} \exp\left\{-\tfrac{1}{2\tau^2} f^T(I_p-\gamma N)f\right\} \end{aligned} \tag{4.7}$$

where N is the neighborhood matrix given by $N=[n_{ij}]$ as $n_{ii}=0\ \forall\ i$ and

$$n_{ij} = \begin{cases} 1 & \text{if } \ i \text{ and } j \text{ are neighbors} \\ 0 & \text{otherwise} \end{cases} \tag{4.8}$$

The value of γ is appropriately chosen to avoid singularity of the matrix $(I_p-\gamma N)$. Also the prior distributions on σ^2 and τ^2 are chosen as

$$\begin{aligned} p(\sigma^2) &\propto (\sigma^2)^{-(\nu+1)}\, e^{-\beta/\sigma^2} \\ p(\tau^2) &\propto (\tau^2)^{-(\kappa+1)}\, e^{-\delta/\tau^2}. \end{aligned} \tag{4.9}$$

In other words, $\sigma^2 \sim$ inverse-gamma (ν,β) and $\tau^2 \sim$ inverse-gamma (κ,δ). Under this formulation, the posterior distribution of f given the observed data g can be shown to be Gibbsian on account of conjugacy under linear degradation. Before describing the proposed MRF-MAP scheme in detail, it is important to formulate a precise definition of the term *known degradation* in the present context. In image processing, generally, the term *known degradation* commonly refers to a mathematically known degradation (or blurring) function which is applied to the *entire* true image and which in conjunction with the (typically additive) noise yields the observed data.

For fracture detection, we assume that the image of the fractured mandible is a degraded version of some true (possibly hypothetical) intact mandible. Consequently, the degradation function needs to be formulated in a manner such that if it is applied to the entire true image, i.e., the CT image of the intact mandible, the resulting image should display a hairline fracture at the desired site while retaining the pixel intensity values of the true image elsewhere. A minor/hairline fracture denotes a loss of bone mass and hence can be modeled as a decrease in the Hounsfield unit (image intensity) at the fracture site. Thus,

from equation (4.5), a simple formulation of the degradation function could be

$$g = Af + \epsilon \tag{4.10}$$

where A is the degradation matrix of order $p \times p$ consisting of non-zero elements only along the diagonal. For the i-th pixel,

$$g_i = a_i f_i + \epsilon_i \tag{4.11}$$

where

$$a_i = \begin{cases} \alpha_i & \text{if} \quad i \text{ is a fracture site} \\ \\ 1 & \text{otherwise} \end{cases} \tag{4.12}$$

for some $\alpha_i \in (0, 1)$. The prior distribution of α_i is assumed to be Gaussian:

$$\alpha_i \sim \mathcal{N}(\alpha_{i0}, \eta^2). \tag{4.13}$$

The choice of parameters α_{i0} and η is made in such a way that the support of the distribution becomes effectively $(0, 1)$. Later in this section, we provide details on the choice of α_{i0}.

Lemma 1: Under the model characterized by equations (4.5)-(4.13), for each fixed value of g, the posterior probability $p(f|g)$ is a Gibbs distribution with an associated energy function given by:

$$U(f|g) = \frac{1}{2\sigma^2}||g - Af||^2 + \frac{1}{2\tau^2} f^T (I_p - \gamma N) f$$

For a detailed proof of the above lemma, the interested reader is referred to Geman and Geman (1984). As a special case of the Lemma 1, we formulate the following claim. Before we formulate the claim, we introduce the parameter f_{i-} to denote the neighborhood of f_i.

Claim 2: Based on the MRF formulation in equation (4.6) and equations (4.10) – (4.12), if we assume $E(f_i|f_{-i}) = \mu(f_{i-})$ then the posterior distribution of f_i given g_i can be shown to be (see Gelman et. al. (2004) for details)

$$f_i|g_i, f_{i-} \sim \mathcal{N}\left(\frac{\frac{a_i g_i}{\sigma^2} + \frac{\mu(f_{i-})}{\tau^2}}{\frac{a_i^2}{\sigma^2} + \frac{1}{\tau^2}}, \frac{1}{\frac{a_i^2}{\sigma^2} + \frac{1}{\tau^2}} \right).$$

Proof The posterior distribution of $f_i|g_i, f_{i-}$ can be expressed as

$$\begin{aligned} p(f_i|g_i, f_{i-}) &\propto \exp\left\{-\frac{1}{2\sigma^2}(g_i - a_i f_i)^2 - \frac{1}{2\tau^2}(f_i - \mu(f_{i-}))^2\right\} \\ &\propto \exp\left[-\tfrac{1}{2}\left\{\left(\frac{a_i^2}{\sigma^2} + \frac{1}{\tau^2}\right) f_i^2 - 2f_i\left(\frac{a_i g_i}{\sigma^2} + \frac{\mu(f_{i-})}{\tau^2}\right)\right\}\right] \\ &\propto \exp\left\{-\frac{\frac{a_i^2}{\sigma^2} + \frac{1}{\tau^2}}{2}\left(f_i - \frac{\frac{a_i g_i}{\sigma^2} + \frac{\mu(f_{i-})}{\tau^2}}{\frac{a_i^2}{\sigma^2} + \frac{1}{\tau^2}}\right)^2\right\} \end{aligned}$$

In claim 2 above it should be noted that $E(f_i|g_i, f_{-i})$ can be seen to be a weighted average of the data and mean of the prior distribution with the weights being functions of reciprocals of σ^2 and τ^2. In other words, the posterior mean is a weighted average of data mean and prior mean where the weights are proportional to their respective precisions.

4.4.2 *Modeling of the Stochastic Degradation Matrix*

A critical issue in the proposed MRF-MAP formulation described above is an appropriate choice for the values of α_i. A necessary prerequisite for choosing an appropriate value of α_i is to acquire some *a priori* knowledge on the shape of the mandible within the selected pixel blocks in each CT slice. The inner contour and outer contour of the portion of the mandible within the selected set of pixel blocks can each be essentially approximated by a quadratic polynomial with a distinct set of coefficients. A quadratic polynomial has the general form:

$$y = c_0 + c_1 x + c_2 x^2. \tag{4.14}$$

To estimate the coefficients c_0, c_1 and c_2 in equation (4.14), we need a set of three points, say, $(x_1, y_1), (x_2, y_2), (x_3, y_3)$ on the inner contour and likewise on the outer contour of the portion of the mandible in the selected set of pixel blocks in each image slice in the CT image stack. The need to obtain so many data points i.e., these six data points per contour (inner or outer) per CT image slice, can be justified as follows:

1. The inner and outer contours of the mandible, appearing in a particular CT image slice, cannot be represented mathematically by a single quadratic polynomial with

the appropriate translational parameters along the x and y axes. This is because the curvatures of the inner contour and the outer contour of the mandible are observed to be quite different.

2. Since the spatial resolution of the image stack is coarser along the z axis (axial direction) compared to the x and y axes, an inner or outer contour of the mandible in two different slices cannot be mathematically approximated by a single quadratic polynomial. This is once again due to difference in curvatures of an inner (or outer) contour in two consecutive CT image slices.

Here we once again re-emphasize a major advantage of the proposed localization phase. Fitting a second degree polynomial to a small portion of the bone surface is more appropriate than to the entire mandible. Although the whole mandible can be roughly viewed as a parabolic shape but it is not exactly a parabola. A small portion of the mandible, on the other hand, can be more accurately represented by a parabola (which is a second degree polynomial). Thus, it is more appropriate to use the second-order polynomial shape approximation within a small pixel lock over an attempt to derive a complex shape model for the entire mandible.

In the current implementation, user interaction is required (via computer mouse clicks) to generate all the required data points. Note that a typical set of three data points can be located anywhere along the inner or outer contour of the portion of the mandible within the chosen set of pixel blocks. Thus, the coefficients of the fitted quadratic polynomial are not particularly sensitive to the choice of the clicked points to the extent that these points lie on the (inner or outer) contour whose equation is being estimated. Once the quadratic polynomial for a contour is determined, a set of points satisfying the polynomial (i.e., set of points along the fitted contour) is generated. Typically, most of the points within the set have high intensity values, since they correspond to bone pixels, whereas only a few have low intensity values since they correspond to pixels at a potential fracture site. For potential fracture pixels at a site i, on a given contour, the value of α can be ideally assumed to be

given by:

$$\alpha_i = g_i / \max_i(g_i) \tag{4.15}$$

where $\max_i(g_i)$ represents the maximum of all the observed pixel intensity values along the contour under consideration. Therefore, it is reasonable to assume the mean of $\alpha_i (= \alpha_{i0})$ to be $g_i / \max_i(g_i)$.

Claim 3: Under the model characterized by equations (4.7), (4.9) and (4.13), it can be shown that the conditional posterior distribution of α_i is given by

$$\alpha_i | g_i, f_i \sim \mathcal{N}\left(\frac{\frac{g_i f_i}{\sigma^2} + \frac{\alpha_{i0}}{\eta^2}}{\frac{f_i^2}{\sigma^2} + \frac{1}{\eta^2}}, \frac{1}{\frac{f_i^2}{\sigma^2} + \frac{1}{\eta^2}} \right).$$

The proof of claim 3 is very similar to the proof of claim 2 given earlier.

Claim 4: Under the model characterized by equations (4.7), (4.9) and (4.13), the posterior distribution of σ^2 is given by

$$\sigma^2 | g, f, \alpha \sim \text{ Inverse-gamma } \left(\frac{p}{2} + \nu, \frac{S + 2\beta}{2} \right).$$

where $S = ||g - Af||^2$ and α is a vector with the elements being the collection of all α_is.

Proof: Note that

$$p(\sigma^2, f, g) \propto p_{\sigma^2}(g|f) p(f|\tau^2) p(\sigma^2)$$

From claim 2, we can use $p(f_i|g_i)$ and write.

$$p(\sigma^2 | f_i, g_i) \quad \propto \quad \left(\sigma^2\right)^{-\frac{p}{2} - (\nu + 1)} \exp\left\{ -\frac{S + 2\beta}{2\sigma^2} \right\}$$

Hence the claim.

Claim 5: Under the model characterized by equations (4.7), (4.9) and (4.13), the posterior distribution of τ^2 is given by

$$\tau^2 | f, g \sim \text{ Inverse-gamma } \left(\frac{p}{2} + \kappa, \frac{f^T (I_p - \gamma N) f + 2\delta}{2} \right).$$

The proof of the above claim is similar to the proof of Claim 4. It can be noted that all conditional posterior distributions in question have closed-form expressions that are known. Under this situation, we can iteratively draw samples from the posterior distribution using Gibbs sampling. The Gibbs sampling procedure ensures convergence to the MAP estimate of the true image after a sufficient number of iterations, Geman and Geman (1984).

4.5 Experimental Results and Analysis

In this section, we will describe experimental results on fracture detection and target mandibular reconstruction using the proposed two-phase MRF-MAP-based scheme within a hierarchical Bayesian setup. The experiments were performed on 7 different real patient CT scans with different resolutions (pixels/mm) and fracture locations. The fractures were observed to appear typically in 3-6 2D slices in each CT image sequence. Each 2D CT image slice is of size 512×512 pixels with a grayscale resolution of 8 bits per pixel.

4.5.1 *Results on Fracture Localization*

Dataset	Block Size	No. of Potential Blocks	No. of Selected Blocks	Correlation Values
$Dataset-1$	$64 \times 64 \times 3$	8	2	$0.022, 0.038$
	$32 \times 32 \times 3$	13	4	$0.007, 0.068, 0.109, 0.153$
$Dataset-2$	$32 \times 32 \times 3$	15	2	$0.012, 0.036$
$Dataset-3$	$32 \times 32 \times 3$	11	2	$0.03, 0.16$
	$32 \times 32 \times 3$	11	3	$0.03, 0.16, 0.23$
$Dataset-4$	$32 \times 32 \times 4$	8	4	$0.002, 0.028, 0.087, 0.144$
$Dataset-5$	$32 \times 32 \times 3$	9	2	$0.029, 0.101$

Table 4.1: Fracture localization

Dataset	Emphysema	Tissue Swelling (pixel distance)	No. of bone pixels	Fractured Half
$Dataset-1$	*Detected in the left half*	–	–	$LEFT$
$Dataset-2$	–	$left: 65, right: 54$	–	$LEFT$
$Dataset-3$	*Detected in the left half*	–	–	$LEFT$
$Dataset-4$	–	$left: 36, right: 79$	–	$RIGHT$
$Dataset-5$	–	$left: 44, right: 76$	–	$RIGHT$
$Dataset-6$	–	–	$left = 5996.0$ $right = 5097.0$	$RIGHT$
$Dataset-7$	–	$left: 60, right: 64$	–	$RIGHT$

Table 4.2: Detection of the fractured half

In the first stage of the proposed two-phase scheme, we localized the fracture by determining the approximate region and the image half, in which it appears. The pixel block width and height (which is a power of 2 in the current implementation) is sought an input from the user via a graphical user interface (GUI). The block thickness is deemed to be the number of 2D slices, in which a broken jaw appears in a given CT sequence. For the CT image sequences used in the current experiments, the pixel block width and height vary between 64×64 and 32×32. Similarly the thickness for the present set of data varies between 3 and 6. However, our scheme is fairly robust to the chosen block dimensions. For example, from Figure 4.1 and Table 3.1, it is evident that the localization works correctly for *dataset* 1 with two different pixel block dimensions given by $64 \times 64 \times 3$ and $32 \times 32 \times 3$. With pixel block dimensions of $64 \times 64 \times 3$, there were altogether 32 pairs of pixel blocks to be examined. For this particular dataset i.e. *dataset* 1, after the two heuristics based on mandible size and bone pixel intensity were applied, only 8 out of 32 blocks were retained. These were then arranged in order of increasing value of the correlation coefficient. Finally, 2 out of these 8 pixel blocks were chosen as potential hairline fracture sites. For the same dataset, with pixel block dimensions of $32 \times 32 \times 3$, the first 13 out of a possible 128 blocks were retained using the aforementioned two heuristics and eventually 3 out of 13 blocks were chosen as potential hairline/minor fracture sites based on the correlation coefficient values. Table 3.1 shows details of our experiments with five different datasets and two different block sizes. It shows the number of selected blocks based on correlation values in each case and typical values of correlation coefficient for potential fracture-containing blocks. Table 3.1 along with Figure 4.1 clearly illustrates that our scheme is not strictly sensitive to (a) the chosen pixel block size and (b) the number of pixel blocks selected as sites of minor/hairline fractures. However, we would like to add the following observations (which hold in most of the cases that)

1. For a mandible of larger size, the size of the pixel block may be chosen to be larger e.g. a 64×64 block size would work better compared to a 32×32 block size for a larger mandible.

2. If number of mandible-containing blocks (k) is smaller, then it is better to be conservative in choosing number of selected blocks (q) as sites of potential fractures e.g. if k turns out to be 2, then it is preferrable to choose q as 2, rather than 1 (a case not shown in Table 3.2).

Next, we discuss the determination of the CT image half containing the potential fracture. Out of the seven datasets we experimented with, we detected the presence of emphysema in two cases (see Table 3.2). The detection of *emphysema* is also marked by the presence of white crosses in Figure 4.1. Note that we have just marked one particular pixel site to denote the occurrence of *emphysema* in each of the 2D slices, where as potentially there are quite a few such emphysema pixels. Obviously, the image half in which emphysema is detected is designated as the one containing fracture site. Table 3.2 demonstrates the detection of *swelling of the soft tissue* for four other datasets. For a given dataset, the average pixel distance of the outer soft tissue boundary from the center of the localized block is computed for both the left and the right image half. The image half for which the average pixel distance is higher, indicates the presence of *soft tissue swelling* and hence the fracture site. The difference in the distance can be quite high such as 30- 40 pixels in case of major swelling (as indicated for *datasets* 4 and 5 in Table 3.2) or could be quite small like under 10 pixels in case of minor swelling (as indicated for *datasets* 1 and 7 in Table 3.2). The major and minor swellings can be found in Figures 4.4 and 4.5 and Figures 4.1) and 4.7 respectively. In one single case (*dataset* 6), we used the count of the number of bone pixels (i.e. pixels with a high intensity value) within the localized blocks (in both the halves) for determining the fracture-containing image half.

Logically, the fracture half (within the localized area) should contain fewer bone pixels compared to its non-fracture-containing symmetric counterpart. As there was no *emphysema* and the *soft tissue swelling* was negligible, the above heuristic was employed to determine the fracture-containing image half for *dataset* 6 (see Figure 4.6 and Table 3.2). The first and second row in each of the Figures 4.1, ..., 4.7 respectively represent a CT image sequence that depicts a mandible with a minor/hairline fracture and the localization of the fracture.

Additionally, in Figure 4.1, the fourth row portrays fracture localization with a different set of pixel block dimensions and with a different number of potential fracture containing blocks (selected, based on correlation values). For all the seven datasets, the localized fractures are highlighted by the surrounding black boxes at the fracture sites.

4.5.2 *Precise fracture detection using MRF-MAP scheme*

Coeffs. of Quad. Poly.	Slice 1		Slice 2		Slice 3	
$(c_0 + c_1x + c_2x^2)$	Outer	Inner	Outer	Inner	Outer	Inner
c_0	-11344	6808	8798	5228.7	-177200	-107.19
c_1	162	75.7	-113.2	67.8	2380	5.79
c_2	-1	2	4	2	-10	-0.02

Table 4.3: Coeffs. of the Quad. Poly. for outer and inner contours over various slices for the dataset 1

In this subsection, we will first discuss in detail the three coefficients, required to fit a second order polynomial to a mandible contour in a 2D slice. From Table 3.3, it is quite clear that for a given slice, the outer and inner contours of a jaw cannot be approximated by two (approximately) parallel quadratic polynomials as all of the three coefficients are widely different. Thus, the user needs to provide six data points per slice (3 data points for the inner contour and 3 data points for the outer contour). Table 3.3 also shows that an outer(inner) contour cannot be approximated over all the slices using the same quadratic polynomial. Thus, the necessity of six points per 2D slice (that contains a fracture) is numerically justified. These points are provided by the user via interactive mouse clicks. It is important to note in this context that the proposed scheme is not sensitive to any specific pattern of mouse clicks. As long as the user clicks on three points on the outer contour and three points on the inner contour in each of the relevant 2D slices, the polynomial fitting procedure and hence subsequent fracture detection works well. This is clearly illustrated in Figure 4.8, where the contour extraction for two different datasets are shown. The centers of the black crosses mark the locations of the mouse clicks. In some cases, even the portion of the mandible contour within the localized area cannot be approximated using a quadratic polynomial.

In such cases, one needs to further restrict the span of mouse clicks within the localized region. Another practical constraint for reducing the spatial extent of the mouse clicks could be excessive intensity inhomogeneity in the CT images. We have used domain knowledge about possible fracture point intensity values to estimate the degradation matrix. So, using equation (4.15), we assign a value of 1 to a_i if α_i is in the range [0.85, 1]; we assign α_i to be the mean of a_i otherwise.

For the MRF-MAP formulation, we chose a first order neighborhood. As far as the statistical parameters are concerned, we chose the shape and scale of the inverse gamma distribution for τ^2 to be 3.0 and 1.0 and for σ^2 to be 3.0 and 0.5. We experimented with both 100 and 1000 as the number of iterations for the possible convergence of the Gibbs sampling procedure (i.e., for the posterior distribution to attain its maximum value) for two datasets. Since, we did not find any notable visual difference in the reconstructed output, we simply decided to use 100 iterations for the remaining datasets, thus reducing the execution time. In fact the completion of each phase in the proposed two-phase scheme was observed to take about 2-3 minutes on a 1.73 GHz Intel© Pentium©-M processor. It is relevant to discuss here why we opted for a partial 3D scheme instead of a pure 3D scheme for the purpose of fracture detection. The localization procedure is essentially a 3D scheme as the correlation is computed over a single 3D pixel block instead of multiple 2D blocks in individual 2D slices. The estimation of the degradation matrix calls for prior polynomial fitting to the contours of the mandible. It is visually intuitive to fit a second-order polynomial to portions of these contours over the localized block in individual 2D slices. Thus, as mentioned earlier, we can avoid the use of complicated shape models for an entire 3D mandible. Furthermore, since the number of slices in which a fracture appears in a stack is often few (typically 3 to 6), a complete 3D scheme would not have resulted in a significant performance improvement in terms of overall execution time. The third row in each of the Figures 4.2, ..., 4.7 and the third as well as the fifth row in the Figure 4.1 show the results of precise fracture detection and visualization using the proposed MRF-MAP scheme. Note that whereas the original intensity values at the fracture pixel sites are low (due to bone loss), the reconstructed (restored) intensity values at these sites are high, (corresponding to the bone pixel inten-

sities) on account of MAP estimation via repetitive Gibbs sampling. The pixels exhibiting large intensity differences (between the reconstructed and the fractured mandible) are highlighted (using the color black) for the purpose of visualization. The detection of the hairline fractures was validated by experienced surgeons. In this context, we would like to briefly discuss about controlling the threshold (indicating the difference in intensity between the input data and the reconstructed data) for fracture visualization. If the user selects a low threshold value, the final fracture appearance will be quite restricted. On the contrary, if the user selects a higher value, the fracture at the proper locations will be displayed more prominently but at the cost of some spurious appearances (due to spatial spread). This fact is well illustrated by Figure 4.6 where the threshold value increases from the third to the fifth row. This procedure for adjusting the threshold can be viewed as a practical knob for controlling the visualization of the fractures. We end this section with a brief discussion on target pattern generation. Figures 4.9, 4.10 and 4.11 depict the generated target patterns for three different datasets. The first row in each of these figures depicts the broken jaw, where as the second row illustrates the *in silico* reconstructed jaw, generated using the proposed MRF-MAP-based approach. It is quite evident that the reconstructions in Figures 4.9 and 4.10 exhibit relatively smooth mandible contours which, in turn, indicate that natural bone healing could be opted for in such cases. However, the reconstructed jaw in Figure 4.11 exhibits a relatively rugged mandible contour which indicates that surgical intervention may be a more appropriate choice in this case.

4.6 Conclusion and Future Work

In this paper, we presented a novel two-phase scheme for simultaneous hairline fracture detection and target pattern generation for fractured human mandibles. The fracture detection scheme is robust to the presence of noise and intensity inhomogeneity, present in the CT images. The reconstructed jaw can be treated as a target pattern which can assist a surgeon in determining whether manual surgical intervention is necessary or whether natural bone healing is adequate. In the first phase of the proposed two-phase approach, the

hairline/minor fractures are localized within the pixel blocks of a given size by analyzing all the image slices in the CT image stack. The CT image analysis exploits the (approximate) bilateral symmetry of the human mandible and uses the statistical correlation coefficient as a measure of intensity mismatch. The CT image half, in which a fracture appears is determined by exploiting domain knowledge, which includes the presence of *emphysema*, *soft tissue swelling* and occurrence of a fewer bone pixels in localized fracture sites. In each of the aforementioned pixel blocks, an MRF-MAP-based approach for hairline/minor fracture detection is employed. The MRF-MAP-based approach surpasses existing conventional techniques based on corner detection or discontinuity detection in the image intensity surface. The hairline fracture detection scheme has obvious diagnostic significance. Since the implicit reconstruction procedure embedded within the proposed MRF-MAP based technique is designed to mimic the natural bone healing process in the absence of any surgical intervention, the proposed scheme has an important prognostic significance as well. In addition to the aforementioned clinical significance, the problem of hairline fracture detection also has certain noteworthy aspects of theoretical interest from the perspective of computer vision and pattern recognition research. This is primarily because we are faced with the challenging task of modeling a spatially localized degradation resulting from a mathematically unknown degradation function. Thus, we first computed the degradation matrix from the input data by fitting quadratic polynomial functions to the inner and outer contours of the mandible. This was followed by the application of the Gibbs sampling procedure for the MAP probability estimation. In the current implementation, the degradation matrix A is modeled as a stochastic entity, by imposing a normal distribution on the α values. Additionally, we used the inverse-gamma distribution for both τ^2 (variance of the prior distribution) and σ^2 (variance of the data model) in order to conform to the hierarchical Bayesian estimation paradigm.

Our future plan is to incorporate a higher degree of automation within the existing scheme. The current scheme for quadratic polynomial approximation of the contours of the mandible calls for user interaction via mouse clicks; which could be automated in order to reduce the burden on the user. Another future direction entails automatic identification of the

set of 2D CT image slices containing the fractured mandible from a larger set of CT images. Note that a typical CT image sequence is an aggregation of 2D image slices where only some of the image slices contain the fractured mandible. We also intend to explore the utility of the present MRF-MAP approach as a smoothing tool. Thus, we can potentially use the proposed MRF-MAP scheme as a smoothing tool on instances of well-displaced fractures that have been registered, Bhandarkar et. al. (2004). Subsequently, we can generate a better 3D model of the reconstructed mandible resulting in a more accurate analysis of biomechanical stability. From the statistical perspective, we plan to investigate the effects of (a) a higher-order neighborhood (e.g. second order) for the MRF or a different model for the MRF-MAP paradigm (e.g. the SAR model), Li (1995) and (b) an objective assumption on the prior distribution of the nuisance parameters which would make the problem computationally more challenging, Gelman et. al (2004).

Acknowledgment

The authors would like to thank Prof. Ramon Figueroa, Department of Radiology, Medical College of Georgia for providing all the real-patient CT scans and Prof. Ernest W. Tollner, Department of Biological and Agricultural Engineering, University of Georgia for discussing how the present approach can possibly improve biomechanical stability analysis.

4.7 References

[1] Besag, J. (1986). On the Statistical Analysis of Dirty Pictures, *J. Roy. Statist. Soc. B*, vol. 48, no. 3, pp. 259-302.

[2] Bhandarkar, S. M., Chowdhury, A. S., Tang, Y., Yu, J. and Tollner, E. W. (2004). Surface Matching Algorithms for Computer Aided Reconstructive Plastic Surgery, in *Proc. of IEEE ISBI*, Arlington, VA, pp. 740 - 743.

[3] Chan, A. K. and Zheng, L. (2001). An artificial intelligent algorithm for tumor detection in screening mammogram, *IEEE Trans. Med. Imag.*, vol. 20, no. 7, pp. 559 - 567.

[4] Chen, T. and Metaxas, D. N. (2004). Markov Random Field Models, in *Insight into Images*, T. S. Yoo Ed., A.K. Peters, Wellesey.

[5] Choi, S. M., Lee, J. E., Kim, J. and Kim, M. H. (1997). Volumetric Object Reconstruction Using the 3-D MRF Model-Based Segmentation, *IEEE Trans. Med. Imag.*, vol. 16, no. 6, pp. 887-892.

[6] Chowdhury, A. S., Bhattacharya, A., Bhandarkar, S. M., Datta, G. S., Yu, J. C. and Figueroa, R. (2007). Hairline Fracture Detection using MRF and Gibbs Sampling, *Proc. of IEEE WACV*, Autin, TX, p56.

[7] Chowdhury, A. S., Bhandarkar, S. M., Datta, G. and Yu, J.C. (2006). Automated Detection of Stable Fracture Points In Computed Tomography Image Sequences, *Proc. of IEEE ISBI*, Arlington, VA, pp. 1320 - 1323.

[8] Donnelley, M. and Knowles, G. (2005). Automated Bone Fracture Detection, in *Proc. of SPIE Medical Imaging*, San Diego, CA, pp. 955 - 966.

[9] Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004) *Bayesian Data Analysis*, Chapman & Hall-CRC, NY.

[10] Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. 6, pp. 721 - 741.

[11] Giannoudis, P. V. and Dinopoulos, H. (2005). Current concepts of the inflammatory response after major trauma: an update, *Injury*, vol. 36, no. 1, pp. 229-230.

[12] Harris, C. J. and Stephens, M. (1988) A combined corner and edge detector, in *Proc. 4th Alvey Vision Conference*, Manchester, UK, pp. 147-151.

[13] Jia, Y. and Jiang, Y. (2006) Active Contour Model with Shape Constraints for Bone Fracture Detection, in *Proc. of IEEE CGIV*, Sydney, Australia, pp. 90-95.

[14] King, R.E., Scianna, J.M. and Petruzzelli, G.J. (2004). Mandible Fracture Patterns: a Suburban Trauma Center Experience, *Am. J. of Otolaryngol.*, vol. 25, no. 5, pp. 301-307.

[15] Lee, S.J., Rangarajan, A. and Gindi, G. (1995). Bayesian Image Reconstruction in SPECT Using Higher Order Mechanical Models as Priors, *IEEE Trans. Med. Imag.*, vol. 14, no. 4, pp. 669 -680, Dec. 1995.

[16] Li, S. Z. (1995) *Markov random field modeling in computer vision*, Springer-Verlag London, UK.

[17] Lum, V., Leow, W., Chen, Y., Howe, T. and Png, M. (2005) Combining classifiers for bone fracture detection in X-ray images, in *Proc. of IEEE ICIP*, Genoa, Italy, pp. 1149 - 1152.

[18] Martin-Fernandez, M. and Alberola-Lopez, C. (2005). An approach for contour detection of human kidneys from ultrasound images using Markov random fields and active contours, *Med. Image Anal.*, vol. 9, pp. 1 - 23.

[19] Molina, R. (1994). On the Hierarchical Bayesian Approach to Image Restoration: Applications to Astronomical Images, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 11, pp. 1122-1128.

[20] Ogundare, B.O., Bonnick, A. and Bayley, N. (2003) Pattern of Mandibular Fractures in an Urban Major Trauma Center, *J. of Oral and Maxillofac. Surg.*, vol. 61, no. 6, pp. 713-718.

[21] Ozanian, T. and Phillips, R. (2000). Image Analysis for Computer-Assisted Surgery of Hip Fractures, *Med. Image Anal.*, vol. 4, no. 2, pp. 137 - 159.

[22] Popa, C. and Zdunek, R. (2005) Penalized Least-Squares Image Reconstruction For Borehole Tomography, in *Proc. of Algorithmy*, Podbanske, Slovakia pp. 260-269.

[23] Prima, S., Ourselin, S. and Ayache, N. (2002). Computation of the Mid-Sagittal Plane in 3D Brain Images, *IEEE Trans. Med. Imag.*, vol. 21, no. 2, pp. 122-138.

[24] Raj, A., Singh, G. and Zabih, R. (2006). MRFs for MRIs: Bayesian Reconstruction of MR Images via Graph Cuts, in *Proc. of CVPR*, New York, NY, pp. 1061 - 1068.

[25] Richard, F. J. P. (2005). A comparative study of Markovian and variational image-matching techniques in application to mammograms, *Pattern Recognit. Lett.*, vol. 26, pp. 1819 - 1829.

[26] Salli, E., Aronen, H. J., Savolainen, S., Korvenoja, A. and Visa, A. (2001). Contextual Clustering for Analysis of Functional MRI Data, *IEEE Trans. Med. Imag.*, vol. 20, no. 5, pp. 403-414.

[27] Syiam, M., El-Aziem, M. A. and El-Mensshawy, M. (2004) Adgen: Adaptive Interface Agent for X-Ray Fracture Detection, *Internat. J. of Comput. and Inform. Sci.*, vol. 2, no. 3, pp. 143 - 148.

[28] Vodovotz, Y., Chow, C. C., Bartels, J., Lagoa, C., Prince, J. M., Levy, R. M., Kumar, R., Day, J., Rubin, J., Constantine, G., Billiar, T. R., Fink, M. P. and Clermont, G. (2006). In silico models of acute inflammation in animals, *Shock*, vol. 26, no. 3, pp. 235-44.

[29] Yap, D. W. H., Chen, Y., Leow, W. K., Howe, T. S. and Png, M. A. (2004) Detecting Femur Fractures by Texture Analysis of Trabeculae, in *Proc. of ICPR*, Cambridge, UK, pp. 730-733.

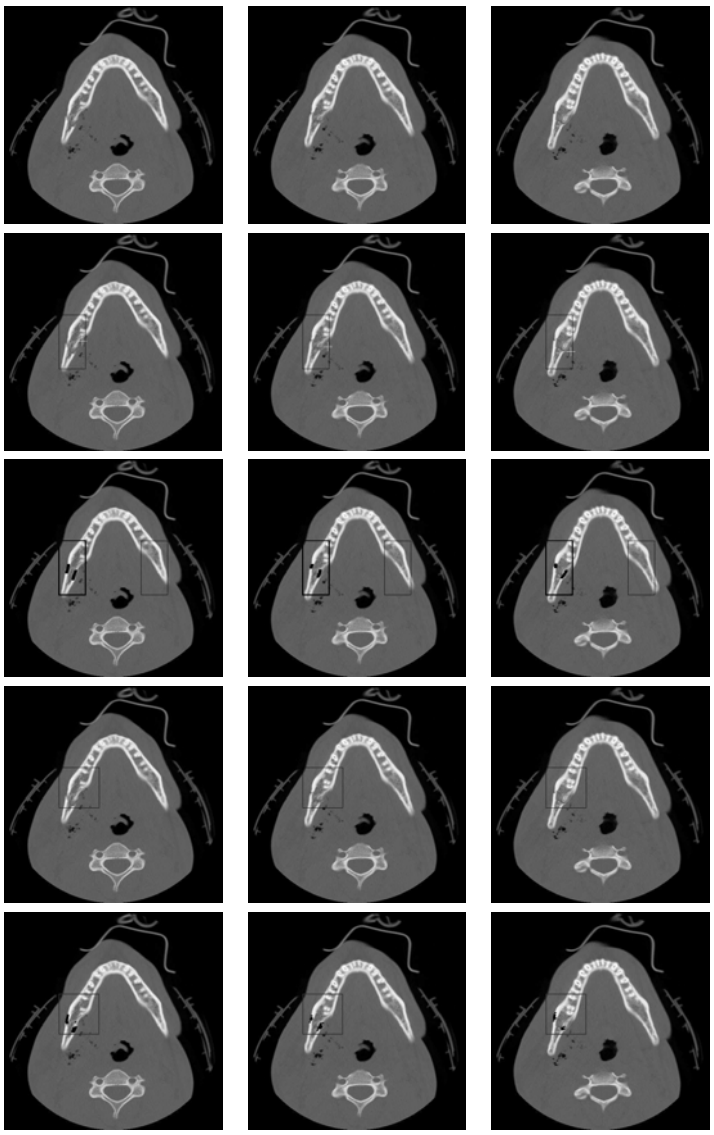

Figure 4.1: Fracture detection for dataset 1 (the topmost row shows the input sequence, the second and third row from the top shows fracture localization and precise detection with a 'block size' of $64 \times 64 \times 3$ and 2 selected blocks (in terms of correlation); the fourth and fifth row respectively show the same with a 'block size' of $32 \times 32 \times 3$ and 4 selected blocks (in terms of correlation); additionally centers of white crosses in the second row mark the detection of *emphysema*)

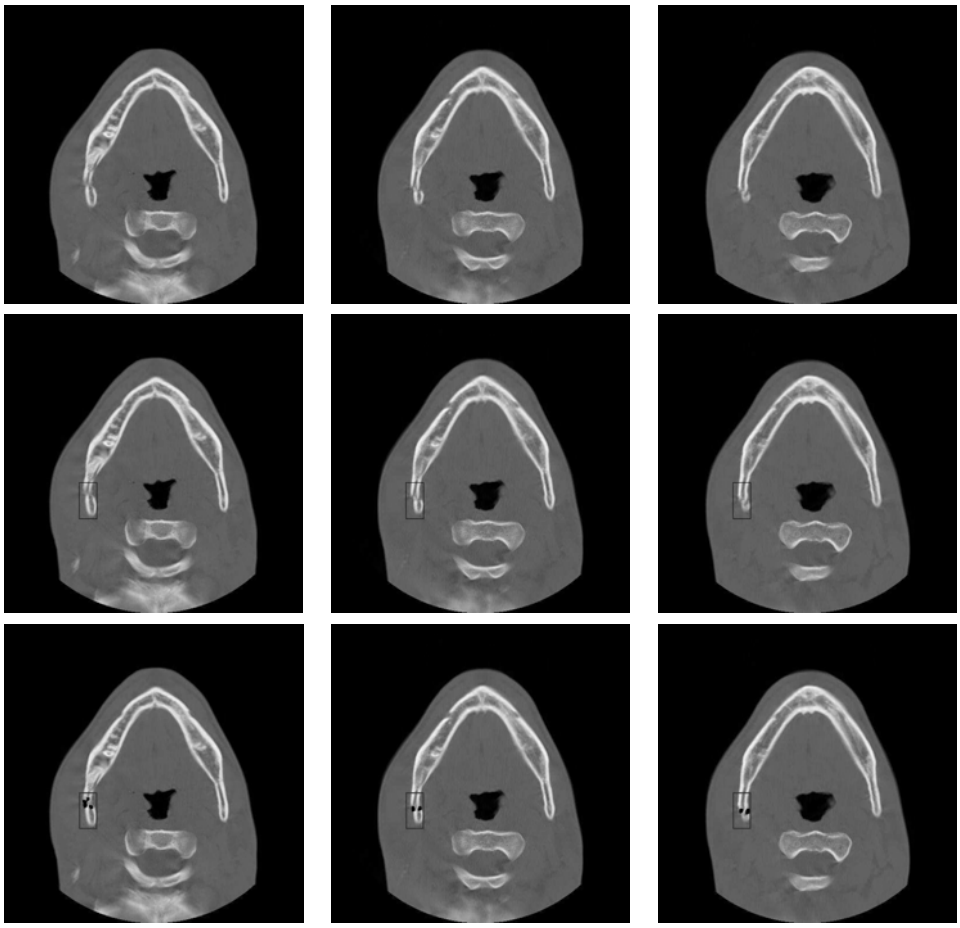

Figure 4.2: Fracture detection for dataset 2 (the top row shows the input sequence, the middle row shows the localization and the bottom row shows the precise detection of the fracture)

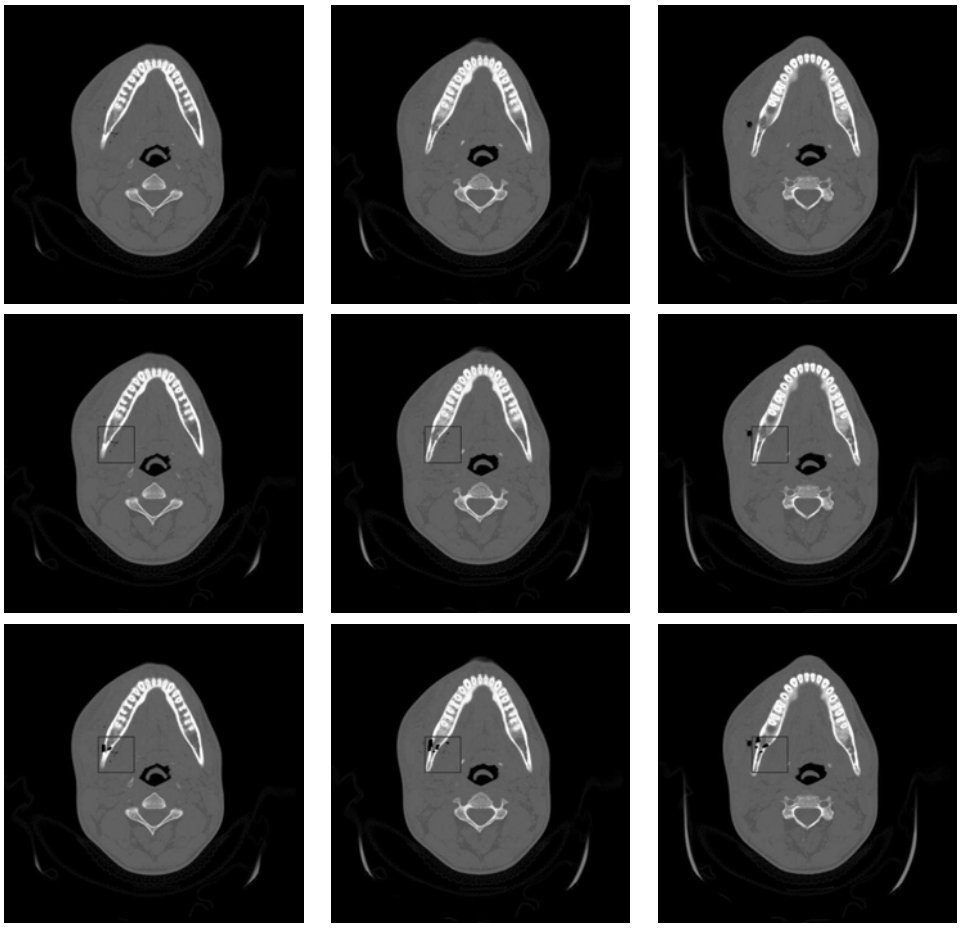

Figure 4.3: Fracture detection for dataset 3 (the top row shows the input sequence, the middle row shows the localization and the bottom row shows the precise detection of the fracture)

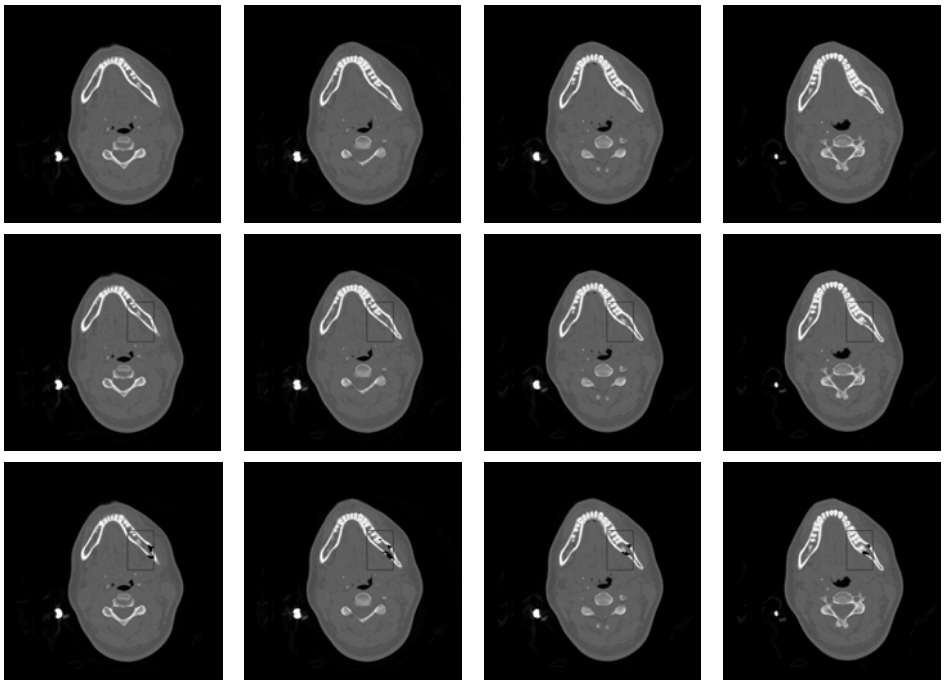

Figure 4.4: Fracture detection for dataset 4 (the top row shows the input sequence, the middle row shows the localization and the bottom row shows the precise detection of the fracture)

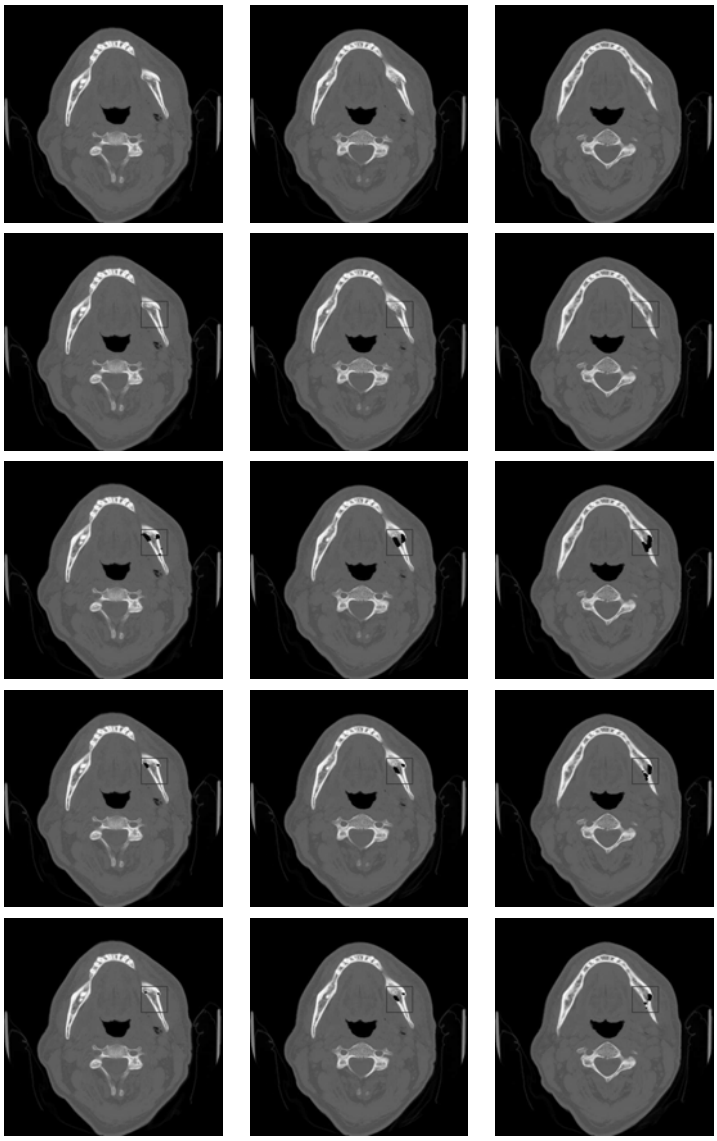

Figure 4.5: Fracture detection for dataset 5 (the topmost row shows the input sequence, the second row from the top shows the localization of the fracture, the third, fourth and fifth rows show the precise detection and visualization of the fracture with successive increase in the threshold value; the threshold value indicates a difference in intensity between the input and the reconstructed data)

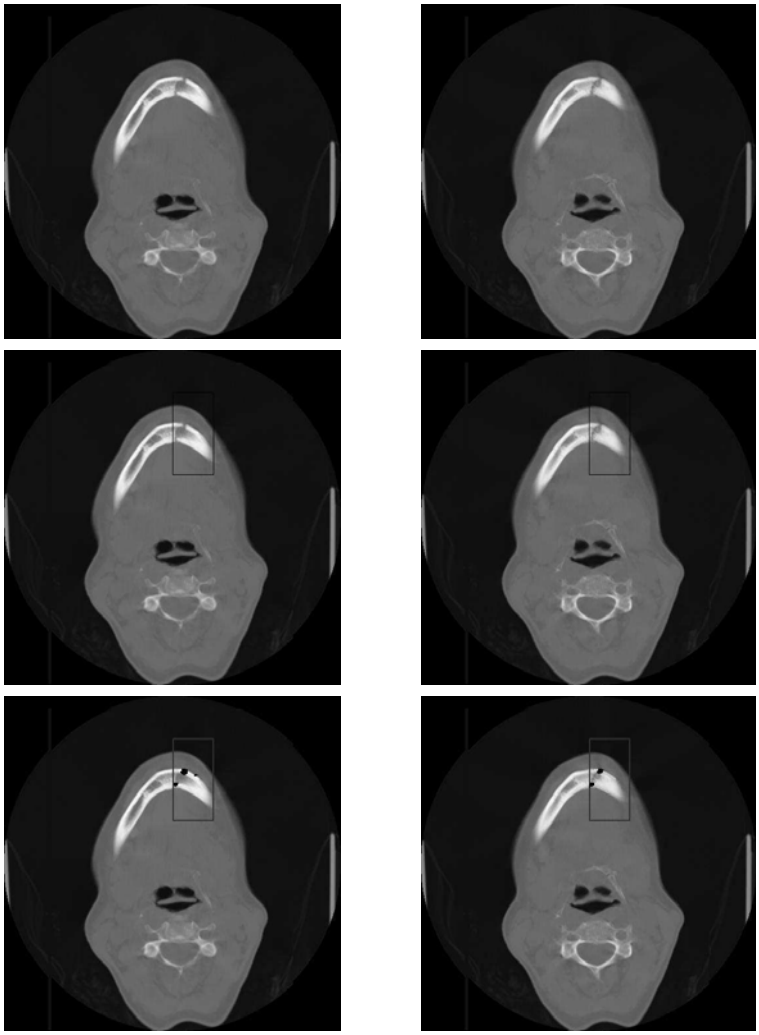

Figure 4.6: Fracture detection for dataset 6 (the top row shows the input sequence, the middle row shows the localization and the bottom row shows the precise detection of the fracture)

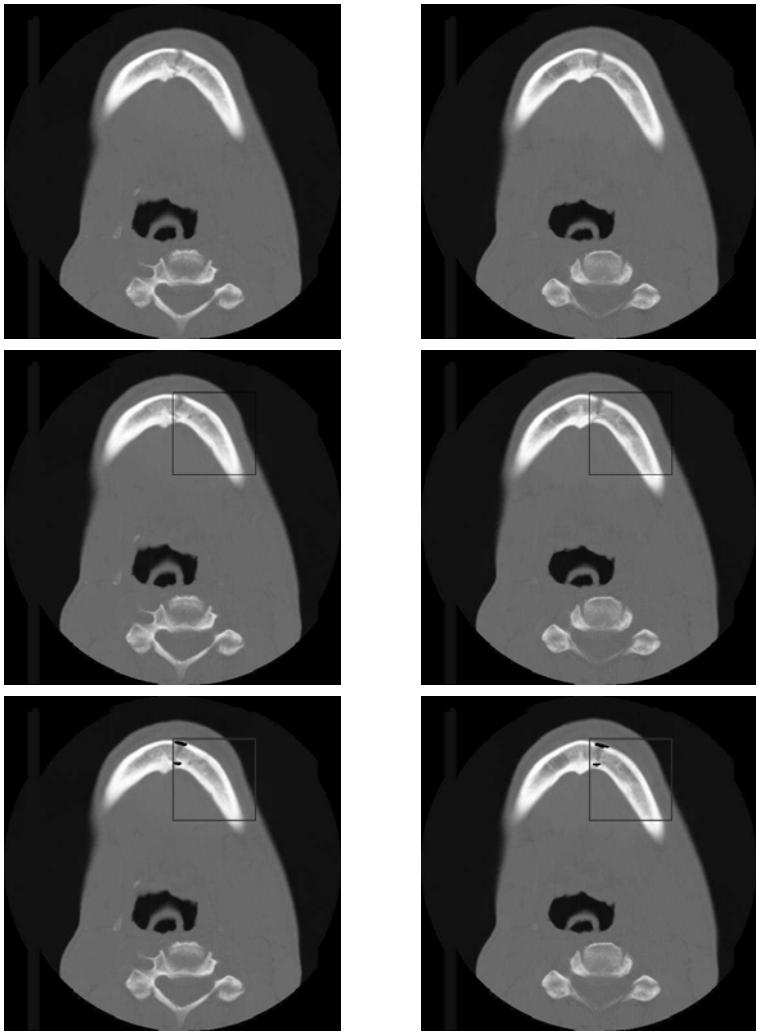

Figure 4.7: Fracture detection for dataset 7 (the top row shows the input sequence, the middle row shows the localization and the bottom row shows the precise detection of the fracture)

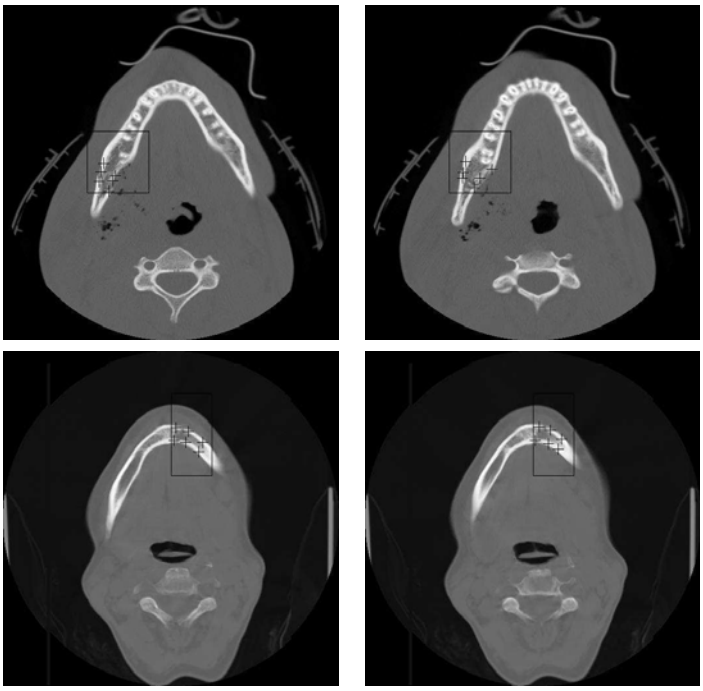

Figure 4.8: Extraction of contour information with mouse clicks (for dataset 1 in the top row and dataset 6 in the bottom row)

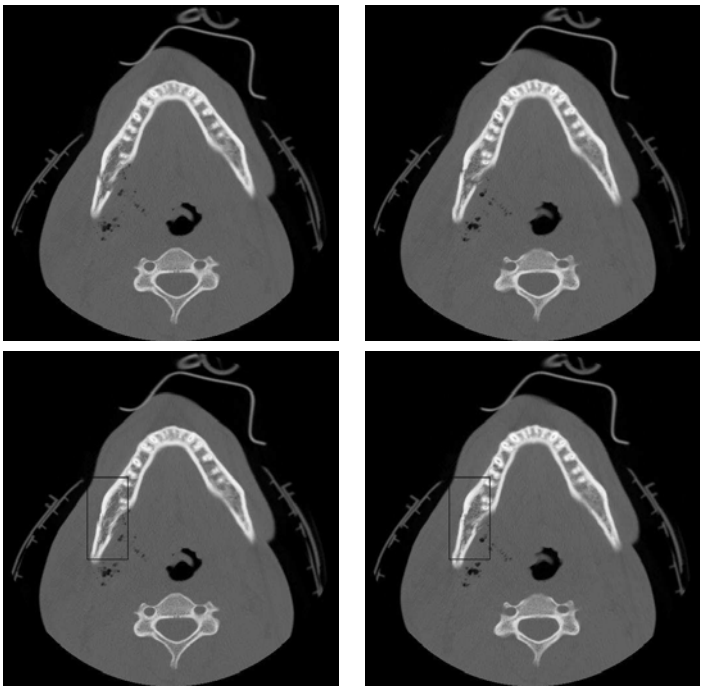

Figure 4.9: Target pattern generation for dataset 1 (the top row shows the fractured jaw and the bottom row shows the reconstructed jaw)

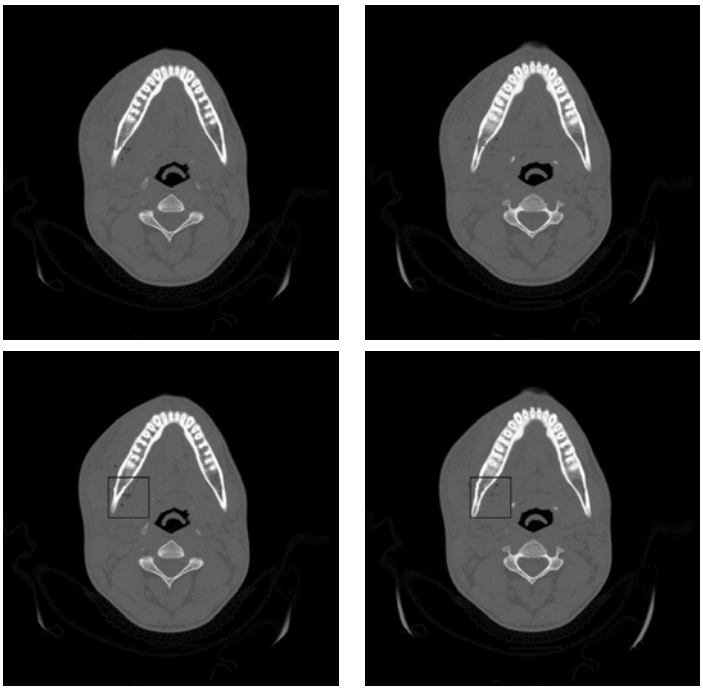

Figure 4.10: Target pattern generation for dataset 3 (the top row shows the fractured jaw and the bottom row shows the reconstructed jaw)

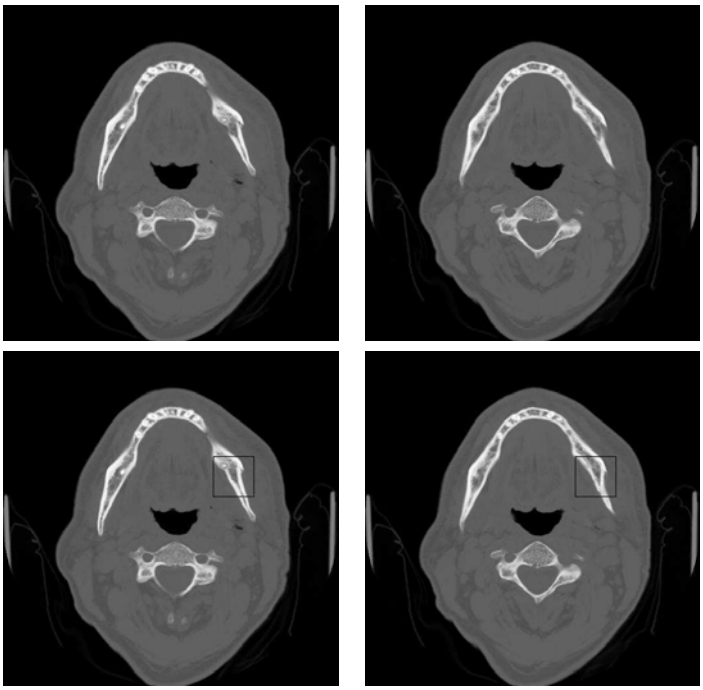

Figure 4.11: Target pattern generation for dataset 5 (the top row shows the fractured jaw and the bottom row shows the reconstructed jaw)

Chapter 5

Conclusions

Three different problems are addressed in this thesis from different fields of statistics and its applications. In this chapter, some brief concluding remarks are discussed regarding each of the three problems and their solutions.

In Chapter 2, controlled branching processes are considered and we provide an estimate of the offspring mean using the conditional weighted least squares approach. In fact, it can be noted that controlled branching processes forms a general class of branching processes, which include the simple branching process and the branching process with immigration as well. Under certain regularity conditions, a suitably normalized version of the estimator we proposed asymptotically follows normal distribution in supercritical and subcritical cases, while the limit distribution of the estimator is a function of a diffusion process in the critical case. It is also noted that in supercritical cases our estimator is asymptotically efficient. Unifying the limit ditribution of our estimator would be an interesting result in this field of research.

In the next chapter, a Bayesian test of zero-inflation is proposed as an alternative to its frequentist counterpart. The most interesting idea in this research is extending the paramter space in a feasible way to tackle the boundary point problem, since testing parameters on the boundary is well known for creating inconvenience in different frequentist and Bayesian methodologies. We propose a Bayesian test, study its finite sample and large sample properties and also provided methodologies to construct credible set in the presence of nuisance paramter. Extending these ideas in the case of data that is coming from several populations, is an area that I am interested to pursue in future.

In Chapter 4, a Bayesian technique of image restoration using Markov random field is applied for detecting and reconstructing hairline fracture in human mandible based on

Computer Tomography images. The fundamental idea in this research is using the stochastic degradation function applied locally only on some part of the image. Generally, degradation functions are designed in such a way that if it is applied to the whole true target image in presence of noise, the result is more or less what we have as data. Here in a two-step way, the fracture is localized using some statistical properties of the image and some domain knowledges and then the fracture is detected and reconstructed within that localized part. We still have to depend on user intervention to find some points on the curvature of the mandible. Although we have shown that our method is robust to that, it would be more appealing if we can automate that part of the procedure. Also from statistical point of view, it would be interesting to perform an objective Bayesian analysis for this problem.

Bibliography

[1] Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *J. R. Statist. Soc. B.* vol 41 pp 113-147.

[2] Besag, J. (1986). On the Statistical Analysis of Dirty Pictures, *J. Roy. Statist. Soc. B*, vol. 48, no. 3, pp. 259-302.

[3] Bhandarkar, S. M., Chowdhury, A. S., Tang, Y., Yu, J. and Tollner, E. W. (2004). Surface Matching Algorithms for Computer Aided Reconstructive Plastic Surgery, in *Proc. of IEEE ISBI*, Arlington, VA, pp. 740 - 743.

[4] Bhat, B. R. and Adke, S. R. (1981). Maximum likelihood estimation for branching processes with immigration. *Adv. Appl. Prob.* vol 13. pp 498-509.

[5] Billingsley, P. (1968). *Probability and Measure*. John Wiley & Sons, New York.

[6] Broek, J. V. D. (1995). A score test for zero inflation on a Poisson distribution. *Biometrics.* vol 51. pp 738-743.

[7] Chan, A. K. and Zheng, L. (2001). An artificial intelligent algorithm for tumor detection in screening mammogram, *IEEE Trans. Med. Imag.*, vol. 20, no. 7, pp. 559 - 567.

[8] Chen, M.-H. and Shao, Q.-M. (1999). Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics.* vol 8. pp 69-92.

[9] Chen, T. and Metaxas, D. N. (2004). Markov Random Field Models, in *Insight into Images*, T. S. Yoo Ed., A.K. Peters, Wellesey.

[10] Choi, S. M., Lee, J. E., Kim, J. and Kim, M. H. (1997). Volumetric Object Reconstruction Using the 3-D MRF Model-Based Segmentation, *IEEE Trans. Med. Imag.*, vol. 16, no. 6, pp. 887-892.

[11] Chow, Y. S. and Teicher H. (1997). *Probability Theory: Independence, Interchangeability, Martingales*. Springer-Verlag, New York. 3rd ed.

[12] Chowdhury, A. S., Bhattacharya, A., Bhandarkar, S. M., Datta, G. S., Yu, J.C. and Figueroa, R. (2007). Hairline Fracture Detection using MRF and Gibbs Sampling, *Proc. of IEEE WACV*, Autin, TX, p56.

[13] Chowdhury, A. S., Bhandarkar, S. M., Datta, G. and Yu, J.C. (2006). Automated Detection of Stable Fracture Points In Computed Tomography Image Sequences, *Proc. of IEEE ISBI*, Arlington, VA, pp. 1320 - 1323.

[14] Clarke, B. and Barron, A. (1994) Jeffreys prior is asymptotically least favorable under entropy loss. *Journal of Statistical Planning and Inference.* vol. 41, pp. 37-60.

[15] Cochran, W. G. (1954). Some methods of strengthening χ^2 tests. *Biometrics.* vol 10. pp 417-451.

[16] Conigliani, C., Castro, J. I. and O'Hagan, A. (2000). Bayesian Assessment of Goodness of Fit against Nonparametric Alternatives. *Canadian Journal of Statistics.* vol 28. pp 327-342.

[17] Datta, G. S. and Mukerjee, R. (2004). Probability Matching Priors: Higher Order Asymptotics. *Springer-Verlag.*

[18] Deng, D and Paul, S. R. (2000). Score test for zero inflation in generalized linear models. *Canadian Journal of Statistics.* vol 28. pp 563-570.

[19] Deng, D and Paul, S. R. (2005). Score tests for zero-inflation and over-dispersion in generalized linear models. *Statistica Sinica.* vol 15. pp 257-276.

[20] Dion, J. P. and Essebbar, B. (1995). On the statistics of controlled branching processes. *Lecture Notes in Statist.*, 99. pp 14-21.

[21] Donnelley, M. and Knowles, G. (2005). Automated Bone Fracture Detection, in *Proc. of SPIE Medical Imaging*, San Diego, CA, pp. 955 - 966.

[22] Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association.* vol 81. pp 709-721.

[23] El-Shaarawi, A. H. (1985). Some goodness-of-fit methods for the Poisson plus added zeros distribution. *Applied and Environmental Microbiology.* vol 49. pp 1304-1306.

[24] Ethier, S. N. and Kurtz, T. G. (1986). *Markov Processes: Characterization and Convergence.* Wiley, New York.

[25] Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004) *Bayesian Data Analysis*, Chapman & Hall-CRC, NY.

[26] Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. 6, pp. 721 - 741.

[27] Ghosh, S. K., Mukhopadhyay, P. and Lu, J. C. (2006). Bayesian analysis of zero-inflated regression models. *Journal of Statistical Planning and Inference.* vol 136. pp 1360-1375.

[28] Giannoudis, P. V. and Dinopoulos, H. (2005). Current concepts of the inflammatory response after major trauma: an update, *Injury*, vol. 36, no. 1, pp. 229-230.

[29] González, M., Martínez, R. and Mota, M. (2005a). On the geometric growth in a class of homogeneous multitype Markov chain, *J. Appl. Prob.* vol 42. pp 1015-1030.

[30] González, M., Martínez, R. and Mota, M. (2005b). On the unlimited growth of a class of homogeneous multitype Markov chains. *Bernoulli.* vol. 11. pp 559-570.

[31] González, M., Martínez, R. and del Puerto, I. (2004). Nonparametric estimation of the offspring distribution and the mean for a controlled branching process. *Test.* vol 13. pp 465-479.

[32] González, M., Martínez, R. and del Puerto, I. (2005). Estimation of the variance for a controlled branching process. *Test.* vol 14. pp 199-213.

[33] González, M., Molina, M. and del Puerto, I. (2002). On the class of controlled branching processes with random control functions. *J. Appl. Prob.* vol 39. pp 804-815.

[34] González, M., Molina, M. and del Puerto, I. (2003). On the geometric growth in controlled branching processes with random control function. *J. Appl. Prob.* vol 40. pp 995-1006.

[35] González, M., Molina, M. and del Puerto, I. (2004). Limiting distribution for subcritical controlled branching processes with random control function. *Stat. & Prob. Lett.* vol 67. pp 277-284.

[36] González, M., Molina, M. and del Puerto, I. (2005). Asymptotic behavior of critical controlled branching processes with random control functions. *J. Appl. Prob.* vol 42. pp 463-477.

[37] Hall, D. B. (2000). Zero-inflated Poisson and binomial regression with random effects: A case study. *Biometrics.* vol 56. pp 1030-1039.

[38] Hall, P. and Heyde, C. C., (1980). *Martingale Limit Theory and Its Application.* Academic Press, San Diego.

[39] Harris, C. J. and Stephens, M. (1988) A combined corner and edge detector, in *Proc. 4th Alvey Vision Conference*, Manchester, UK, pp. 147-151.

[40] Heyde, C. C., (1970). Extension of a result of Seneta for the supercritical Galton-Watson process. *Ann. Math. Statist.* vol 41. pp 739-742.

[41] Heyde, C. C. and Seneta, E., (1971). Analogous of classical limit theorems for the supercritical Galton-Watson process with immigration. *Math. Biosci.* vol 11. pp 249-259.

[42] Heyde, C. C. and Seneta, E., (1972). Estimation theory for growth immigration rates in a multiplicative process. *J. Appl. Prob.* vol 9. pp 235-258.

[43] Heyde, C. C. and Seneta, E., (1974). Notes on 'Estimation theory for growth immigration rates in a multiplicative process.' *J. Appl. Prob.* vol 11. pp 572-577.

[44] Heyde, C. C. and Brown, B. M. (1971). An invariance principle and some convergence rate results for branching processes. *Z. Wahrsch. Verw. Geb.* vol 20. pp 271-278.

[45] Jagers, P. (1975). *Branching Processes with Biological Applications*. Wiley, London.

[46] Jia, Y. and Jiang, Y. (2006) Active Contour Model with Shape Constraints for Bone Fracture Detection, in *Proc. of IEEE CGIV*, Sydney, Australia, pp. 90-95.

[47] Johnson, N. L., Kotz, S., and Kemp, A. W. (1992). *Univariate Discrete Distributions.* Second edition. John Wiley & Sons Inc.

[48] Kallenberg, O. (1997). *Foundations of Modern Probability.* Springer-Verlag, New York.

[49] Klimko, L. A. and Nelson, P. I., (1978). On conditional least squares estimation for stochastic processes. *Ann. Statist.* vol 6. pp 629-642.

[50] King, R.E., Scianna, J.M. and Petruzzelli, G.J. (2004). Mandible Fracture Patterns: a Suburban Trauma Center Experience, *Am. J. of Otolaryngol.*, vol. 25, no. 5, pp. 301-307.

[51] Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics.* vol 34. pp 1-14.

[52] Lee, S.J., Rangarajan, A. and Gindi, G. (1995). Bayesian Image Reconstruction in SPECT Using Higher Order Mechanical Models as Priors, *IEEE Trans. Med. Imag.*, vol. 14, no. 4, pp. 669 -680, Dec. 1995.

[53] Li, S. Z. (1995) *Markov random field modeling in computer vision*, Springer-Verlag London, UK.

[54] Lum, V., Leow, W., Chen, Y., Howe, T. and Png, M. (2005) Combining classifiers for bone fracture detection in X-ray images, in *Proc. of IEEE ICIP*, Genoa, Italy, pp. 1149 - 1152.

[55] Martin-Fernandez, M. and Alberola-Lopez, C. (2005). An approach for contour detection of human kidneys from ultrasound images using Markov random fields and active contours, *Med. Image Anal.*, vol. 9, pp. 1 - 23.

[56] McKendrick, A. G. (1926). Application of mathematics to medical problems. *Proc. Edin. Math. Soc.* vol 44. pp 98-130.

[57] Molina, R. (1994). On the Hierarchical Bayesian Approach to Image Restoration: Applications to Astronomical Images, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 11, pp. 1122-1128.

[58] Ogundare, B.O., Bonnick, A. and Bayley, N. (2003) Pattern of Mandibular Fractures in an Urban Major Trauma Center, *J. of Oral and Maxillofac. Surg.*, vol. 61, no. 6, pp. 713-718.

[59] Ozanian, T. and Phillips, R. (2000). Image Analysis for Computer-Assisted Surgery of Hip Fractures, *Med. Image Anal.*, vol. 4, no. 2, pp. 137 - 159.

[60] Popa, C. and Zdunek, R. (2005) Penalized Least-Squares Image Reconstruction For Borehole Tomography, in *Proc. of Algorithmy*, Podbanske, Slovakia pp. 260-269.

[61] Prima, S., Ourselin, S. and Ayache, N. (2002). Computation of the Mid-Sagittal Plane in 3D Brain Images, *IEEE Trans. Med. Imag.*, vol. 21, no. 2, pp. 122-138.

[62] Qi, Y. and Reeves, J. (2002). On sequential estimation for branching processes with immigration. *Stoch. Proc. Appl.* vol 100. pp 41-51.

[63] Raj, A., Singh, G. and Zabih, R. (2006). MRFs for MRIs: Bayesian Reconstruction of MR Images via Graph Cuts, in *Proc. of CVPR*, New York, NY, pp. 1061 - 1068.

[64] Rao, C. R. and Chakravarti, I. M. (1956). Some small sample tests of significance for a Poisson distribution. *Biometrics.* vol 12. pp 264-282.

[65] Richard, F. J. P. (2005). A comparative study of Markovian and variational image-matching techniques in application to mammograms, *Pattern Recognit. Lett.*, vol. 26, pp. 1819 - 1829.

[66] Rissanen, J. (1983). A universal prior for integers and estimation by minimum description length. *Ann. Statist.* vol. 11 no. 2, pp 416-431.

[67] Salli, E., Aronen, H. J., Savolainen, S., Korvenoja, A. and Visa, A. (2001). Contextual Clustering for Analysis of Functional MRI Data, *IEEE Trans. Med. Imag.*, vol. 20, no. 5, pp. 403-414.

[68] Self, S. G. and Liang, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association.* vol 82. pp 605-610.

[69] Sevast'yanov, B. A. and Zukov, A. (1974). Controlled Branching Processes. *Theor. Prob. Appl.* vol 19. Issue 1. pp 14-21.

[70] Shete, S and Sriram, T. N. (2003). A note on estimation in multitype supercritical branching processes with immigration. *Sankhyā.* vol 65. part 1. pp 107-121.

[71] Silvapulle, M. S. and Silvapulle, P. (1995). A score test against one-sided alternatives. *Journal of the American Statistical Association.* vol 90. no 429. pp 342-349.

[72] Smoluchowski, M. (1916). Drei vortrage uber diffusion Brownsche bewegung und wagulation von kolloidtelchen. *Physik. Zeits.* vol 17. pp 557-585.

[73] Sriram, T. N., Basawa, I. V. and Huggins, R. (1991). Sequential estimation for branching processes with immigration. *Ann. Statist.* vol. 19. pp 2232-2243.

[74] Syiam, M., El-Aziem, M. A. and El-Mensshawy, M. (2004) Adgen: Adaptive Interface Agent for X-Ray Fracture Detection, *Internat. J. of Comput. and Inform. Sci.*, vol. 2, no. 3, pp. 143 - 148.

[75] Venkataraman, K. N. (1982). A time series approach to the study of the simple subcritical Galton-Watson process with immigration. *Adv. Appl. Prob.* vol 14. pp 1-20.

[76] Venkataraman, K. N. and Nanthi, K. (1982). A limit theorem on a subcritical Galton-Watson process with immigration. *Ann. Prob.* vol 10. pp 1069-1074.

[77] Vodovotz, Y., Chow, C. C., Bartels, J., Lagoa, C., Prince, J. M., Levy, R. M., Kumar, R., Day, J., Rubin, J., Constantine, G., Billiar, T. R., Fink, M. P. and Clermont, G. (2006). In silico models of acute inflammation in animals, *Shock*, vol. 26, no. 3, pp. 235-44.

[78] Wei, C. Z. and Winnicki, J. (1989). Some asymptotic results for the branching process with immigration. *Stoch. Proc. Appl.* vol 31. pp 261-282.

[79] Wei, C. Z. and Winnicki, J. (1990). Estimation of the mean in the branching process with immigration. *Ann. Statist.* vol 18. pp 1757-1773.

[80] Yanev, N. M. (1975). Conditions for degeneracy of ϕ-branching processes with random ϕ. *Theor. Prob. Appl.* vol 28. pp 481-491

[81] Yap, D. W. H., Chen, Y., Leow, W. K., Howe, T. S. and Png, M. A. (2004) Detecting Femur Fractures by Texture Analysis of Trabeculae, in *Proc. of ICPR*, Cambridge, UK, pp. 730-733.