Summer 1997

Mark-Recapture Creel Survey and Survival Models

Shampa Saha
Old Dominion University

MARK-RECAPTURE CREEL SURVEY AND SURVIVAL MODELS

by

SHAMPA SAHA

M.Sc. May 1987, Bangalore University
Bangalore, India

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirement for the Degree of

DOCTOR OF PHILOSOPHY

COMPUTATIONAL AND APPLIED MATHEMATICS

OLD DOMINION UNIVERSITY
August 1997

Approved by:

Ram C. Dahiya (Director)

Dayanand N. Naik (Member)

Larry Lee (Member)

Edward P. Markowski (Member)

ABSTRACT

MARK-RECAPTURE CREEL SURVEY AND SURVIVAL MODELS

Shampa Saha
Old Dominion University, 1997
Director: Dr. Ram C. Dahiya

In this dissertation, we consider a model based approach to the estimation of exploitation rate of a fish population by combining mark-recapture procedures with a creel survey. We also consider the analysis of a proportional hazards survival model for randomly censored observations, known as the Koziol-Green model. The model assumes that the lifetime survivor function is a power of the censored time survivor function.

In Chapter 2, we introduce the model based approach to the estimation of the exploitation rate of a fish population by combining mark-recapture procedures with a creel survey. We assume that in the beginning of a fishing season M number of fishes are captured, marked and then released back into the population. Also, there are N available sampling units out of which n are sampled by the creel survey agents. The agents observe the number of fish captured with and without tags in each sampled unit. We make two basic assumptions, (1) the number captured in each unit follows a Poisson distribution and (2) the number recaptured given the number captured in each unit follows a binomial distribution. We obtain the maximum likelihood estimator and the moment estimator of the exploitation rate. We also compare the performance of these two estimators. In Chapter 3, the model based approach of Chapter 2 is extended to the case where the space-time units of the fisheries are stratified according to the different fishing areas or seasons.

In Chapter 4, we consider the Koziol-Green survival model with the lifetimes following a Weibull distribution. We consider the Bayesian analysis of this model by specifying parametric prior distributions for the parameters of the model. The method of Gibbs sampler is used for Bayesian computations. The Bayes estimator is compared with the maximum likelihood estimator of the survivor function. In Chapter 5, we incorporate covariates in the Koziol-Green model and study their effect on the lifetimes. We consider maximum likelihood estimation of the parameters of this model.

Dedicated

to

Baba

ACKNOWLEDGMENTS

I wish to express my gratitude to my advisor Dr. Ram C. Dahiya for his advice, guidance and critical review of this manuscript. I would like to thank my committee members Dr. Dayanand Naik and Dr. Larry Lee for their useful suggestions. I also thank Dr. Edward P. Markowski for serving on my dissertation committee.

I am very grateful to Dr. Ram C. Tiwari for his invaluable advice and help. My sincere thanks to Rodney Pierce and Andrew Bindman of the State of Minnesota, Department of Natural Resources for sharing their data with me. I would like to thank Hassan Lakkis for his time and effort in the initial stages of my research.

I wish to thank my friends Justine Shults, Alka and Surya Dhakar for their support and making my stay in Norfolk comfortable. I greatly appreciate the kindness and help given to me by Gayle Tarkelsen and Barbara Jeffrey.

Words cannot express my gratitude to my husband and best friend Akhil for his patience and many sacrifices. Finally, I would like to thank my parents Rekha and Bhabatosh Saha, parents-in-law Shankuntala and Shrikrishan Vaish, my brother Debashish, sister Sujata, sister-in-law Alpana and brother-in-law Amit for their love and constant encouragement.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Mark-Recapture Procedures

Capture-recapture or mark-recapture procedures are used to estimate population parameters in wildlife and fisheries sciences. Usually, the parameters of interest in such kind of studies are the population size, survival rate, mortality rate and exploitation rate (in case of fisheries studies). A typical mark-recapture experiment consists of sampling the population under study say k times, usually $k \geq 2$. Each time, an unmarked animal caught is uniquely marked and previously marked animals have their captures recorded and all animals are released back into the population. In some cases animals may be injured by capture, in which case they will not be released and are recorded as losses on capture. At the end of the study, we assume that the experimenter has the complete capture history of every animal handled.

There are several mark-recapture models classified as closed or open population models and they are defined as follows:

Closed Population: A closed population is one where permanent additions (births or immigrants) or deletions (deaths or emigrants) do not occur i.e., the population has a constant size during the entire study.

Open Population: An open population is one where permanent additions or deletions occur during the study.

A typical mark-recapture study provides two distinct types of information that can be used to estimate the parameters of interest. (1) Information obtained from the recapture of marked animals and (2) information obtained from comparing numbers of marked and unmarked animals captured at each sampling time. Data from (1) can be used to estimate survival rates, whereas data from (1) and (2) are necessary to estimate population sizes.

The Lincoln-Petersen Model

The Lincoln-Petersen model is the simplest form of mark-recapture models. Here a sample of n_1 animals is caught, marked and then released. Later, a sample of n_2 animals is captured, of which m_2 are marked animals. Then the estimator of the population size N is obtained based on the notion that the ratio of marked to the total animals in the second sample should reflect the same ratio in the population, so that

$$\frac{m_2}{n_2} = \frac{n_1}{N}$$

$$\Rightarrow \hat{N} = \frac{n_1 n_2}{m_2}.$$

The model makes the following assumptions:

(1) The population is closed to additions and deletions.

(2) All animals are equally likely to be captured in each sample.

(3) Marks are not lost and are not overlooked by the observers.

Several other closed population models have been developed by relaxing assumptions (2) and (3). Some time-dependent models have also been developed.

The Jolly-Seber Model

In many mark-recapture studies it is not possible to assume that the population is closed to permanent additions or deletions. In that case we have to consider open population models. The Jolly-Seber model is the most basic open population model.

The Jolly-Seber model allows estimation of population size at each sampling time as well as estimation of survival rates and number of births between sampling times.

The Jolly-Seber model requires the following assumptions:

(1) All animals in the population at a particular sampling time have the same probability of capture.

(2) Every marked animal present in the population immediately after a particular sampling time has the same probability of survival until the next sampling time.

(3) Marks are not overlooked or lost at any time.

(4) Samples are chosen instantaneously and each release is made immediately after sampling is completed.

The estimation of parameters under the Jolly-Seber model involves standard methods and the estimators are simple and intuitive. The population size estimators have the same form as in the Lincoln-Petersen model. Survival rate estimators are based on the ratios of the estimated numbers of marked animals at successive times. The number of additions between two sampling periods is a simple function of the two population size estimates and the survival rate for the same period. The Jolly-Seber model allows for injured animals not returned to the population.

Tag Return Models

Usually, it is of interest for government agencies to estimate the annual exploitation rate of a fish population for recreational and commercial fisheries. One may also be interested in estimating the annual survival rate, the natural mortality rate and the size of the fish population. In recent years, there have been several "tag return models" for fish tags returned by recreational or commercial fishers to estimate the above mentioned quantities. See Brownie et al. (1985) for a complete review of tag return models. Papers by Jagielo (1991) and Pollack et al. (1991) emphasize tag return models run in conjunction with angler surveys. Tags may be solicited from anglers on-site by a survey agent or voluntarily reported by anglers to a fishery agency. We first briefly review the existing tag return models presented in Brownie et al. (1985) and Pollack et al. (1991).

The paper by Pollack et al. (1991) forms the basis of this review. Let us consider the possible fates of a fish tagged at the beginning of the year:

S = the finite annual survival rate or the probability of surviving the year,

u = the annual exploitation rate or the probability of being harvested during the year,

λ = the tag reporting rate or the probability that a tag will be found and reported to the fisheries biologists, given that the fish has been harvested.

Furthermore, if we assume that all the fish killed are retrieved by the anglers, we have

$$v = 1 - S - u$$

as the finite natural mortality rate or the probability of dying from natural causes in the presence of fishing mortality. The data thus obtained supplies information directly only about harvested fish whose tags are reported. Hence the product $f = \lambda u$, the tag recovery rate, is estimable. We need additional information, such as that generated by reward tags or creel surveys or port samples to estimate λ and u. Consider the set up where we have multi-year taggings and recoveries and the animals are not stratified according to age-class. Following are the set of four models devised by Brownie et al. (1985):

Model 1: This is the most general model with the matrix of expected recovery numbers given in Table 1.1, for three tagging years and four recovery years. Here, for the ith year, S_i is the year-specific annual survival rate, f_i^* is the year-specific annual recovery rate for newly tagged fish and f_i is the year specific annual recovery rate for previously tagged fish. There may be a need to have separate recovery rates f_i^* and f_i for previously and newly tagged fish because fishing may begin before all tagging is completed, or the newly tagged fish may be more difficult to capture, or reporting rates may differ near and away from initial tagging sites.

Model 2: This is a special case of Model 1, where $f_i^* = f_i$ for all i. That is, all tagged fish irrespective of whether they are newly tagged or previously tagged have the same recovery rates in a given year.

Table 1.1: Matrix of Expected Recoveries.

	Expected number of recoveries in year			
Year, number tagged	1	2	3	4
1, N_1	$N_1 f_1^*$	$N_1 S_1 f_2$	$N_1 S_1 S_2 f_3$	$N_1 S_1 S_2 S_3 f_4$
2, N_2		$N_2 f_2^*$	$N_2 S_2 f_3$	$N_2 S_2 S_3 f_4$
3, N_3			$N_3 f_3^*$	$N_3 S_3 f_4$

Model 3: This is a special case of Model 2, where $S_i = S$ for all i. That is, all tagged fish have the same annual survival rate over all the years in the study.

Model 4: This is a special case of Model 3, where all tagged fish have constant annual survival and constant annual recovery rates over all the years in the study.

Model 1 is the most general model and Model 4 is the most restricted model. Brownie et al. (1985) provide a computer program, ESTIMATE, that determines the best model and estimates its survival and recovery rates, numerically. The following are the model assumptions: (1) The tagged sample is representative of the target population. If tagging takes place in areas with very heavy fishing pressure, it could give the appearance of high recovery and low survival rates for the entire region. Hence, to avoid this, tagging should be dispersed over a wide area of each region under study and should be in proportion to the population density in the area. Also, the assumption that tagged fish mix thoroughly throughout the whole area is usually unrealistic. (2) There is no tag loss. Nelson et al. (1980) through simulation studies found that tag loss produces a negative bias in survival and recovery rate estimates. This problem could be overcome by a double tagging study to estimate the tag loss and hence adjust for the survival and recovery rate estimates (see Seber 1982, page 94). (3) Survival rates are not influenced by tagging. If tagging increases mortality substantially then the survival estimates will not apply to untagged fish. (4) The year (fishing season) of tag recovery is correctly tabulated. If anglers report tags from fish caught in previous years, it produces a positive bias on survival estimates.

(5) The fate of each tagged fish is independent of the fate of other tagged fish. This assumption is violated in almost all practical applications of tag return models. This will not cause any model bias in any of the estimators, but it will mean that true sampling variances are larger than those given by the statistical models. (6) All tagged fish within an identifiable class have the same annual survival and recovery probabilities. Simulation studies of Nichols et al. (1982), Pollock and Ravelling (1982) revealed that if only recovery rates are heterogeneous, survival estimates are not biased and recovery rate estimates can be averaged for the population (if the tagging sample is random). If the survival probabilities are heterogeneous over the population then the survival rate estimators generally will have a negative bias, which is more serious when the average survival rate is high and the study is short. In theory, survival rate estimators could have a positive bias if segments of the population have markedly different survival rates but similar recovery rates.

Pollock et al. (1991) modified the models proposed by Brownie et al. (1985) by allowing tags to be solicited by survey agents or other biologists, which is a more realistic structure for fisheries studies. There is a certain unknown probability δ that a tag will be solicited. If we define the recovery rate of solicited tags as $f_s = u\delta$ and the recovery rate of unsolicited tags as $f_r = u(1-\delta)\lambda$, then S, f_s and f_r are estimable quantities and their estimates can be obtained using the program SURVIV (White, 1983). From these estimates we can obtain the estimate of the exploitation rate u, if we know or can estimate the reporting rate λ. Estimate of u is given by

$$\hat{u} = \hat{f}_s + \frac{\hat{f}_r}{\lambda}. \tag{1.1}$$

The expected value of $\hat{u}$ is

$$\begin{aligned} E(\hat{u}) &\approx u\delta + \frac{u(1-\delta)\lambda}{\lambda} \\ &= u\delta + u(1-\delta) \\ &= u. \end{aligned} \tag{1.2}$$

So $\hat{u}$ will be unbiased for large samples. Notice that it is not necessary to estimate δ because it drops out when we use the above equation. Estimation of λ is

crucial and we can achieve this in two different ways:

(i) Use of Tag Return Rewards: We can use two types of tags in a special study to estimate λ, one (control) that offers no reward and another one that offers reward for tag returns. This approach was developed by Henny and Burnham (1976) and Conroy and Blandin (1984). An important assumption here is that all recaptured fish with special reward tags are reported either voluntarily or via solicitation. Ideally, the reward notice should be displayed prominently on the tag so that it is not likely to be overlooked, although it may present operational difficulties in practical situations. Another assumption is that angler behavior does not change in response to the study. For more details see Pollock et al. (1991).

(ii) Use of Angler Surveys: We can estimate the tag reporting rate λ with an on-site angler survey. That is, we can use a creel survey or port sampling scheme. When the survey agent is interviewing anglers or commercial fishermen and checking their catch, we assume that the probability of a tag being reported is one; whereas, when the survey agent is not interviewing, the angler or commercial fishermen report the tags with probability λ ($0 \leq \lambda \leq 1$). The estimator of λ is

$$\hat{\lambda} = \frac{R_h}{\hat{R} - R_s} \tag{1.3}$$

with estimator of the variance of $\hat{\lambda}$ given by

$$\hat{V}(\hat{\lambda}) = \frac{\hat{\lambda}(1-\hat{\lambda})}{\hat{R} - R_s} + \frac{\hat{\lambda}(1-\hat{\lambda})\hat{V}(\hat{R})}{(\hat{R} - R_s)^3} \tag{1.4}$$

where

R_h = the number of tags recovered by anglers or commercial fishermen that are reported to the fisheries scientist in the absence of solicitation,

$\hat{R}$ = estimate of the total number of tags recovered by anglers or commercial fishermen,

R_s = number of tags recovered by anglers or commercial fishermen that were solicited by the survey agent and

$\hat{R} - R_s$ = the estimated number of tags recovered by anglers or commercial fishermen that are available to be reported with probability λ.

The total number of tags, R, recovered by anglers or commercial fishermen is estimated from a creel survey or port sample that runs concurrently with the tagging study. The method of estimation of R and its variance depends on the nature of the sampling scheme used. The fisheries scientist expands the number of tags found by the agent to the number that would have been found if the agents were present all the time, that is, the agents carry out a census of the fishery. For the derivation of the variance of $\hat{\lambda}$, see Pollock et al. (1991). The method depends on three important assumptions: (1) the agents and the anglers or commercial fishermen do not miss any tags on fish that are examined, (2) the solicited tags are all reported by the anglers or commercial fishermen and (3) the survey design is based on probability sampling so that the estimator of R does not suffer from model bias.

1.2 Survival Models

Radio Telemetry Survival Methods

Another field of tagging studies is radio-tagging which is becoming a popular method of following wild animals for a variety of purposes including survival analysis. An animal is captured by trap, dart gun or by some other method is fitted with a small radio transmitter and released. After release, the animal's unique radio signal can be monitored until the animal dies or its signal is lost which is a form of right censoring. As only marked animals are followed, it is usually possible to estimate only survival rates, not population sizes or recruitment rates. Traditional survival analysis methods like the Kaplan-Meier distribution free methods and the Cox's proportional hazards model have been used by Pollock et al. (1989).

Following Pollock (1984) and Pollock et al. (1989), we consider a random sample of n radio tagged animals. All tagged animals are monitored regularly so that the exact times of death are known. We assume that there is a fixed area to cover and if an animal with a functional radio is present, it is found with probability one. Let $T_1, \ldots, T_n$ form a set of survival times from tagging to death. We assume that these constitute a random sample from some probability distribution with density

function $f(t)$ and survivor function $S(t)$. Let $C_1, \ldots, C_n$ form a set of corresponding censoring times that constitutes a random sample from some probability distribution with density function $g(t)$ and survivor function $G(t)$. Let

$$Z_i = min(T_i, C_i) \text{ and } \delta_i = I(T_i \leq C_i), \quad i = 1, \ldots, n \tag{1.5}$$

where I is an indicator function which takes the value 0 if the observation is censored $(T_i > C_i)$ and 1 if it is not censored $(T_i \leq C_i)$.
The survivor function of the lifetimes is given by

$$S(t) = P(T > t). \tag{1.6}$$

The hazard function is the instantaneous rate of failure or death at time t given that the individual survives until time t and it is given by

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} \\ &= \frac{f(t)}{S(t)}. \end{aligned} \tag{1.7}$$

Then the likelihood function of the parameters involved in the model is given by

$$L = \prod_{i=1}^{n} [G(z_i)\, f(z_i)]^{\delta_i}\, [S(z_i)\, g(z_i)]^{1-\delta_i}. \tag{1.8}$$

If $g(t)$ and $G(t)$ do not involve any parameters of interest then the terms involving the censoring distribution can be ignored and part of the likelihood function given by

$$L_1 = \prod_{i=1}^{n} f(z_i)^{\delta_i}\, S(z_i)^{1-\delta_i} \tag{1.9}$$

is used to estimate the parameters involved in L_1.

If one assumes a parametric form for $f(t)$, standard maximum likelihood inference could be carried out. We can also find the nonparametric estimator of the survivor function which is the product-limit estimator (PLE), also known as the

Kaplan-Meier estimator. Suppose there are observations on n individuals and that there are k ($k \leq n$) distinct times $t_1 < t_2 < \cdots < t_k$ at which deaths occur. Let d_j be the number of deaths at time t_j and n_j be the number of individuals or animals at risk at time t_j, that is, the number of individuals alive and uncensored just prior to t_j. Then the Kaplan-Meier estimator of the survivor function is given by

$$\hat{S}(t) = \prod_{j:t_j<t} \frac{n_j - d_j}{n_j}. \tag{1.10}$$

The estimator for the variance of the Kaplan-Meier estimator of the survivor function is given by

$$\hat{V}[\hat{S}(t)] = [\hat{S}(t)]^2 \sum_{j:t_j<t} \frac{d_j}{n_j(n_j - d_j)}. \tag{1.11}$$

Mostly, nonparametric and semi-parametric survival methods have been considered for the analysis of radio-telemetry data.

The Proportional Hazards Model

In many situations, a biologist may be interested in the influence of important covariates on the survival process or on the lifetimes. An important class of models to study the association of covariates with survival times is the proportional hazards model (Cox, 1972). For this model the hazard function takes the form

$$h(t, x) = h_0(t)\, g(x) \tag{1.12}$$

where x is a vector of covariates; $h_0(t)$ is the baseline hazard function, i.e., the hazard function for an individual for whom $g(x) = 1$ when $x = 0$. Usually, $g(x)$ is assumed to be of the form

$$g(x) = exp(x'\beta) \tag{1.13}$$

where β a vector of unknown regression coefficients. Under this model, as the name suggests, the ratio of hazard functions for two individuals with specified vector of

covariates does not vary with time. In other words, different individuals have proportional hazard functions.

The proportional hazards model assumes that the covariates have a multiplicative effect on the hazard function. This assumption seems to hold in many situations. There are several methods in the literature to determine whether the proportional hazards model is appropriate for any particular situation. If we assume the baseline hazard function to be of parametric form (e.g. Weibull distribution, exponential distribution, etc.), then the proportional hazards models are known as parametric regression models.

1.3 Bayesian Methods

In the Bayesian approach to statistical inference problems, an effort is made to utilize all previously available information and combine it with new information to form the basis for statistical procedures. Bayes theorem describes the formal mechanism used to combine the new information with previously available information.

Bayes Theorem

Let $y' = (y_1, \ldots, y_n)$ be a vector of n observations whose probability distribution $p(y \mid \theta)$ depends on the values of k parameters $\theta' = (\theta_1, \ldots, \theta_k)$. Suppose θ has a probability density function (pdf) $p(\theta)$. Then

$$p(y \mid \theta)p(\theta) = p(y, \theta) = p(\theta \mid y)p(y). \tag{1.14}$$

Given the observed data y, the conditional distribution of θ is given by

$$p(\theta \mid y) = \frac{p(y \mid \theta)p(\theta)}{p(y)} \tag{1.15}$$

and

$$p(y) = E[p(y \mid \theta)] = c^{-1} = \begin{cases} \int p(y \mid \theta)\, p(\theta) d\theta & \text{if } \theta \text{ continuous} \\ \sum p(y \mid \theta)\, p(\theta) & \text{if } \theta \text{ discrete} \end{cases} \tag{1.16}$$

where the integral or sum is taken over the admissible range of θ and $E[f(\theta)]$ is the mathematical expectation of $f(\theta)$ with respect to θ. We can write (1.15) as

$$p(\theta \mid y) = c\, p(y \mid \theta)\, p(\theta). \tag{1.17}$$

The statements of (1.15) or (1.17) are usually known as the Bayes theorem. The pdf $p(\theta)$ is called the the prior distribution of θ and it tells us what is known about θ without any information from the data. In other words, $p(\theta)$ is the distribution *apriori*. Also $p(\theta \mid y)$ is called the posterior distribution of θ given y or the distribution *aposteriori* and it tells us what is known about θ given knowledge of the data. The quantity c is merely a normalizing constant necessary to ensure that the posterior distribution $p(\theta \mid y)$ integrates or sums to one.

Given the data y, $p(y \mid \theta)$ in (1.17) can be regarded as a function of θ, when so regarded, it is called the likelihood function of θ for given y and can be written as $L(\theta \mid y)$. Thus the Bayes theorem can be written as

$$p(\theta \mid y) \propto L(\theta \mid y)p(\theta). \tag{1.18}$$

The likelihood function $L(\theta \mid y)$ plays a very important role in Bayes theorem, since it is the function through which the data y modifies prior knowledge of θ.

Prior Distributions

The Bayesian approach depends on assigning probability distributions not only to data variables like y, but also to parameters like θ. Bayesian analysis is based on quantifying whatever feelings (however vague) we may have about θ having a pdf π, before having looked at data y. This distribution is then updated by the data via the Bayes theorem described by (1.15) and (1.17) with the resulting posterior distribution reflecting a blend of the information in the data and the prior distribution. So, it is of importance to determine an appropriate form of the prior distribution π.

Usually, prior distributions are specified based on information accumulated from past studies or from the opinions of subject-area experts. Two main methods of specifying the prior distributions are given below.

(1) We can restrict π to some familiar parametric distributional family $\pi(\theta|\eta)$, choosing η such that the result matches our true prior belief. If reasonable approximations to our prior belief could be managed by using a particular distribution which will lead to a posterior distribution of a nice form and which would also simplify computations to a great extent, then it would be sensible on our part to employ this distribution. It is this notion that leads to the concept of conjugate priors.

In many situations, it may be possible to select a member of that family which is conjugate to the likelihood function $L(\theta \mid y)$. In other words, choosing a prior distribution that leads to a posterior distribution $p(\theta \mid y)$ belonging to the same family of prior distributions. Such types of priors are known as *conjugate priors.*

(2) Many times, we may not be able to find reliable information on θ or we may prefer an inference based solely on data. Suppose we could find a distribution $\pi(\theta)$ that contained no information about θ or in other words, it did not favor one value of θ over another. We refer to such prior distributions as *noninformative priors* and we could infer that all of the information resulting in the posterior $p(\theta \mid y)$ arose from the data.

For example, suppose the parameter space is discrete and finite, i.e., $\Theta = \{\theta_1, \ldots \theta_l\}$ and $\pi(\theta) = 1/l$ for $i = 1, \ldots, l$. Clearly, $\pi(\theta)$ does not favor any one of the value of θ over any other value and in this sense it is noninformative for θ. Also, if we have a continuous bounded parameter space $\Theta = [c, d]$ where $-\infty < c < d < \infty$ and $\pi(\theta) = 1/(d - c)$ for $c < \theta < d$. Then $\pi(\theta)$ is the noninformative prior for θ.

The *Jeffreys prior* offers a fairly easy way of determining noninformative priors. In the univariate case, the prior is given by

$$\pi(\theta) = [I(\theta)]^{1/2} \tag{1.19}$$

where $I(\theta)$ is the expected Fisher information in the model, namely

$$I(\theta) = -E\left[-\frac{\partial^2 \log L(\theta \mid y)}{\partial \theta^2}\right]. \tag{1.20}$$

One can see that the form of the likelihood helps to determine the prior distribution in (1.19). There are several other methods for constructing informative and noninformative priors, see Berger (1985, Chapter 3).

In specifying prior distribution. Box and Tiao (1973. page 23) suggest that all that is important is that the data dominates whatever information is contained in the prior. As long as this happens, the precise form of the prior is not very important. Next, we give a brief description of the subsequent chapters of this dissertation.

In Chapter 2, we introduce a model based approach to the estimation of the exploitation rate of a fish population by combining mark-recapture procedures with a creel survey, which is an angler survey. In literature, most of the tag return models which run in conjunction with angler surveys are predominantly design based methods. We consider maximum likelihood estimation and moment estimation of the exploitation rate and also compare the two methods of estimation. Our model is for a closed population and it is a simple and useful model in the sense that we do not have to rely on voluntary tag returns or rewards.

In Chapter 3, we extend our model based approach to the case where we stratify the space-time units of the fisheries according to fishing seasons or different fishing areas. Stratification of the space-time units of the fisheries may be carried out for (1) administrative convenience, (2) gain in precision of the estimates of the population quantities or (3) it may be of interest to obtain the estimates of the parameters of the strata themselves.

In Chapter 4, we consider the analysis of a survival model for randomly censored observations where the lifetime survivor function is a power of the censored time survivor function with the lifetimes following a Weibull distribution. This survival model is also known as the Koziol-Green model and it is a proportional hazards model. The Koziol-Green survival model is useful in medical and reliability studies. This model could also be used for radio-telemetry survival data. We also consider the Bayesian analysis of this model by specifying parametric prior distributions for the parameters of the model. The Bayes estimates of the parameters of the model are obtained by the method of the Gibbs Sampler. We use simulated data to compare the maximum likelihood estimator (MLE) with the Bayes estimator of the survivor function. We show an application of this model to a real life medical data.

In Chapter 5. we incorporate covariates in the Koziol-Green survival model to

study the effect of covariates on the lifetimes. If the data is randomly censored and shows evidence that the lifetime survivor function is a power of the censored time survivor function then the model considered here would be appropriate to study the association of covariates with lifetimes. This model is particularly useful when there is a substantial amount of censored observations in the data set. We consider maximum likelihood estimation of the parameters of the model and we adopt a stepwise procedure to select the most important subset of covariates from the available set of covariates.

Note that, throughout this dissertation, the notation MLE will also be used to represent maximum likelihood estimate when working with the observed data. Similarly, PLE denotes product-limit estimator or estimate.

Chapter 2

Estimation of Exploitation Rate by Combining Mark-Recapture Procedures with a Creel Survey

2.1 Introduction

In this chapter, we introduce a model based approach to the estimation of the exploitation rate of a fish population by combining mark-recapture procedures with a creel survey. We derive the maximum likelihood estimator and the moment estimator of the exploitation rate and compare them through a simulation study.

We use the following notations in this chapter:

F=population size

M= number of fish marked

N= total number of sampling units

n= number of sampled units

c_i = number of fish caught in unit i

r_i = number of fish recaptured in unit i

R= number of fish recaptured in all N units

C= number of fish caught in all N units

$u = \frac{C}{F}$ = exploitation rate.

Note that $C = \sum_{i=1}^{N} c_i$, $R = \sum_{i=1}^{N} r_i$. Let $r. = \sum_{i=1}^{n} r_i$ and $c. = \sum_{i=1}^{n} c_i$. We define the following concepts:

Creel survey is an on-site intercept survey design where the survey agents interview anglers or commercial fishermen on-site about their catch.

Sampling unit is a portion of a day or an entire day in a fishing season.

Pierce et al. (1995) estimated the exploitation rate of native northern pike in seven small north central Minnesota lakes in view of the historical decline in the population of northern pike. They have combined a mark-recapture procedure with a creel survey to estimate the exploitation rate. They mark M fish at the beginning of the fishing season. During the season a simple random creel survey is done with a portion of a day as the sampling unit. The problem was treated as a two-stage sampling scheme with the mark-recapture experiment as stage 1 and the creel survey as stage 2. The number of fish caught in a sampling unit is assumed to be distributed as a Poisson random variable. The exploitation rate u is given by $u = C/F$. A standard estimate of u is $\tilde{u} = R/M$, but we cannot observe R. Hence, an estimate of u is $\hat{u} = \hat{R}/M$ where $\hat{R}$ is an estimator of R and it is given by

$$\hat{R} = N\frac{\sum_{i=1}^{n} r_i}{n} = \frac{Nr.}{n}$$

which is the simple random sample estimator of the population total.

The variance of $\hat{u}$ was calculated using the formula for a two stage sampling scheme (Cochran 1977, page 277),

$$V(\hat{u}) = E_1(V_2(\hat{u} \mid c.)) + V_1(E_2(\hat{u} \mid c.))$$

where V_1, E_1 are the variance and expectation under the mark-recapture and V_2, E_2 are the variance and expectation under the creel survey. They obtained variance of $\hat{u}$ as

$$V(\hat{u}) = \frac{u}{M}(\frac{N}{n} - u). \tag{2.1}$$

Here we adopt a model based approach to the estimation of the exploitation rate u for a fish population by combining a mark-recapture procedure with a creel survey under the same scenario as considered by Pierce et al. (1995). We assume that fish are tagged at the beginning of the season. We estimate the total number of fish caught and the total number of tags returned based on a simple random creel survey. This approach is especially useful when the fisheries biologists do not rely on voluntary tag returns or rewards.

We assume that there are N available sampling units where a sampling unit may be a day or a portion of a day in the fishing season. The creel agents choose a simple random sample of n sampling units and they observe the number of fish caught and the number recaptured in the sampled unit i. We make the following distributional assumptions:

(i) c_i are i.i.d. (independent and identically distributed) Poisson(λ) where λ is the average number of fish caught in unit i and it remains the same for all the sampling units.

(ii) $r_i|c_i$ is Binomial(c_i, p) where $p = \frac{M}{F}$ is the probability of a fish getting recaptured in unit i and we assume p to be the same in all the sampling units.

Due to the distributional assumptions that we have made above, we observe the following:

(a) If all N units are sampled, then R and C are sufficient statistics for λ and p.

(b) Based on a sample of size n, $r.$ and $c.$ are complete sufficient for p and λ giving UMVUE (uniformly minimum variance unbiased estimator) of p and λ as $\hat{p} = \frac{r.}{c.}$ and $\hat{\lambda} = \frac{c.}{n}$, respectively.

Note that $E(\hat{p}) = E_{c.}\left[E(\frac{r.}{c.}|c.)\right] = E_{c.}\left[\frac{c.p}{c.}\right] = p.$

(c) Based on the notion that the ratio of the total number of fish recaptured to the total number captured is same as the ratio of the total number of fish marked to the total number of fish in the population, we obtain $\tilde{u}$ as an estimator of u based on all N sampling units. Hence

$$\frac{M}{F} = \frac{R}{C} \Rightarrow \frac{C}{F} = \frac{R}{M} \Rightarrow \tilde{u} = \frac{R}{M}.$$

We can estimate F from the above relationship. Our main aim is to estimate u which boils down to estimating R and C.

(d) Since $c.$ and $r.$ are sufficient for λ and p, we obtain the joint distribution of $(C, c., R, r.)$ and use this to estimate R and C. Now the joint probability mass function of $(c., C-c., r., R-r.)$ is given by

$$f(c., C-c., r., R-r.) = \frac{e^{-n\lambda}(n\lambda)^{c.}}{c.!}\frac{e^{-(N-n)\lambda}[(N-n)\lambda]^{C-c.}}{(C-c.)!}\times \binom{c.}{r.}p^{r.}q^{c.-r.}\binom{C-c.}{R-r.}p^{R-r.}q^{C-c.-R+r.} \quad (2.2)$$

where $q = 1-p$. Note that $c.$ and $C-c.$ are independent, also $r.$ and $R-r.$ are independent. These facts are used in obtaining (2.2).

Note that (2.2) is also the joint probability mass function of (C, $c.$, R, $r.$). Since R and C are not observed, we need to find the conditional distribution of ($r.$, $c.$ | R, C). This is given by

$$g(r., c.|R, C) = \frac{f(c., C-c., r., R-r.)}{g_1(R|C)\, g_2(C)}$$

where $(R|C)$ is distributed as Binomial(C, p) and C is distributed as Poisson$(N\lambda)$. Then

$$g(r., c.|R, C) = \binom{C}{c.}\theta^{c.}(1-\theta)^{C-c.}\frac{\binom{c.}{r.}\binom{C-c.}{R-r.}}{\binom{C}{R}} \quad (2.3)$$

where $\theta = \frac{n}{N}$.

From (2.3), it is obvious that $c.|C$ is Binomial(C, θ) and $(r.|c., R, C)$ is hypergeometric. Also note that $c.$ and $r.$ are sufficient statistics for C and R. This can be shown by looking at the joint distribution of $(c_1, c_2, ..., c_n, r_1, r_2, ..., r_n|C, R)$.

In Section 2.2, we consider the estimation of the exploitation rate u. In Section 2.2.1, we obtain the maximum likelihood estimator of u. In Section 2.2.2, we consider the moment estimation of u. Section 2.2.3 gives the confidence interval for the exploitation rate u. In Section 2.3, we conduct a simulation study to compare the maximum likelihood estimator with the moment estimator of u and we also check the assumption of normality for the asymptotic distribution of the moment estimator of u. In Section 2.4, we analyze a real data set using our model based approach.

2.2 Estimation of the Exploitation Rate u

In this section, we obtain the MLE and the moment estimator of the exploitation rate u and also construct a confidence interval for u.

2.2.1 Maximum Likelihood Estimation of u

We have seen in Section 2.1 that the estimation of u and F boils down to estimating R and C. Therefore we concentrate on the maximum likelihood estimation of R and C and we do this in Theorem 2.1.

Theorem 2.1 *The maximum likelihood estimators of R and C are given by*

$$\tilde{\tilde{R}} = [\tilde{R}] \text{ and } \tilde{\tilde{C}} = [\tilde{C}], \text{ respectively}$$

(i) when $c_\cdot \neq r_\cdot$

$$\tilde{R} = N\frac{r_\cdot}{n} + \frac{r_\cdot}{c_\cdot - r_\cdot}, \tag{2.4}$$

$$\tilde{C} = N\frac{c_\cdot}{n} + \frac{r_\cdot}{c_\cdot - r_\cdot}, \tag{2.5}$$

(ii) when $c_\cdot = r_\cdot$

$$\tilde{R} = \tilde{C} = \frac{N}{n}c_\cdot \tag{2.6}$$

where $[\tilde{R}]$ denotes the greatest integer less than or equal to $\tilde{R}$ and the same definition holds for $[\tilde{C}]$.

Proof: To find the MLEs of R and C, we use the method of integer maximization of Dahiya (1981). Considering R and C as parameters to be estimated, the likelihood function of (R, C) is given by

$$L(R, C) = g(r_\cdot, c_\cdot | R, C) \tag{2.7}$$

where $g(r_\cdot, c_\cdot | R, C)$ is given by (2.3).

Case(i) $c_\cdot \neq r_\cdot$.

The method of integer maximization involves solving the two equations given by

$$L(R,C) = L(R-1,C) \tag{2.8}$$

and

$$L(R,C) = L(R,C-1). \tag{2.9}$$

Further simplification of $L(R,C)$ is given by

$$L(R,C) = a(1-\theta)^C \frac{R!\,(C-R)!}{(R-r.)!\,(C-c.-R+r.)!} \tag{2.10}$$

where a is a constant which is independent of R and C. Now from (2.8), we have

$$\begin{aligned} 1 &= \frac{L(R,C)}{L(R-1,C)} \\ &= \frac{R!\,(C-R)!}{(R-r.)!(C-c.-R+r.)!}\frac{(R-1-r.)!\,(C-c.-R+1+r.)!}{(R-1)!\,(C-R+1)!} \\ &= \frac{R(C-c.-R+1+r.)}{(C-R+1)(R-r.)} \end{aligned}$$

$$\Rightarrow \tilde{R} = \frac{r.}{c.}(\tilde{C}+1). \tag{2.11}$$

Similarly from (2.9), we have

$$\begin{aligned} 1 &= \frac{L(R,C)}{L(R,C-1)} \\ &= \frac{(1-\theta)^C\,(C-R)!}{(C-c.-R+r.)!}\frac{(C-1-c.-R+r.)!}{(1-\theta)^{C-1}\,(C-1-R)!} \\ &= \frac{(1-\theta)\,(C-R)}{(C-c.-R+r.)} \end{aligned}$$

$$\Rightarrow \theta(\tilde{C}-\tilde{R}) = (c.-r.). \tag{2.12}$$

Using (2.11) for $\tilde{R}$ in (2.12), we get

$$\tilde{C} = N\frac{c.}{n} + \frac{r.}{c.-r.}.$$

Finally, substituting for $\tilde{C}$ given in (2.5) in (2.11), we obtain

$$\tilde{R} = N\frac{r.}{n} + \frac{r.}{c.-r.}.$$

Case(ii) $c. = r.$

Note that C is an upper bound on R. Now $L(R,C) = a(1-\theta)^C R(R-1)...(R-c.+1)$. It is obvious that $L(R,C)$ increases as R increases. Hence $R = C$ will maximize $L(R,C)$ for a given C. Now using $R = C$ in $L(R,C)$, we have

$$L(C,C) = a(1-\theta)^C \frac{C!}{(C-c.)!}$$

$$1 = \frac{L(C,C)}{L(C-1,C-1)} \Rightarrow \tilde{C} = \frac{N}{n}c.$$

hence

$$\tilde{R} = \tilde{C} = \frac{N}{n}c.$$

Also note that if $c. = 0$ then

$$\tilde{R} = \tilde{C} = 0.$$

This completes the proof of the theorem.

Corollary 2.1 *The maximum likelihood estimator of the exploitation rate u is*

$$\tilde{u} = \frac{\tilde{\tilde{R}}}{M} = \frac{[\tilde{R}]}{M} \tag{2.13}$$

where $\tilde{R}$ is given by (2.4) in the case of $c. \neq r.$ and by (2.6) in the case of $c. = r.$.

Proof: The proof of the corollary follows from Theorem 2.1 and the definition of u.

2.2.2 Moment Estimation of u

In this section, we consider moment estimation of u. As we have noted earlier that estimation of u boils down to estimating C and R, hence in Theorem 2.2 we consider the moment estimation of C and R.

Theorem 2.2 *The moment estimator of C and its variance are given by*

$$\hat{C} = N\frac{c.}{n} \tag{2.14}$$

and

$$V(\hat{C}) = \frac{N^2\lambda}{n}. \tag{2.15}$$

respectively.

The moment estimator of R and its variance are given by

$$\hat{R} = N\frac{r.}{n} \tag{2.16}$$

and

$$V(\hat{R}) = \frac{N^2\lambda p}{n}, \tag{2.17}$$

respectively.

Proof: The moment estimators of C and R can be obtained from

$$C = E(C) = N\lambda \Rightarrow \hat{C} = N\hat{\lambda} = N\frac{c.}{n}$$

and

$$R = E(R) = E_C E(R|C) = E_C(Cp) = N\lambda p$$
$$\Rightarrow \hat{R} = N\hat{p}\hat{\lambda} = N\frac{r.}{n}.$$

Note that the moment estimators are not integers but can be rounded to the nearest integers.

Variances of $\hat{R}$ and $\hat{C}$ can be easily obtained as follows:

$$V(\hat{R}) = \frac{N^2}{n^2}V(r.) = \frac{N^2\lambda p}{n}$$

and

$$V(\hat{C}) = \frac{N^2}{n^2}V(c.) = \frac{N^2\lambda}{n},$$

respectively. The above results are based on the fact that the marginal distribution of $r.$ is Poisson$(n\lambda p)$ and that of $c.$ is Poisson$(n\lambda)$. Which follow from our assumptions that c_i are i.i.d. Poisson(λ) and $r_i|c_i$ is Binomial(c_i, p) for $i = 1, \ldots, n$. This completes the proof of the theorem.

Corollary 2.2 *The moment estimator of u is*

$$\hat{u} = \frac{1}{M}\frac{Nr.}{n} \tag{2.18}$$

and $\hat{u}$ is an unbiased estimator of u with variance

$$V(\hat{u}) = \frac{N^2\lambda}{MFn}. \tag{2.19}$$

Proof: Using the moment estimators of R and C which were obtained in Theorem 2.2 and $\hat{F} = \frac{M}{\hat{p}} = \frac{Mc.}{r.}$, we obtain

$$\hat{u} = \frac{\hat{C}}{\hat{F}} = \frac{\hat{R}}{M} = \frac{Nc.}{n}\frac{r.}{Mc.} = \frac{1}{M}\frac{Nr.}{n}.$$

For obtaining the mean and variance of $\hat{u}$, we only need to look at mean and variance of $r_.$. Consider,

$$\begin{aligned} E(r.) &= E_{c.}E(r.|c.) = E_{c.}(pc.) = n\lambda p \\ \Rightarrow E(\hat{u}) &= \frac{N}{Mn}n\lambda p = \frac{N\lambda p}{M} \\ \Rightarrow E(\hat{u}) &= \frac{N\lambda}{F}. \end{aligned} \tag{2.20}$$

Note that u is a random variable and

$$E(u) = \frac{E(C)}{F} = \frac{N\lambda}{F}. \tag{2.21}$$

Hence from (2.20) and (2.21), $\hat{u}$ is an unbiased estimator of u since $E(\hat{u} - u) = 0$.

Using the fact that the marginal distribution of $r.$ is Poisson$(n\lambda p)$, we can easily obtain $V(\hat{u})$ which is given by

$$V(\hat{u}) = \frac{N^2}{M^2n^2}V(r.) = \frac{N^2}{M^2n^2}n\lambda p = \frac{N^2}{M^2n}\frac{\lambda M}{F} = \frac{N^2\lambda}{MFn}.$$

This completes the proof of the corollary.

Since u is a random variable, we need to find $V(\hat{u} - u) = E(\hat{u} - u)^2$ which is derived in Theorem 2.3 given below.

Theorem 2.3

$$E(\hat{u} - u)^2 = \frac{N\lambda}{MF}[\frac{N}{n} - \frac{M}{F}]. \tag{2.22}$$

Proof: Let us consider

$$E(\hat{u} - u)^2 = E(\hat{u}^2) - 2E(\hat{u}u) + E(u^2).$$

Since

$$E(\hat{u}u) = E_u[uE(\hat{u}|u)] = E(u^2),$$

we have

$$E(\hat{u}-u)^2 = E(\hat{u}^2) - E(u^2). \tag{2.23}$$

Now, since $u = \frac{C}{F}$ and $\hat{u} = \frac{Nr.}{Mn}$, we have

$$E(u^2) = \frac{E(C^2)}{F^2} = \frac{N\lambda(N\lambda+1)}{F^2} \tag{2.24}$$

and

$$E(\hat{u}^2) = \frac{N^2}{M^2n^2}E(r.^2) = \frac{N^2}{M^2n^2}n\lambda p(1+n\lambda p). \tag{2.25}$$

From (2.23), (2.24) and (2.25), we have

$$\begin{aligned} E(\hat{u}-u)^2 &= \frac{N^2}{M^2n}\lambda\frac{M}{F}(1+n\lambda\frac{M}{F}) - \frac{N\lambda(N\lambda+1)}{F^2} \\ &= \frac{1}{F^2M^2n}[MN^2\lambda(F+n\lambda M) - M^2nN\lambda(N\lambda+1)] \\ &= \frac{MN\lambda}{F^2M^2n}[NF+n\lambda MN - MnN\lambda - Mn] \\ &= \frac{N\lambda}{nMF}[N - \frac{nM}{F}] \\ &= \frac{N\lambda}{MF}[\frac{N}{n} - \frac{M}{F}]. \end{aligned}$$

Therefore we have obtained (2.22) and this completes the proof of the theorem.

Next, in Theorem 2.4 we derive the conditional variance of $\hat{u}$ given u and show that its expected value is same as $E(\hat{u}-u)^2$.

Theorem 2.4 *The conditional variance of $\hat{u}$ given u is given by*

$$V(\hat{u}|u) = \frac{u}{M}\left[\frac{N}{n} - \frac{M}{C}u\right] \tag{2.26}$$

and

$$E[V(\hat{u}|u)] = E(\hat{u}-u)^2. \tag{2.27}$$

Proof: Consider

$$V(\hat{u}|u) = E(\hat{u}^2|u) - u^2. \tag{2.28}$$

Also,

$$E(\hat{u}^2|u) = \frac{N^2}{M^2n^2}E(r.^2|u) = \frac{N^2}{M^2n^2}E(r.^2|C) \tag{2.29}$$

and

$$E(r.^2|C) = E_R\left[[E(r.^2|R)]|C\right]. \tag{2.30}$$

Note that $(r.|R)$ is distributed as Binomial(R,θ) and $(R|C)$ is distributed as Binomial(C,p) where $\theta = \frac{n}{N}$ and $p = \frac{M}{F}$. We can write

$$E(r.^2|R) = R\frac{n}{N}(1-\frac{n}{N}) + R^2\frac{n^2}{N^2}. \tag{2.31}$$

Substituting (2.31) in (2.30), we get

$$\begin{aligned} E(r.^2|C) &= E_R[(R\frac{n}{N}(1-\frac{n}{N}) + R^2\frac{n^2}{N^2})|C] \\ &= C\frac{Mn}{FN}(1-\frac{n}{N}) + \frac{n^2}{N^2}[C\frac{M}{F}(1-\frac{M}{F}) + C^2\frac{M^2}{F^2}] \\ &= u\frac{nM}{N} - u\frac{M^2n^2}{FN^2} + u^2\frac{M^2n^2}{N^2} \end{aligned} \tag{2.32}$$

and substituting (2.32) in (2.29), we obtain

$$E(\hat{u}^2|u) = u\frac{N}{nM} - \frac{u}{F} + u^2. \tag{2.33}$$

Finally, from (2.33) and (2.28) we get

$$V(\hat{u}|u) = \frac{u}{M}\left[\frac{N}{n} - \frac{M}{F}\right] = \frac{u}{M}\left[\frac{N}{n} - \frac{M}{C}u\right].$$

Therefore, we have (2.26).

Now, in order to show (2.27), we have

$$E[V(\hat{u}|u)] = E(\frac{Nu}{Mn} - \frac{u}{F}) = \frac{N^2\lambda}{MnF} - \frac{N\lambda}{F^2} = \frac{N\lambda}{MF}\left[\frac{N}{n} - \frac{M}{F}\right] = E(\hat{u}-u)^2.$$

This completes the proof of the theorem.

As we have noted earlier, it is more appropriate to compute $E(\hat{u}-u)^2$ instead of variance of $\hat{u}$, so, it is important to obtain an unbiased estimator of $E(\hat{u}-u)^2$ and we do this in Theorem 2.5.

Theorem 2.5 *The unbiased estimator of* $E(\hat{u}-u)^2$ *is given by*

$$h(r.,c.)=\frac{Nr.}{nM^2}\left[\frac{N}{n}-\frac{r.-1}{c.-1}\right]. \tag{2.34}$$

Proof: Considering the facts that $(r.|c.) \sim$ Binomial$(c.,p)$, $r. \sim$ Poisson$(n\lambda p)$, $c. \sim$ Poisson$(n\lambda)$ and

$$E(\hat{u}-u)^2=\frac{N\lambda}{MF}(\frac{N}{n}-\frac{M}{F})=\frac{Np\lambda}{M^2}(\frac{N}{n}-p)=\frac{N}{M^2}(\frac{N}{n}p\lambda-p^2\lambda). \tag{2.35}$$

Now, it is clear that in order to obtain the unbiased estimator of $E(\hat{u}-u)^2$, we need to derive the unbiased estimators of $p\lambda$ and $p^2\lambda$.
Note that $\frac{r.}{n}$ is unbiased for $p\lambda$. Starting with

$$\hat{p}^2\hat{\lambda}=\frac{r.^2}{c.^2}\frac{c.}{n}=\frac{r.^2}{nc.},$$

we have

$$E\left(\frac{r.^2}{nc.}\right)=\frac{1}{n}E_{c.}\left[\frac{E(r.^2|c.)}{c.}\right]=\frac{1}{n}E_{c.}\left[\frac{(c.pq+c.^2p^2)}{c.}\right]=\lambda p^2+\frac{pq}{n} \tag{2.36}$$

where $q=1-p$, $\hat{p}=r./c.$ and $\hat{\lambda}=c./n$.

Therefore from (2.36), we can see that the bias in $(\frac{r.^2}{nc.})$ is $\frac{pq}{n}$ as an estimator of $p^2\lambda$.
Now $\frac{c.\hat{p}\hat{q}}{(c.-1)}$ is unbiased estimator of pq, since

$$E(\frac{c.\hat{p}\hat{q}}{c.-1})=E_{c.}[\frac{c.}{c.-1}E(\hat{p}\hat{q}|c.)]=E_{c.}(pq)=pq$$

where $\hat{q}=1-\hat{p}$. Note that $\frac{c.\hat{p}\hat{q}}{c.-1}=\frac{r.(c.-r.)}{c.(c.-1)}$, hence

$$\frac{r.^2}{nc.}-\frac{r.(c.-r.)}{nc.(c.-1)}=\frac{r.(r.-1)}{n(c.-1)} \tag{2.37}$$

is an unbiased estimator of $p^2\lambda$. Substituting the unbiased estimators of $p\lambda$ and $p^2\lambda$ in (2.35), the unbiased estimator of $E(\hat{u}-u)^2$ is given by (2.34). This completes the proof of the above theorem.

We can also show that $E[h(r.,c.)|C] = V(\hat{u}|u)$ and this is done in Theorem 2.6.

Theorem 2.6

$$E[h(r.,c.)|C] = V(\hat{u}|u) \tag{2.38}$$

where $h(r.,c.)$ is given by (2.34).

Proof: Note that $C \sim$Poisson$(N\lambda)$. Now simplifying (2.26), we have

$$V(\hat{u}|u) = \frac{Nu}{Mn} - \frac{u}{F} = \frac{NC}{MnF} - \frac{C}{F^2} \tag{2.39}$$

and simplifying (2.34), we have

$$h(r.,c.) = \frac{N^2 r.}{n^2M^2} - \frac{N}{nM^2}\frac{r.(r.-1)}{c.-1}. \tag{2.40}$$

Consider

$$E(r.|C) = E_{c.}[E(r.|c.)|C] = E_{c.}(c.\ p|C) = C\frac{n}{N}p = \frac{Mn}{N}u. \tag{2.41}$$

From (2.41), we obtain

$$E(\frac{N^2 r.}{M^2n^2}|C) = \frac{N}{Mn}u. \tag{2.42}$$

Consider

$$\begin{aligned} E\left[\frac{Nr.(r.-1)}{nM^2(c.-1)}|C\right] &= E_{c.}\left[\frac{N}{nM^2}E(\frac{r.(r.-1)}{(c.-1)}|c.)|C\right] \\ &= E_{c.}\left[\frac{N}{nM^2}\frac{c.pq + c.^2p^2 - c.p}{(c.-1)}|C\right] \\ &= \frac{N}{nM^2}E_{c.}\left[\frac{c.(p-p^2+c.p^2-p)}{c.-1}|C\right] \\ &= \frac{Np^2}{nM^2}E_{c.}[c.|C] = \frac{Np^2Cn}{nM^2N} = \frac{C}{F^2}. \end{aligned} \tag{2.43}$$

Therefore from (2.40), (2.42) and (2.43), we get

$$\begin{aligned} E[h(r.,c.)|C] &= E\left[\frac{N^2 r.}{n^2M^2}|C\right] - E\left[\frac{N}{nM^2}\frac{r.(r.-1)}{c.-1}|C\right] \\ &= \frac{Nu}{Mn} - \frac{u}{F} \\ &= V(\hat{u}|u). \end{aligned}$$

This completes the proof of the theorem.

Since p is expected to be very small, we expect $r./(c.-r.)$ to be a small fraction. Hence we can concentrate on the moment estimator instead of the maximum likelihood estimator of u. Moreover, it is easier to derive the distributional properties of $\hat{u}$.

2.2.3 Confidence Interval of u

From the results of Section 2.2.2, it follows that

$$\frac{\hat{u}-u}{\sqrt{E(\hat{u}-u)^2}} \to N(0,1) \text{ as } n \to \infty$$

and we can use this to find the asymptotic confidence interval of u. In fact, $\hat{u}$ is a constant multiple of $r.$ which in turn is a Poisson$(n\lambda p)$ variable. Hence, we know the exact distribution of $\hat{u}$. However, $\hat{u}-u$ has a complicated probability distribution since $r.$ is not independent of C which is involved in u.

Therefore a large sample $100(1-\alpha)\%$ confidence interval of u is given by

$$\hat{u} \pm z_{\frac{\alpha}{2}}\sqrt{h(r.,c.)} \tag{2.44}$$

where $h(r.,c.)$ is an unbiased estimator of $E(\hat{u}-u)^2$ given by (2.34) and $z_{\frac{\alpha}{2}}$ is the upper $100(\frac{\alpha}{2})$th percentile of the standard normal distribution.

2.3 Simulation Study

In this section, we present the results of a simulation study which compares the moment estimator with the MLE of u and we check the assumption of asymptotic

normality for the distribution of the statistic

$$X = \frac{\hat{u} - u}{\sqrt{E(\hat{u} - u)^2}} \tag{2.45}$$

where $E(\hat{u} - u)^2$ is given by (2.22). We also check the assumption of asymptotic normality for the distribution of the statistic

$$X' = \frac{\hat{u} - u}{\sqrt{h(r., c.)}} \tag{2.46}$$

where $h(r., c.)$ is an unbiased estimator of $E(\hat{u} - u)^2$ given by (2.34). Since X' is the statistic which will be considered in practice, we think it is of interest to study the validity of the distributional assumptions of X'.

We use the Shapiro-Wilk test to test the assumption of normality for the distribution of X and X' given by (2.45) and (2.46), respectively . Table 2.1 gives the biases of the moment estimator and MLE of u. It also gives the ratio of the mean square errors of the MLE and the moment estimator of u for different values of M, p, n and λ. Tables 2.2 and 2.3 give the values of the Shapiro-Wilk test statistic denoted by W and the corresponding p-values for different values of M, p, n and λ for X and X', respectively.

Note that $\hat{F} = M\frac{c.}{r.}$ and $E(\hat{F})$ is not finite since $E(\frac{1}{r.}|c.)$ is infinity. This can be avoided if we define $\hat{F} = F_0$, a fixed number, when $r. = 0$. Note that $P(r. = 0)$ is very small. However, variance of $\hat{F}$ can be computed. For simulations, $\hat{F}$ will have to be defined to be some large but fixed constant whenever $r. = 0$. Otherwise, $\hat{F} = \infty$ whenever $r. = 0$.

For the simulation study, we vary the values of M (number of fish marked or tagged), p (probability of capturing a tagged fish in any unit i), n (number of units sampled) and λ (the average catch rate in any unit i). We fix the total population size $F = 15000$ and the number of available sampling units $N = 150$.

We generate $c.$ from Poisson$(n\lambda)$, $C - c.$ from Poisson$((N - n)\lambda)$, $r.|c.$ from Binomial$(c., p)$ and $(R - r.)|(C - c.)$ from Binomial$(C - c., p)$. We generate $K = 1000$ such samples for a fixed F, M, N, n and λ. The MLE of u is computed using (2.13)

Table 2.1: Biases and Ratios of the MSEs of the MLE and Moment Estimator of u.

M	p	n	λ	$Bias(Moment)$	$Bias(MLE)$	$Ratio(MSEs)$
500	0.033	15	0.6	0.0000936	0.0000956	1.0018165
500	0.033	15	2	-0.0003	-0.0003	1
500	0.033	25	0.6	0.0000751	0 0.0000751	1
500	0.033	50	0.6	-0.000086	-0.000086	1
500	0.033	100	0.6	0.0004231	-0.000553	0.9734267
500	0.033	100	1	0.0005007	-0.000525	0.9899933
1000	0.067	15	0.6	0.0002534	0.0002534	1
1000	0.067	15	4	-0.000338	-0.000338	1
1000	0.067	25	1	-0.00025	-0.00025	1
1000	0.067	100	0.6	0.0001269	-0.000386	1.023187
1000	0.067	100	1	0.000015	0.000015	1
2500	0.17	15	0.6	-0.00015	-0.000138	1.0078171
2500	0.17	15	2	-0.00002	-0.00002	1
2500	0.17	25	0.6	-0.00008	-0.000078	1.0020531
2500	0.17	25	2	0.0000701	0.0000701	1
2500	0.17	50	0.6	-0.000071	-0.000071	1
2500	0.17	100	0.6	0.0000885	-0.000108	0.9998873
2500	0.17	100	2	0.0000789	0.0000789	1
3500	0.23	15	0.6	-0.000104	-0.000081	1.0157645
3500	0.23	15	4	0.0000405	0.0000405	1
3500	0.23	25	0.6	-0.00009	-0.000084	1.0062969
3500	0.23	25	2	0.0000485	0.0000485	1

Table 2.1(continued): Biases and Ratios of the MSEs of the MLE and Moment Estimator of u.

M	p	n	λ	$Bias(Moment)$	$Bias(MLE)$	$Ratio(MSEs)$
3500	0.23	50	0.6	-0.00005	-0.00005	1.0016363
3500	0.23	50	1	0.0000361	0.0000361	1
3500	0.23	75	0.5	-0.000077	-0.000077	1
3500	0.23	75	1	-0.000052	-0.000052	1
3500	0.23	100	0.6	0.0000575	-0.000075	1.0183756
3500	0.23	100	4	0.0000539	-0.000092	0.9979816
5000	0.33	15	0.6	-0.000146	-0.000098	1.0299431
5000	0.33	15	2	8.1333E-6	8.1333E-6	1.0036773
5000	0.33	25	0.6	-0.000066	-0.000041	1.0206465
5000	0.33	25	1	0.0000145	0.0000253	1.0092577
5000	0.33	50	0.6	-0.000045	-0.000038	1.0139726
5000	0.33	50	4	-0.000032	-0.000032	1
5000	0.33	75	0.6	-0.000015	-0.000012	1.0094264
5000	0.33	75	2	-5.6E-6	-5.6E-6	1
5000	0.33	100	0.6	0.0000451	-4.867E-6	1.0694747
5000	0.33	100	4	0.0000463	-5.067E-6	1.0255549

and the moment estimator of u is computed using (2.18). The biases of the moment estimator and MLE of u are

$$Bias(Moment) = \frac{1}{K}\sum_{j=1}^{K}(\hat{u}_j - u) \tag{2.47}$$

and

$$Bias(MLE) = \frac{1}{K}\sum_{j=1}^{K}(\tilde{u}_j - u), \tag{2.48}$$

respectively. The mean square errors (MSEs) of the moment estimator and MLE of u are

$$MSE(Moment) = \frac{1}{K}\sum_{j=1}^{K}(\hat{u}_j - u)^2 \tag{2.49}$$

and

$$MSE(MLE) = \frac{1}{K}\sum_{j=1}^{K}(\tilde{u}_j - u)^2, \tag{2.50}$$

respectively. The ratio of the two MSEs is

$$Ratio(MSEs) = \frac{MSE(MLE)}{MSE(Moment)}. \tag{2.51}$$

We can see from Table 2.1 that the moment estimator of u performs moderately better than the MLE of u whenever we sample large number of units ($n \geq 75$) and p is sufficiently large (at least 0.23). In most of the cases, both the estimators show negligible negative biases.

Table 2.2 shows that for a small $p = 0.033$, the distribution of X is close to $N(0,1)$ when n is at least 50 and λ is at least 8. When p is increased to 0.067, the distribution of X is close to $N(0,1)$ for $n = 25$ and $\lambda = 4$. For $p = 0.17$, the distribution of X is close to $N(0,1)$ for $n = 15$ and $\lambda = 4$. When p is increased to 0.23, the distribution of X is close to $N(0,1)$ for n as small as 15 and $\lambda = 2$. These results indicate that if p is small, it would be appropriate to sample a large number of units.

From Table 2.3, we can see that the convergence of distribution of X' given in (2.46) is much slower than X. For the distribution of X' to be close to normal, p has to be sufficiently large (at least 0.13) and one needs to sample a large number of units.

Table 2.2: Shapiro-Wilk Test of Normality for X.

M	p	n	λ	W	$p-value$
500	0.033	15	2	0.856427	0.0001
500	0.033	15	8	0.960543	0.0001
500	0.033	25	8	0.98006	0.0012
500	0.033	50	4	0.981252	0.0065
500	0.033	50	8	0.989722	0.9150
500	0.033	75	2	0.982177	0.0201
500	0.033	75	4	0.990694	0.9638
500	0.033	100	2	0.983107	0.0527
1000	0.067	15	8	0.981926	0.0150
1000	0.067	25	4	0.983289	0.0624
1000	0.067	50	2	0.983719	0.0908
1000	0.067	75	2	0.991274	0.9797
1000	0.067	100	1	0.984959	0.2220
2500	0.17	15	4	0.983955	0.1099
2500	0.17	25	2	0.98591	0.3700
2500	0.17	50	1	0.986603	0.4930
2500	0.17	75	0.6	0.983828	0.0993
2500	0.17	100	0.6	0.986823	0.5325
3500	0.23	15	2	0.98257	0.0308
3500	0.23	25	1	0.982103	0.0185
3500	0.23	50	1	0.986633	0.4983
3500	0.23	100	0.6	0.985634	0.3237

Table 2.3: Shapiro-Wilk Test of Normality for X'.

M	p	n	λ	W	$p-value$
500	0.033	15	8	0.922441	0.0001
500	0.033	25	10	0.895131	0.0001
500	0.033	75	12	0.96493	0.0001
500	0.033	75	25	0.977147	0.0001
1000	0.067	75	22	0.982095	0.0183
1000	0.067	100	15	0.981946	0.0154
2000	0.133	50	15	0.982186	0.0203
2000	0.133	75	10	0.982077	0.0179
2000	0.133	100	8	0.983315	0.0639
2500	0.17	50	13	0.982977	0.0465
2500	0.17	75	8	0.982376	0.0250
2500	0.17	100	5	0.981817	0.0132
3000	0.2	75	5	0.982076	0.0179
3500	0.23	50	8	0.983345	0.0656
3500	0.23	75	4	0.982013	0.0166
3500	0.23	100	4	0.98258	0.0311

2.4 Data Analysis

In this section, we analyze a data set using our model based method. The data was collected from Medicine lake, Minnesota by the State of Minnesota, Department of Natural Resources in 1988-1989 to study the exploitation of northern pikes. This data has been analyzed in the paper by Pierce et al. (1995). The data was collected from May, 1988 to February, 1989. The sampling unit is a 5 hour (summer) or 3 hour (winter) period. Pierce et al. (1995) used only the data from May to October and also considered two strata (weekday vs weekend and holiday). For our analysis we do not distinguish between the weekday and weekend or holiday. Table 2.4 gives the data that was collected from Medicine Lake. Table 2.5 gives the moment estimate of u and an estimate of its variance and also a 95% confidence interval of u. It is likely that we may obtain an under-estimate of the actual number of fish caught, since the creel clerks may concentrate more on monitoring tagged fish captured by the anglers.

Table 2.4: Data from Medicine Lake, Minnesota.

M	N	n	Observed captures ($c.$)	Observed recaptures ($r.$)
854	741	157	243	32

Table 2.5: Estimates of the Annual Exploitation Rate (u).

$\hat{u}$	0.1769
$\hat{E}(\hat{u}-u)^2 = h(r.,c.)$	0.0009509
95% Confidence interval of u	$(0.1164, 0.2373)$

Chapter 3

Estimation of Exploitation Rate by Stratifying the Space-Time Units of the Fishery

3.1 Introduction

In this chapter, we extend our model based approach of Chapter 2 to the case where we stratify the space-time units (N) of the fishery to obtain an estimate of the exploitation rate u. We may stratify the space-time units according to different fishing seasons (summer, fall, winter and spring). If we are estimating the exploitation rate of a fish population in a large water body (lake, pond, etc.) which spreads over a large area (perhaps over two or three neighboring states), then we would like to stratify the space-time units according to the different areas (states). Since each area will have their own fishery office and it would be of interest to obtain the exploitation rate estimate in that particular area (state) in addition to the estimate of the overall exploitation rate of the fish population. Similarly if stratification of the space-time units is done according to the fishing seasons then we would probably like to obtain

the exploitation rate estimates for each season in addition to the estimate of the annual exploitation rate.

In stratified sampling the population of N units is divided into L nonoverlapping strata of known sizes $N_1, N_2, \ldots, N_L$ such that together they comprise the whole of the population i.e., $N_1 + N_2 + \cdots + N_L = N$. There are many reasons why one may adopt a stratification scheme; some of the important reasons are: (1) If the population is heterogeneous then stratifying the population into homogeneous strata will reduce the variance of the population estimators and hence increase the precision of the estimators. (2) Stratification may be done because of administrative convenience; for example, the agency conducting the survey may have field offices, each of which can supervise the survey for a part of the population. (3) It may be of interest and importance to estimate the parameters for the strata themselves, hence yielding greater information.

If we adopt a stratified random sampling scheme i.e., simple random samples of size $n_1, n_2, \ldots, n_L$ are chosen from the L different strata, then the population mean μ can be estimated by

$$\bar{y}_{st} = \frac{\sum_{h=1}^{L} N_h \bar{y}_h}{N}$$

instead of the sample mean

$$\bar{y} = \frac{\sum_{h=1}^{L} n_h \bar{y}_h}{n}$$

where $\bar{y}_h$ is the sample mean of stratum h and $n = n_1 + \cdots + n_L$. We would prefer $\bar{y}_{st}$ as an estimator of the population mean, since in $\bar{y}_{st}$ the estimators from the individual strata receive their correct weights N_h/N (see Cochran 1977, page 91). The variance of $\bar{y}_{st}$ is given by

$$V(\bar{y}_{st}) = \sum_{h=1}^{L} \frac{N_h^2}{N^2} V(\bar{y}_h)$$

where $V(\bar{y}_h)$ is the variance of the simple random sample mean $\bar{y}_h$ from stratum h.

For our model discussed in Chapter 2, we now stratify the N available sampling units into L strata of sizes $N_1, N_2, \ldots, N_L$ such that $N_1 + N_2 + \cdots + N_L = N$. We choose simple random samples of sizes $n_1, n_2, \ldots, n_L$ from L different strata such

that $n_1 + n_2 + \cdots + n_L = n$. We observe the number of fish caught and the number of tagged fish captured in the sampled units in each stratum. Then the total catch and the total number of recaptures in each stratum are estimated and all the stratum estimates are pooled to obtain the estimate of the overall exploitation rate u.

We use the following notations in this chapter:

F = population size

M= number of fish marked

N_h= total number of sampling units in stratum h

n_h= number of units sampled in stratum h

r_{ih} = number of fish recaptured in unit i in stratum h

c_{ih} = number of fish captured in unit i in stratum h

R_h= number of fish recaptured in all the N_h units in stratum h

C_h= number of fish captured in all the N_h units in stratum h

C= total number of fish captured in all the L strata

R= total number of fish recaptured in all the L strata

$u = \frac{C}{F}$ = exploitation rate.

Note that $C_h = \sum_{i=1}^{N_h} c_{ih}$ and $R_h = \sum_{i=1}^{N_h} r_{ih}$.

Let $r_{.h} = \sum_{i=1}^{n_h} r_{ih}$, $c_{.h} = \sum_{i=1}^{n_h} c_{ih}$, $c_{..} = \sum_{h=1}^{L} \sum_{i=1}^{n_h} c_{ih}$,

$r_{..} = \sum_{h=1}^{L} \sum_{i=1}^{n_h} r_{ih}$, $C = \sum_{h=1}^{L} C_h$, $R = \sum_{h=1}^{L} R_h$, $N = \sum_{h=1}^{L} N_h$

and $n = \sum_{h=1}^{L} n_h$.

Simple random samples of n_h units are chosen from N_h units in stratum h and sampling is done independently in each stratum. We make the following assumptions:

(i) c_{ih} are i.i.d. Poisson(λ_h) for $h = 1, \ldots, L$.

(ii) $r_{ih}|c_{ih}$ is Binomial(c_{ih}, p) for $h = 1, \ldots, L$ where $p = \frac{M}{F}$.

We assume p the probability of a recapture to be the same for all the units in all the strata.

We note the following:

(a) $r_{.h}$ and $c_{.h}$ are complete and sufficient for p and λ_h.

(b) $\hat{p} = \frac{r_{..}}{c_{..}}$ and $\hat{\lambda}_h = \frac{c_{.h}}{n_h}$ are UMVUE of p and λ_h, respectively. Also $\tilde{p} = \frac{r_{.h}}{c_{.h}}$ is the UMVUE of p based on the fact that $r_{.h}|c_{.h}$ is also Binomial$(c_{.h}, p)$.

Note that

$$E(\hat{p}) = E_{c_{..}}\left[E(\frac{r_{..}}{c_{..}}|c_{..})\right] = E_{c_{..}}\left[\frac{c_{..}p}{c_{..}}\right] = p, \tag{3.1}$$

also

$$E(\tilde{p}) = E_{c_{.h}}\left[E(\frac{r_{.h}}{c_{.h}}|c_{.h})\right] = E_{c_{.h}}\left[\frac{c_{.h}p}{c_{.h}}\right] = p. \tag{3.2}$$

(c) Our main aim is to estimate u which boils down to estimating R_h and C_h for $h = 1, \ldots, L$.

Since

$$u = \sum_{h=1}^{L} \frac{C_h}{F}$$

and

$$\frac{M}{F} = \frac{R}{C} = \frac{\sum_{h=1}^{L} R_h}{\sum_{h=1}^{L} C_h}$$

$$\Rightarrow \frac{\sum_{h=1}^{L} C_h}{F} = \frac{\sum_{h=1}^{L} R_h}{M},$$

hence

$$\hat{u} = \frac{\sum_{h=1}^{L} \hat{R}_h}{M}. \tag{3.3}$$

(d) Since $c_{.h}$ and $r_{.h}$ are sufficient for λ_h and p respectively, we obtain the joint distribution of $(C_h, c_{.h}, R_h, r_{.h})$ and use this to estimate R_h and C_h. The joint distribution of $(c_{.h}, C_h - c_{.h}, r_{.h}, R_h - r_{.h})$ is given by

$$f(c_{.h}, C_h - c_{.h}, r_{.h}, R_h - r_{.h}) = \frac{e^{-n_h\lambda_h}(n_h\lambda_h)^{c_{.h}}}{c_{.h}!} \frac{e^{-(N_h-n_h)\lambda_h}[(N_h-n_h)\lambda_h]^{C_h-c_{.h}}}{(C_h-c_{.h})!} \times \binom{c_{.h}}{r_{.h}} p^{r_{.h}} q^{c_{.h}-r_{.h}} \binom{C_h-c_{.h}}{R_h-r_{.h}} p^{R_h-r_{.h}} q^{C_h-c_{.h}-R_h+r_{.h}} \tag{3.4}$$

where $q = 1 - p$.

Note that $c_{.h}$ and $C_h - c_{.h}$ are independent. Also $r_{.h}$ and $R_h - r_{.h}$ are independent. These facts are used in obtaining (3.4).

Note that (3.4) also gives the joint probability mass function of $(C_h,\ c_{.h},\ R_h,\ r_{.h})$. Since, R_h and C_h are not observed, we need to find the conditional distribution of $(r_{.h},\ c_{.h} \mid R_h,\ C_h)$. This is given by

$$h(r_{.h}, c_{.h}|R_h, C_h) = \frac{f(c_{.h}, C_h - c_{.h}, r_{.h}, R_h - r_{.h})}{g_1(R_h|C_h)\, g_2(C_h)}$$

where $(R_h|C_h)$ is distributed as Binomial(C_h, p) and C_h is distributed as Poisson$(N_h\lambda_h)$. Then

$$h(r_{.h}, c_{.h}|R_h, C_h) = \binom{C_h}{c_{.h}} \theta_h^{c_{.h}} (1-\theta_h)^{(C_h - c_{.h})} \frac{\binom{c_{.h}}{r_{.h}}\binom{C_h - c_{.h}}{R_h - r_{.h}}}{\binom{C_h}{R_h}} \tag{3.5}$$

where $\theta_h = \frac{n_h}{N_h}$.

From (3.5), it is obvious that $c_{.h}|C_h$ is Binomial(C_h, θ_h) and $(r_{.h}|c_{.h}, R_h, C_h)$ is hypergeometric. Also note that $c_{.h}$ and $r_{.h}$ are sufficient statistics for C_h and R_h.

In Section 3.2, we consider maximum likelihood estimation of u under stratification and in Section 3.3, we consider the moment estimation of u. In Section 3.4, we present the results of a simulation study similar to the study done in Chapter 2. In Section 3.5, we analyze the data set considered in Chapter 2 by stratifying the data into four strata.

3.2 Maximum Likelihood Estimation of u

As we have noted in Chapter 2, finding the MLE of u boils down to finding the MLEs of R_h and C_h for $h = 1, \ldots, L$. Considering R_h and C_h for $h = 1, \ldots, L$ as parameters to be estimated, we find MLEs by integer maximization of (3.5) with respect to R_h and C_h.

In order to use the method of Dahiya (1981) for integer maximization, we write

$$L(R_h, C_h) = h(r_{.h}, c_{.h}|R_h, C_h)$$

where $L(R_h, C_h)$ is the likelihood function of (R_h, C_h). The integer maximization involves solving the two equations given by

$$L(R_h, C_h) = L(R_h - 1, C_h)$$

and

$$L(R_h, C_h) = L(R_h, C_h - 1).$$

Following the same steps as in Theorem 2.1, the MLEs of C_h and R_h for $h = 1, \ldots, L$ when $c_{.h} \neq r_{.h}$ are given by

$$\tilde{C}_h = N_h \frac{c_{.h}}{n_h} + \frac{r_{.h}}{c_{.h} - r_{.h}} \tag{3.6}$$

and

$$\tilde{R}_h = N_h \frac{r_{.h}}{n_h} + \frac{r_{.h}}{c_{.h} - r_{.h}}. \tag{3.7}$$

When $c_{.h} = r_{.h}$, note that C_h is an upper bound on R_h. Now $L(R_h, C_h) = a(1-\theta_h)^{C_h} R_h(R_h - 1)\ldots(R_h - c_{.h} + 1)$ where a does not depend on C_h and R_h. It is obvious that $L(R_h, C_h)$ increases as R_h increases. Hence $R_h = C_h$ will maximize $L(R_h, C_h)$ for a given C_h. Now using $R_h = C_h$ in $L(R_h, C_h)$, we have

$$L(C_h, C_h) = a(1-\theta_h)^{C_h} \frac{C_h!}{(C_h - c_{.h})!}$$

$$1 = \frac{L(C_h, C_h)}{L(C_h - 1, C_h - 1)} \Rightarrow \tilde{C}_h = \frac{N_h}{n_h} c_{.h}.$$

Hence

$$\tilde{R}_h = \tilde{C}_h = \frac{N_h}{n_h} c_{.h}. \tag{3.8}$$

Also, if $c_{.h} = 0$ then

$$\tilde{R}_h = \tilde{C}_h = 0.$$

Note that the integer MLE $\tilde{\tilde{C}}_h$ of C_h and $\tilde{\tilde{R}}_h$ of R_h are given by $\tilde{\tilde{C}}_h = [\tilde{C}_h]$ and $\tilde{\tilde{R}}_h = [\tilde{R}_h]$, respectively. Where $[a]$ denotes the greatest integer less than or equal to a and $\tilde{C}_h$, $\tilde{R}_h$ are given by (3.6), (3.7) and (3.8).

Hence the MLE of u is given by

$$\tilde{u} = \frac{1}{M} \sum_{h=1}^{L} \tilde{\tilde{R}}_h. \tag{3.9}$$

3.3 Moment Estimation of u

Here we consider the moment estimation of u under stratification. As we have seen previously, to obtain the moment estimator of u, we need to first obtain the moment estimators of C_h and R_h for $h = 1, \ldots, L$.

The moment estimators of C_h and R_h can be obtained from

$$C_h = E(C_h) = N_h\lambda_h \Rightarrow \hat{C}_h = N_h\hat{\lambda}_h = N_h\frac{c_{.h}}{n_h} \tag{3.10}$$

and

$$R_h = E(R_h) = E_{C_h}E(R_h|C_h) = E_{C_h}(C_hp) = N_hp\lambda_h$$
$$\Rightarrow \hat{R}_h = N_h\hat{\tilde{p}}\hat{\lambda}_h \tag{3.11}$$

where $\hat{\tilde{p}}$ is an estimator of p. Note that the moment estimators are not integers but can be rounded to the nearest integers.

Since p is expected to be very small, we expect $r_{.h}/(c_{.h} - r_{.h})$ to be a small fraction. Ignoring this fraction, we can concentrate on moment estimators instead of MLEs of R_h and C_h. We can obtain two unbiased moment estimators of u depending on what we choose as an estimator for p.

Case(i) If we choose $\tilde{p} = r_{.h}/c_{.h}$ to be an estimator of p then an unbiased estimator of u is

$$\hat{u}_1 = \frac{1}{M}\sum_{h=1}^{L}\frac{N_h}{n_h}r_{.h} \tag{3.12}$$

and

$$\begin{aligned} E(\hat{u}_1) &= E\left(\frac{1}{M}\sum_{h=1}^{L}\frac{N_h}{n_h}r_{.h}\right) \\ &= \frac{1}{M}\sum_{h=1}^{L}N_h\lambda_h\, p \\ &= E(u). \end{aligned} \tag{3.13}$$

Note that u is a random variable and

$$E(u) = \frac{1}{F}\sum_{h=1}^{L}E(C_h) = \frac{1}{F}\sum_{h=1}^{L}N_h\lambda_h = \frac{1}{M}\sum_{h=1}^{L}N_h\lambda_hp. \tag{3.14}$$

Hence, $\hat{u}_1$ is an unbiased estimator of u since $E(\hat{u}_1 - u) = 0$. It can be shown that the marginal distribution of $r_{.h}$ is Poisson$(n_h\lambda_h p)$. As in the case of without stratification, we can derive

$$E(\hat{u}_1 - u)^2 = E(\hat{u}_1^2) - 2E(\hat{u}_1 u) + E(u^2)$$

but

$$E(\hat{u}_1 u) = E_u[uE(\hat{u}_1|u)] = E(u^2),$$

hence

$$E(\hat{u}_1 - u)^2 = E(\hat{u}_1^2) - E(u^2). \tag{3.15}$$

Now, since

$$u = \frac{\sum_{h=1}^{L} C_h}{F}$$

and

$$\hat{u}_1 = \frac{1}{M}\sum_{h=1}^{L}\frac{N_h}{n_h} r_{.h},$$

we have

$$\begin{aligned} E(u^2) &= E(\frac{1}{F}\sum_{h=1}^{L} C_h)^2 \\ &= \frac{1}{F^2}E(\sum_{h=1}^{L} C_h^2 + 2\sum_{h=1}^{L-1}\sum_{k=h+1}^{L} C_h C_k) \\ &= \frac{1}{F^2}(\sum_{h=1}^{L} N_h\lambda_h(N_h\lambda_h + 1) + 2\sum_{h=1}^{L-1}\sum_{k=h+1}^{L} N_h N_k \lambda_h \lambda_k). \end{aligned} \tag{3.16}$$

It can be easily shown that

$$E(\hat{u}_1^2) = \frac{1}{M^2}(\sum_{h=1}^{L}\frac{N_h^2}{n_h}\lambda_h p(n_h\lambda_h p + 1) + 2\sum_{h=1}^{L-1}\sum_{k=h+1}^{L} N_h N_k \lambda_h \lambda_k p^2). \tag{3.17}$$

Using (3.15), (3.16) and (3.17) we get

$$E(\hat{u}_1 - u)^2 = \frac{1}{MF}\sum_{h=1}^{L} N_h\lambda_h[\frac{N_h}{n_h} - \frac{M}{F}]. \tag{3.18}$$

We can derive the unbiased estimator of $E(\hat{u}_1 - u)^2$ on the same lines as we did in Theorem 2.5. Hence the unbiased estimator of $E(\hat{u}_1 - u)^2$ is

$$h_1(r_{.h}, c_{.h}) = \frac{1}{M^2}\sum_{h=1}^{L}\frac{N_h}{n_h} r_{.h}(\frac{N_h}{n_h} - \frac{r_{.h} - 1}{c_{.h} - 1}). \tag{3.19}$$

As in Section 2.2.3, without stratification case, a large sample 100(1-α)% confidence interval of u is given by

$$\hat{u}_1 \pm z_{\frac{\alpha}{2}}\sqrt{h_1(r_{.h}, c_{.h})}$$

where $h_1(r_{.h}, c_{.h})$ is given by (3.19).

Case(ii) If we choose $\hat{p} = \frac{r_{..}}{c_{..}}$ to be an estimator of p then the estimator of u is given by

$$\hat{u}_2 = \frac{r_{..}}{Mc_{..}} \sum_{h=1}^{L} \frac{N_h}{n_h} c_{.h} \tag{3.20}$$

and

$$\begin{aligned} E(\hat{u}_2) &= E(\frac{r_{..}}{M\, c_{..}} \sum_{h=1}^{L} \frac{N_h}{n_h} c_{.h}) \\ &= \frac{1}{M} \sum_{h=1}^{L} \frac{N_h}{n_h} E(\frac{r_{..}\, c_{.h}}{c_{..}}) \\ &= \frac{1}{M} \sum_{h=1}^{L} \frac{N_h}{n_h} E_{c_{..}} E(\frac{r_{..}\, c_{.h}}{c_{..}} | c_{..}) \\ &= \frac{1}{M} \sum_{h=1}^{L} \frac{N_h}{n_h} E_{c_{..}} \frac{1}{c_{..}} E(r_{..}\, c_{.h} | c_{..}) \\ &= \frac{1}{M} \sum_{h=1}^{L} N_h \lambda_h\, p \\ &= E(u), \end{aligned} \tag{3.21}$$

since $c_{.h}|c_{..}$ is Binomial($c_{..}$, $\frac{n_h\lambda_h}{\sum_{h=1}^{L} n_h\lambda_h}$). Hence $\hat{u}_2$ is also an unbiased estimator of u.

It is more appropriate to use $\hat{p}$ rather than $\tilde{p}$ as an estimator of p since we are assuming the probability of recapture to be the same in all the strata. Hence we prefer to use $\hat{u}_2$ rather than $\hat{u}_1$ as an estimator of the exploitation rate u. Also, simulation results in Section 3.4 show that $\hat{u}_2$ has a smaller mean square error (MSE) as compared to $\hat{u}_1$.

As in the case of without stratification, we derive $E(\hat{u}_2 - u)^2$ rather than the variance of $\hat{u}_2$. In Theorem 3.1 we obtain $E(\hat{u}_2 - u)^2$.

Theorem 3.1

$$E(\hat{u}_2 - u)^2 = \frac{1}{M^2}[p(1-p)(\lambda-1)(\sum_{h=1}^{L} a_h p_{.h})^2 + p(1-p)\sum_{h=1}^{L} a_h^2 p_{.h} + \lambda p^2 \sum_{h=1}^{L} a_h(a_h-1)p_{.h}] \quad (3.22)$$

where

$$a_h = \frac{N_h}{n_h},$$

$$p_{.h} = \frac{n_h\lambda_h}{\sum_{h=1}^{L} n_h\lambda_h},$$

$$\lambda = \sum_{h=1}^{L} n_h\lambda_h$$

and $\hat{u}_2$ *is given in (3.20).*

Proof: Note that

$$E(\hat{u}_2 - u)^2 = E(\hat{u}_2^2) - E(u^2), \quad (3.23)$$

$c_{..}$ is Poisson($\sum_{h=1}^{L} n_h\lambda_h$), $r_{..}|c_{..}$ is Binomial($c_{..}, p$) and $c_{.h}|c_{..}$ is Binomial($c_{..}, \frac{n_h\lambda_h}{\sum_{h=1}^{L} n_h\lambda_h}$).

Hence

$$\begin{aligned} E(\hat{u}_2^2) &= E(\frac{r_{..}}{Mc_{..}}\sum_{h=1}^{L}\frac{N_h}{n_h}c_{.h})^2 \\ &= \frac{1}{M^2}E(\sum_{h=1}^{L}\frac{N_h}{n_h}\frac{r_{..}}{c_{..}}c_{.h})^2 \\ &= \frac{1}{M^2}E(\sum_{h=1}^{L}\frac{N_h^2}{n_h^2}\frac{r_{..}^2}{c_{..}^2}c_{.h}^2 + 2\sum_{h=1}^{L-1}\sum_{k=h+1}^{L}\frac{N_h}{n_h}\frac{N_k}{n_k}\frac{r_{..}^2}{c_{..}^2}c_{.h}c_{.k}) \\ &= \frac{1}{M^2}[\sum_{h=1}^{L}\frac{N_h^2}{n_h^2}E(\frac{r_{..}^2}{c_{..}^2}c_{.h}^2) + 2\sum_{h=1}^{L-1}\sum_{k=h+1}^{L}\frac{N_h}{n_h}\frac{N_k}{n_k}E(\frac{r_{..}^2}{c_{..}^2}c_{.h}c_{.k})]. \end{aligned} \quad (3.24)$$

But

$$\begin{aligned} E\left(\frac{r_{..}^2}{c_{..}^2}c_{.h}^2\right) &= E_{c_{..}}\frac{1}{c_{..}^2}E\left(c_{.h}^2 r_{..}^2|c_{..}\right) \\ &= E_{c_{..}}\frac{1}{c_{..}^2}[(c_{..}p_{.h}(1-p_{.h}) + c_{..}^2 p_{.h}^2)(c_{..}p(1-p) + c_{..}^2 p^2)] \\ &= E[p_{.h}(1-p_{.h})p(1-p) + c_{..}(p^2 p_{.h}(1-p_{.h}) + p_{.h}^2 p(1-p)) + c_{..}^2 p_{.h}^2 p^2] \\ &= (p_{.h}(1-p_{.h}) + \lambda p_{.h}^2)(p(1-p) + \lambda p^2) + \lambda p_{.h}^2 p^2. \end{aligned} \quad (3.25)$$

To obtain (3.25), we have made use of the fact that given $c_{..}$, $r_{..}$ and $c_{.h}$ are independently distributed. Next we derive

$$\begin{aligned} E\left(\frac{r_{..}^2}{c_{..}^2}c_{.h}c_{.k}\right) &= E_{c_{..}}\frac{1}{c_{..}^2}E\left(c_{.h}c_{.k}r_{..}^2|c_{..}\right) \\ &= E_{c_{..}}\frac{1}{c_{..}^2}[(c_{..}^2 p_{.h}\, p_{.k} - c_{..}\, p_{.h}\, p_{.k})(c_{..}\, p(1-p) + c_{..}^2\, p^2)] \\ &= E[-p_{.h}p_{.k}p(1-p) + c_{..}(-p^2 p_{.h}p_{.k} + p_{.h}p_{.k}p(1-p)) + c_{..}^2\, p_{.h}p_{.k}p^2] \\ &= (p(1-p) + \lambda p^2)(\lambda p_{.h}p_{.k} - p_{.h}p_{.k}) + \lambda p^2 p_{.h}p_{.k}. \end{aligned} \tag{3.26}$$

In deriving (3.26) we have used the fact that $(c_{.h}, c_{.k}|c_{..})$ and $(r_{..}|c_{..})$ are independently distributed. Also for given $c_{..}$, the joint probability mass function of $c_{.h}$ and $c_{.k}$ is Multinomial$(c_{..}, p_{.h}, p_{.k})$. From (3.24), (3.25) and (3.26) we get

$$\begin{aligned} E(\hat{u}_2^2) &= \frac{1}{M^2}[\sum_{h=1}^{L} a_h^2 E(\frac{r_{..}^2}{c_{..}^2}c_{.h}^2) + 2\sum_{h=1}^{L-1}\sum_{k=h+1}^{L} a_h a_k E(\frac{r_{..}^2}{c_{..}^2}c_{.h}c_{.k})] \\ &= \frac{1}{M^2}[(p(1-p) + \lambda p^2)\sum_{h=1}^{L} a_h^2(p_{.h}(1-p_{.h}) + \lambda p_{.h}^2) + \lambda p^2 \sum_{h=1}^{L} a_h^2 p_{.h}^2 \\ &\quad +2(p(1-p) + \lambda p^2)\sum_{h=1}^{L-1}\sum_{k=h+1}^{L} a_h a_k(\lambda p_{.h}p_{.k} - p_{.h}p_{.k}) \\ &\quad +2\lambda p^2 \sum_{h=1}^{L-1}\sum_{k=h+1}^{L} a_h a_k p_{.h}p_{.k}]. \end{aligned} \tag{3.27}$$

Now rewriting (3.16) we have

$$\begin{aligned} E(u^2) &= \frac{1}{F^2}(\sum_{h=1}^{L} N_h\lambda_h)(\sum_{h=1}^{L} N_h\lambda_h + 1) \\ &= \frac{1}{F^2}[(\lambda\sum_{h=1}^{L} a_h p_{.h})(\lambda\sum_{h=1}^{L} a_h p_{.h} + 1)] \\ &= \frac{p^2}{M^2}[\lambda^2(\sum_{h=1}^{L} a_h p_{.h})^2 + \lambda\sum_{h=1}^{L} a_h p_{.h}] \\ &= \frac{1}{M^2}[\lambda^2 p^2(\sum_{h=1}^{L} a_h p_{.h})^2 + \lambda p^2\sum_{h=1}^{L} a_h p_{.h}]. \end{aligned} \tag{3.28}$$

Hence from (3.23), (3.27) and (3.28), we have

$$
\begin{aligned}
E(\hat{u}_2 - u)^2 &= \frac{1}{M^2}([p + (\lambda - 1)p^2][\sum_{h=1}^{L} a_h^2 p_{.h} + (\lambda - 1)(\sum_{h=1}^{L} a_h p_{.h})^2] \\
&\quad + \lambda p^2 (\sum_{h=1}^{L} a_h p_{.h})^2 - \lambda^2 p^2 (\sum_{h=1}^{L} a_h p_{.h})^2 - \lambda p^2 \sum_{h=1}^{L} a_h p_{.h}) \\
&= \frac{1}{M^2}(p \sum_{h=1}^{L} a_h^2 p_{.h} + p(\lambda - 1)(\sum_{h=1}^{L} a_h p_{.h})^2 + p^2(\lambda - 1) \sum_{h=1}^{L} a_h^2 p_{.h} \\
&\quad + p^2(\lambda - 1)^2 (\sum_{h=1}^{L} a_h p_{.h})^2 - p^2 \lambda (\lambda - 1)(\sum_{h=1}^{L} a_h p_{.h})^2 - \lambda p^2 \sum_{h=1}^{L} a_h p_{.h}) \\
&= \frac{1}{M^2}((\sum_{h=1}^{L} a_h p_{.h})^2 [(\lambda - 1)p(1 + p(\lambda - 1) - \lambda p)] + p \sum_{h=1}^{L} a_h^2 p_{.h} \\
&\quad + p^2 \lambda \sum_{h=1}^{L} a_h^2 p_{.h} - p^2 \sum_{h=1}^{L} a_h^2 p_{.h} - \lambda p^2 \sum_{h=1}^{L} a_h p_{.h}) \\
&= \frac{1}{M^2} p(1 - p)(\lambda - 1)(\sum_{h=1}^{L} a_h p_{.h})^2 + p(1 - p) \sum_{h=1}^{L} a_h^2 p_{.h} \\
&\quad + \lambda p^2 \sum_{h=1}^{L} a_h (a_h - 1) p_{.h}.
\end{aligned}
$$

Hence we obtain (3.22) and this completes the proof of the theorem.

It is of interest to find the unbiased estimator of $E(\hat{u}_2 - u)^2$ and we do this in Theorem 3.2.

Theorem 3.2 *The unbiased estimator of* $E(\hat{u}_2 - u)^2$ *is given by*

$$
h_2(r_{.h}, c_{.h}) = \frac{1}{M^2}[(\frac{r_{..}}{c_{..}} \sum_{h=1}^{L} \frac{N_h}{n_h} c_{.h})^2 - (\sum_{h=1}^{L} \frac{N_h}{n_h} r_{.h})^2 + \sum_{h=1}^{L} \frac{N_h}{n_h} r_{.h} (\frac{N_h}{n_h} - \frac{r_{.h} - 1}{c_{.h} - 1})]. \tag{3.29}
$$

Proof: Note that, $\hat{u}_2^2$ is an unbiased estimator of $E(\hat{u}_2^2)$, so to find an unbiased estimator of $E(\hat{u}_2 - u)^2$ it is enough to find an unbiased estimator of $E(u^2)$. Note the following:

$$
E(\frac{r_{.h}^2 - r_{.h}}{n_h^2}) = \lambda_h^2 p^2, \tag{3.30}
$$

$$
E(\frac{r_{.h}}{n_h}) = p\lambda_h \tag{3.31}
$$

and

$$
E(\frac{r_{.h}(r_{.h} - 1)}{n_h(c_{.h} - 1)}) = p^2 \lambda_h. \tag{3.32}
$$

Hence the unbiased estimator of $E(u^2)$ using (3.30), (3.31) and (3.32) is

$$\begin{aligned}\hat{E}(u^2) &= \frac{1}{M^2}[\sum_{h=1}^{L}\frac{N_h^2}{n_h^2}r_{.h}(r_{.h}-1)+2\sum_{h=1}^{L-1}\sum_{k=h+1}^{L}\frac{N_h}{n_h}\frac{N_k}{n_k}r_{.h}r_{.k}+\sum_{h=1}^{L}\frac{N_h}{n_h}\frac{r_{.h}(r_{.h}-1)}{(c_{.h}-1)}] \\ &= \frac{1}{M^2}[(\sum_{h=1}^{L}\frac{N_h}{n_h}r_{.h})^2-\sum_{h=1}^{L}\frac{N_h}{n_h}r_{.h}(\frac{N_h}{n_h}-\frac{r_{.h}-1}{c_{.h}-1})]. \qquad (3.33)\end{aligned}$$

Hence from (3.27) and (3.33) we have

$$h_2(r_{.h},\, c_{.h}) = \frac{1}{M^2}[(\frac{r_{..}}{c_{..}}\sum_{h=1}^{L}\frac{N_h}{n_h}c_{.h})^2-(\sum_{h=1}^{L}\frac{N_h}{n_h}r_{.h})^2+\sum_{h=1}^{L}\frac{N_h}{n_h}r_{.h}(\frac{N_h}{n_h}-\frac{r_{.h}-1}{c_{.h}-1})].$$

This completes the proof of the theorem.

Confidence Interval of u

We can construct a confidence interval of u by considering $\hat{u}_2$ to be the estimator of u. As mentioned earlier and the simulation results of Section 3.4 show that $\hat{u}_2$ has the least MSE as compared to the MLE and the other moment estimator $\hat{u}_1$ of u. It follows that

$$Y = \frac{\hat{u}_2 - u}{\sqrt{E(\hat{u}_2-u)^2}} \rightarrow N(0,1) \text{ as } n \rightarrow \infty. \qquad (3.34)$$

Hence a large sample 100(1-α)% confidence interval for u is given by

$$\hat{u}_2 \pm z_{\frac{\alpha}{2}}\sqrt{h_2(r_{.h},\, c_{.h})}$$

where $h_2(r_{.h},\, c_{.h})$ is given by (3.29).

3.4 Simulation Study

In this section, we compare the MLE $\tilde{u}$ given by (3.9), the moment estimators $\hat{u}_1$ given by (3.12) and $\hat{u}_2$ given by (3.20) of u. We also check the assumption of normality for the distribution of the statistic Y given in (3.34).

For the simulation study we assume the population size to be $F = 15000$. We consider three strata with sizes $N_1 = 125$, $N_2 = 105$ and $N_3 = 120$. We vary the values of M, the strata sample sizes n_1, n_1, n_3 and the average catch rates of the three strata, λ_1, λ_2 and λ_3. We generate the total observed catch ($c_{.h}$) and the total

observed recaptures ($r_{.h}$) in each stratum in the same manner as we did in Section 2.3. The MSE of each estimator is computed, by generating K=1000 samples, in the same way as we did in Section 2.3. The MSE of $\hat{u}_1$ is represented by MSE($\hat{u}_1$), MSE of $\hat{u}_2$ is represented by MSE($\hat{u}_2$) and MSE of MLE is represented as MSE(MLE). We define the following ratios of MSEs:

$$RU1 = MSE(\hat{u}_1)/MSE(\hat{u}_2), \tag{3.35}$$

$$RU2 = MSE(MLE)/MSE(\hat{u}_1) \tag{3.36}$$

and

$$RU3 = MSE(MLE)/MSE(\hat{u}_2). \tag{3.37}$$

Table 3.1 gives the ratios $RU1$, $RU2$ and $RU3$. We can see that the moment estimator $\hat{u}_2$ performs better than $\hat{u}_1$ and the MLE of u for all values of p. The MSE of $\hat{u}_1$ and the MSE of MLE are close for most cases.

Table 3.2 gives the values of Shapiro-Wilk test statistic (W) and the corresponding p-values for testing the assumption of normality for the asymptotic distribution of Y. We can see that for small values of p, it is imperative to sample large number of units from each stratum, for the asymptotic distribution of Y to be close to $N(0,1)$. We also checked the assumption of normality for the asymptotic distribution of Y by replacing $E(\hat{u}_2 - u)^2$ by its unbiased estimator $h_2(r_{.h}, c_{.h})$. We found the convergence of the distribution of Y in this case to be very slow. Since in all practical situations only an estimate of $E(\hat{u}_2 - u)^2$ will be available, we need to sample large number of units from each stratum for the asymptotic distribution of Y to be close to $N(0,1)$.

Table 3.1: Ratios of the MSEs of the Moment Estimators and the MLE of u.

p	n_1	n_2	n_3	λ_1	λ_2	λ_3	$RU1$	$RU2$	$RU3$
0.033	5	10	8	1.5	1	1.2	1.1553952	0.9852173	1.1383153
0.033	5	10	8	2	2.1	1.5	1.1584459	0.9900778	1.1469516
0.033	5	10	8	4	2	3.5	1.084626	0.9963156	1.0806299
0.033	41	50	35	1.5	1	1.2	0.9971051	0.9522328	0.9494762
0.033	41	50	35	4	2	3.5	1.0320238	0.9795325	1.0109008
0.067	5	10	8	4	2	3.5	1.1359066	0.9965985	1.1320427
0.067	20	15	21	2	2.1	1.5	1.0030873	1.0044495	1.0075505
0.067	20	15	21	8	6	4.5	1.0062164	1.0091773	1.0154507
0.067	41	50	35	2	2.1	1.5	1.0594631	0.9952559	1.0544369
0.067	41	50	35	8	6	4.5	1.0373242	1.0289903	1.0673965
0.167	5	10	8	2	2.1	1.5	1.1170693	1.0009484	1.1181288
0.167	20	15	21	2	2.1	1.5	1.0024904	1.0106171	1.013134
0.167	20	15	21	8	6	4.5	0.9999995	1.002695	1.0026946
0.167	41	50	35	1	1.5	1.2	1.0341983	1.0100545	1.0445967
0.167	41	50	35	4	2	3.5	1.0451043	0.9977598	1.0427631
0.233	5	10	8	2	2.1	1.5	1.1249345	1.0061599	1.1318639
0.233	20	15	21	4	2	3.5	1.0032173	1.0056003	1.0088356
0.233	41	50	35	1.5	1	1.2	1.0275223	1.0281708	1.0564684
0.233	41	50	35	4	2	3.5	1.0343155	1.0137309	1.0485175

Table 3.2: Shapiro-Wilk Test of Normality for Y.

p	n_1	n_2	n_3	λ_1	λ_2	λ_3	W	$p-value$
0.033	5	10	8	2	2.1	1.5	0.903873	0.0001
0.033	41	50	35	1.5	1	1.2	0.964728	0.0001
0.033	41	50	35	4	2	3.5	0.982709	0.0356
0.067	5	10	8	4	2	3.5	0.97345	0.0001
0.067	20	15	21	4	2	3.5	0.982935	0.0446
0.067	41	50	35	2	2.1	1.5	0.981975	0.0159
0.067	41	50	35	4	2	3.5	0.985653	0.3268
0.167	5	10	8	2	2.1	1.5	0.980378	0.0019
0.167	5	10	8	8	6	4.5	0.984594	0.1755
0.167	20	15	21	1	1.5	1.2	0.985056	0.2354
0.233	5	10	8	4	2	3.5	0.980654	0.0028
0.233	20	15	21	2	2.1	1.5	0.984935	0.2187
0.233	41	50	35	1.5	1	1.2	0.989209	0.8742

3.5 Data Analysis

In this section, we analyze the data on Medicine Lake, Minnesota; which was also considered in Section 2.4. Here we stratify the sampling units (N) into four strata. Table 3.3 gives the stratified data and also gives the information about the available sampling units in each stratum (N_h), the number of units that was sampled in each stratum (n_h), the total observed catch in each stratum ($c_{.h}$) and the total number of observed recaptures ($r_{.h}$) in each stratum.

It would be of interest to know the exploitation rates of the northern pike in the summer, fall and winter seasons. In Table 3.4, we present the moment estimates ($\hat{u}_h$) of the exploitation rates (u_h) for the four strata. The estimates ($\hat{u}_h$) are obtained using the moment estimator (2.18) and corresponding interval estimates are obtained using (2.44). We can see from Table 3.4 that the exploitation rate of the northern pike is the highest during the summer months, as expected. In Table 3.5, we present the moment estimate ($\hat{u}_2$) and interval estimate of the annual exploitation rate.

Table 3.3: Stratified Data from Medicine Lake, Minnesota.

Strata	N_h	n_h	$c_{.h}$	$r_{.h}$
May 14-July 8	168	38	130	15
July 9-Sept 5	177	41	88	14
Sept 6-Nov 30	165	37	21	3
Dec 1-Feb 15	231	41	4	0
Total	$N = 741$	$n = 157$	$c_{..} = 243$	$r_{..} = 32$

Table 3.4: Estimates of the Strata Exploitation Rates.

Strata	$\hat{u_h}$	$\hat{E}(\hat{u}_h - u_h)^2$	95% confidence interval
May 14-July 8	0.07765	0.0003921	(0.0388, 0.11647)
July 9-Sept 5	0.07077	0.0003454	(0.0344, 0.10719)
Sept 6-Nov 30	0.01567	0.00008	(-0.00186, 0.03319)
Dec 1-Feb 15	0	0	(0, 0)

Table 3.5: Estimates of the Annual Exploitation Rate (u).

$\hat{u}_2$	0.1651
$\hat{E}(\hat{u}_2 - u)^2 = h_2(r_{.h}, c_{.h})$	0.001157
95% confidence interval of u	(0.098453, 0.2318)

Chapter 4

Analysis of Koziol-Green Model with the Assumption of Weibull Lifetimes

4.1 Introduction

In this chapter, we consider the maximum likelihood estimation and Bayesian analysis of the Koziol-Green model with the assumption of Weibull failure times or lifetimes. We also compare the maximum likelihood estimator with the Bayes estimator of the survivor function of lifetimes.

The Koziol-Green model is a special competing risks model with proportional hazards and the model is described in the following manner. Let $T_1, T_2, \ldots, T_n$ be independent and identically distributed random variables denoting lifetimes with a continuous distribution function say F and $C_1, C_2, \ldots, C_n$ be the corresponding independent and identically distributed random variables denoting censoring times with a continuous distribution function say G. Let

$$Z_i = min(T_i, C_i) \text{ and } \delta_i = I(T_i \leq C_i) \tag{4.1}$$

and (Z_i, δ_i) for $i = 1, \ldots, n$ be the observed data.

Let $S_T(t) = P(T > t)$ be the survivor function of the lifetimes and $S_C(t) = P(C > t)$ be the survivor function of the censoring times. The Koziol-Green model assumes

$$S_C(t) = [S_T(t)]^\gamma \tag{4.2}$$

where γ is a parameter that controls the amount of censoring, i.e.,

$$P(T > C) = \frac{\gamma}{1+\gamma}, \qquad \gamma > 0. \tag{4.3}$$

The statement (4.2) is true if and only if the random variables $Z_i's$ and $\delta_i's$ given in (4.1) are independent (see Allen (1963)).

This kind of a proportional hazards model has been used in censored data problems in medical and industrial life testing studies by various authors (Cox D.R. (1959), Efron (1967), Koziol and Green (1976), Chen et al. (1982) and Chen-H'sinchen (1984)) to develop statistical procedures and to study their properties because of the tractable nature of the model. Koziol and Green (1976) used this model to test the hypothesis that an oestrogen treatment of prostatic cancer was ineffective. They developed a Cramer-von Mises type statistic to test the hypothesis. Since then, this model is popularly known as 'Koziol-Green' model.

The Koziol-Green model can also arise as a result of the structure of the system under study. A good motivational example provided by Chen et al. (1982) is the following: suppose one is studying a two component system and the system is a series system which means that the system functions if and only if both components are functioning. Further, let component one itself be a series system of K_1 independent and identically distributed subcomponents with corresponding lifetimes $T_1, T_2, \ldots, T_{K_1}$ each having distribution F (say) and component two is also a series system of K_2 independent and identically distributed subcomponents with corresponding lifetimes $C_1, C_2, \ldots, C_{K_2}$ each having distribution F. Let

$T = min(T_1, T_2, \ldots, T_{K_1})$ and $C = min(C_1, C_2, \ldots, C_{K_2})$ then the pair (T, C) follows a proportional hazards model described by (4.1) and (4.2). Further, if we have n such systems and we observe $Z's$ and $\delta's$ where $Z's$ are the failure times of the system and the $\delta's$ tell us whether the failure of component one or component two caused failure of the system. Then on the basis of the $Z's$ and $\delta's$ we can estimate F, the failure distribution of component one.

The Koziol-Green model is used in medical studies where one is interested in studying the failure due to a particular cause, when there is a possibility that failure could occur due to other causes also, i.e., we have a competing risks situation.

In this chapter, we consider a parametric Koziol-Green model with the underlying lifetime distribution being Weibull. In statistical analysis, parametric models are often used to model the time until an event occurs. Such events may, for instance, be the onset of certain chronic disease in a medical study, the failure of a component in industrial life testing, the performance of a certain task in a learning experiment in psychology or a change of residence in a demographic study.

The Weibull distribution is one of the most commonly used parametric distribution to model failure or lifetime data with monotone (increasing or decreasing) hazard rates. Its probability density function (pdf) is given by

$$f_T(t \mid \theta, \beta) = \frac{\beta}{\theta^\beta} t^{\beta-1} exp(-(\frac{t}{\theta})^\beta) \tag{4.4}$$

where $\theta > 0$ is the scale, $\beta > 0$ is the shape parameter and $t > 0$. The corresponding survivor and hazard functions are

$$S_T(t) = exp(-(\frac{t}{\theta})^\beta) \tag{4.5}$$

and

$$h_T(t) = h_T(t \mid \theta, \beta) = \frac{\beta\, t^{\beta-1}}{\theta^\beta}. \tag{4.6}$$

respectively. The hazard rate is increasing for $\beta > 1$, decreasing for $0 < \beta < 1$ and constant for $\beta = 1$, the latter corresponding to an exponential pdf with mean θ.

Hence under the assumptions (4.1), (4.2) and (4.5), for randomly censored data, the survivor function and probability density function of censoring times $C_i's$ are

$$S_C(t) = [exp(-(\frac{t}{\theta})^\beta)]^\gamma = exp(-\gamma\,(\frac{t}{\theta})^\beta) \tag{4.7}$$

and

$$f_C(t) = \frac{\gamma\,\beta}{\theta^\beta}\,t^{\beta-1}exp(-\gamma\,(\frac{t}{\theta})^\beta), \tag{4.8}$$

respectively.

In Section 4.2, we discuss the maximum likelihood estimation of the parameters θ, β, γ and the survivor function of lifetimes $S_T(t)$. Section 4.2.1 gives the asymptotic variance-covariance matrix of the estimators. In Section 4.3, we consider the Bayesian analysis of the model. Section 4.3.1 gives the prior distributions and the posterior distributions. We also discuss briefly the motivation behind the choice of the prior distributions. In Section 4.3.2, we discuss the Gibbs sampler technique. We use it to obtain the Bayes estimates of the posterior pdfs of the scale and shape parameters of the model and of the survivor function for a specified time. Sections 4.3.3, 4.3.4, 4.3.5 and 4.3.6 give conditional posterior distributions, algorithms for sampling from them and estimation of posterior moments and other quantities of interest using the Gibbs sampler. In Section 4.3.7, we give a numerical example. In Section 4.3.8, we compare the MLE with the Bayes estimator of the survivor function of the lifetimes by a simulation study. In Section 4.3.9, we consider a real life example.

4.2 Maximum Likelihood Estimation

The likelihood function for the Koziol-Green model described by (4.1), (4.2) and with the assumption of Weibull lifetimes is

$$\begin{aligned} L(\theta,\ \beta,\ \gamma \mid \underline{z},\ \underline{\delta}) &= \prod_{i=1}^{n}[S_C(z_i)\ f_T(z_i)]^{\delta_i}\ [S_T(z_i)\ f_C(z_i)]^{1-\delta_i} \\ &= \frac{\beta^n}{\theta^{n\,\beta}} \prod_{i=1}^{n} z_i^{\beta-1} \gamma^{n-\sum_{i=1}^{n}\delta_i} exp(-\frac{(1+\gamma)}{\theta^\beta}\sum_{i=1}^{n} z_i^\beta). \end{aligned} \tag{4.9}$$

The likelihood function in (4.9) is obtained using (4.4), (4.5), (4.7) and (4.8) where $\underline{z} = (z_1, z_2, \ldots, z_n)$ and $\underline{\delta} = (\delta_1, \delta_2, \ldots, \delta_n)$.
The natural logarithm of the likelihood in (4.9) is

$$\ln L = n\log\beta - n\beta\log\theta + (\beta-1)\sum_{i=1}^{n}\ln z_i + (n - \sum_{i=1}^{n}\delta_i)\log\gamma - \frac{(1+\gamma)}{\theta^\beta}\sum_{i=1}^{n} z_i^\beta \tag{4.10}$$

where $L = L(\theta,\ \beta,\ \gamma \mid \underline{z},\ \underline{\delta})$.

To get the MLEs of θ, β and γ, we partially differentiate the log-likelihood given by (4.10) with respect to θ, β and γ and equate the partial derivatives to zeros. Hence the estimating equations are

$$\frac{\partial \ln L}{\partial \theta} = 0$$

$$\Rightarrow (1+\gamma)\beta\frac{\sum_{i=1}^{n} z_i^\beta}{\theta^\beta} - n\beta = 0, \tag{4.11}$$

$$\frac{\partial \ln L}{\partial \beta} = 0$$

$$\Rightarrow \frac{n}{\beta} - n\ln\theta + \sum_{i=1}^{n}\ln z_i - \frac{(1+\gamma)}{\theta^\beta}[\sum_{i=1}^{n}(\ln z_i - \ln\theta) z_i^\beta] = 0, \tag{4.12}$$

and

$$\frac{\partial \ln L}{\partial \gamma} = 0$$

$$\Rightarrow \frac{(n - \sum_{i=1}^{n} \delta_i)}{\gamma} - \frac{\sum_{i=1}^{n} z_i^{\beta}}{\theta^{\beta}} = 0$$

$$\Rightarrow \gamma = \frac{(n - \sum_{i=1}^{n} \delta_i)\, \theta^{\beta}}{\sum_{i=1}^{n} z_i^{\beta}}. \tag{4.13}$$

Substituting for γ given by (4.13) in equations (4.11) and (4.12), we have

$$(\sum_{i=1}^{n} z_i^{\beta} + (n - \sum_{i=1}^{n} \delta_i)\, \theta^{\beta}) \frac{\beta}{\theta^{\beta}} - n\, \beta = 0 \tag{4.14}$$

and

$$\frac{n}{\beta} - n \ln \theta + \sum_{i=1}^{n} \ln z_i - (\frac{\sum_{i=1}^{n} z_i^{\beta} + (n - \sum_{i=1}^{n} \delta_i)\, \theta^{\beta}}{\sum_{i=1}^{n} z_i^{\beta} \theta^{\beta}})(\sum_{i=1}^{n} \ln(\frac{z_i}{\theta}) z_i^{\beta}) = 0. \tag{4.15}$$

We solve equations (4.14) and (4.15) simultaneously to obtain the MLEs of θ and β. No explicit expression exists for the MLEs of θ and β, equations (4.14) and (4.15) have to be solved by a numerical procedure. We employ the Newton-Raphson method to solve the equations for θ and β. The MLE of γ can be obtained by substituting the MLEs of θ and β in (4.13). Let $\tilde{\theta}$, $\tilde{\beta}$ and $\tilde{\gamma}$ represent the MLEs of θ, β and γ, respectively. The MLE of the survivor function in (4.5) is given by

$$\tilde{S}(t) = exp(-(\frac{t}{\tilde{\theta}})^{\tilde{\beta}}). \tag{4.16}$$

4.2.1 Variances and Covariances of Estimators

The asymptotic variance-covariance matrix of the MLEs of the parameters θ, β and γ is obtained by inverting the information matrix whose elements are the negative expected values of the second order derivatives of the log-likelihood given in (4.10). For a sufficiently large sample we can estimate the expected values by their MLEs. Accordingly, we have the estimator of asymptotic variance-covariance matrix as

$$\begin{bmatrix} I_{11} & I_{12} & I_{13} \\ I_{21} & I_{22} & I_{23} \\ I_{31} & I_{32} & I_{33} \end{bmatrix}^{-1} = \begin{bmatrix} \hat{V}(\tilde{\theta}) & \hat{C}ov(\tilde{\theta},\tilde{\beta}) & \hat{C}ov(\tilde{\theta},\tilde{\gamma}) \\ \hat{C}ov(\tilde{\theta},\tilde{\beta}) & \hat{V}(\tilde{\beta}) & \hat{C}ov(\tilde{\beta},\tilde{\gamma}) \\ \hat{C}ov(\tilde{\theta},\tilde{\gamma}) & \hat{C}ov(\tilde{\beta},\tilde{\gamma}) & \hat{V}(\tilde{\gamma}) \end{bmatrix}. \tag{4.17}$$

Where

$$I_{11} = -\frac{\partial^2 \ln L}{\partial \theta^2}\bigg|_{\tilde{\theta},\tilde{\beta},\tilde{\gamma}} = -\frac{n\beta}{\theta^2} + \frac{(1+\gamma)\beta(\beta+1)}{\theta^{\beta+2}} \sum_{i=1}^{n} z_i^{\beta}\bigg|_{\tilde{\theta},\tilde{\beta},\tilde{\gamma}} \tag{4.18}$$

$$\begin{aligned} I_{12} &= I_{21} = -\frac{\partial^2 \ln L}{\partial\theta\partial\beta}\bigg|_{\tilde{\theta},\tilde{\beta},\tilde{\gamma}} \\ &= \frac{n}{\theta} - \frac{(1+\gamma)}{\theta^{\beta+1}}[\beta \sum_{i=1}^{n} z_i^{\beta} \ln z_i + (1 - \beta \ln \theta) \sum_{i=1}^{n} z_i^{\beta}]\bigg|_{\tilde{\theta},\tilde{\beta},\tilde{\gamma}}, \end{aligned} \tag{4.19}$$

$$I_{13} = I_{31} = -\frac{\partial^2 \ln L}{\partial\theta\partial\gamma}\bigg|_{\tilde{\theta},\tilde{\beta},\tilde{\gamma}} = -\frac{\beta}{\theta^{\beta+1}} \sum_{i=1}^{n} z_i^{\beta}\bigg|_{\tilde{\theta},\tilde{\beta},\tilde{\gamma}}, \tag{4.20}$$

$$I_{22} = -\frac{\partial^2 \ln L}{\partial\beta^2}\bigg|_{\tilde{\theta},\tilde{\beta},\tilde{\gamma}} = \frac{n}{\beta^2} + \frac{(1+\gamma)}{\theta^{\beta}}[\sum_{i=1}^{n} z_i^{\beta}(\ln z_i - \ln\theta)^2]\bigg|_{\tilde{\theta},\tilde{\beta},\tilde{\gamma}}, \tag{4.21}$$

$$I_{23} = I_{32} = -\frac{\partial^2 \ln L}{\partial\beta\partial\gamma}\bigg|_{\tilde{\theta},\tilde{\beta},\tilde{\gamma}} = -\frac{1}{\theta^{\beta}} \sum_{i=1}^{n} (\ln\theta - \ln z_i) z_i^{\beta}\bigg|_{\tilde{\theta},\tilde{\beta},\tilde{\gamma}} \tag{4.22}$$

and

$$I_{33} = -\frac{\partial^2 \ln L}{\partial\gamma^2}\bigg|_{\tilde{\theta},\tilde{\beta},\tilde{\gamma}} = \frac{1}{\gamma^2}(n - \sum_{i=1}^{n} \delta_i)\bigg|_{\tilde{\theta},\tilde{\beta},\tilde{\gamma}}. \tag{4.23}$$

Equation (4.17) is strictly valid only for large samples as noted by Cohen (1965) and Lemon (1975), it also works well in case of moderately large samples.

4.3 Bayesian Analysis

For the Bayesian analysis of the Koziol-Green model with Weibull lifetimes, we consider two sets of prior distributions and we obtain the Bayes estimates of the posterior pdfs of the parameters of the model and the survivor function by the method of Gibbs sampler.

4.3.1 Prior and Posterior Distributions

(i) The first set of prior distributions that we consider for θ, β and γ are given below.

We assume the distribution of θ^{β} given β to be inverted gamma (IG(a, b)) with shape parameter $a > 0$ and scale parameter $b > 0$ such that the prior distribution of θ given β is given by

$$\pi_1(\theta|\beta) = \beta \frac{b^a}{\Gamma a} exp(-\frac{b}{\theta^{\beta}}) \frac{1}{\theta^{\beta\, a+1}}, \quad \theta > 0. \tag{4.24}$$

One advantage of choosing this prior distribution is that the conditional posterior distribution of θ given β, γ and the data will also be inverted gamma and this will help us to generate samples from the conditional distribution and hence facilitate the implementation of the Gibbs sampler and simplify the Bayes estimation of θ.

The prior pdf $\pi_1(\beta)$ of β is assumed to be a log-concave density. There are several log-concave densities available in the literature, see Devroye (1986, page 287) for more information. Here we assume the prior distribution to be a truncated normal distribution with pdf

$$\pi_1(\beta) \propto exp(-\frac{\beta^2}{2\,\sigma^2}), \qquad 0 < c_0 \leq \beta \leq d_0 < \infty. \tag{4.25}$$

For the same reasons as mentioned above, we assume the prior distribution of β to be log-concave so that the conditional posterior distribution of β given θ, γ and data will also be log-concave and this will help us in the implementation of the Gibbs sampler. We can easily show that $\pi_1(\beta)$ given in (4.25) is log-concave.

Consider

$$\ln \pi_1(\beta) = C^* - \frac{\beta^2}{2\,\sigma^2}. \tag{4.26}$$

By differentiating (4.26) twice with respect to β, we have

$$\frac{\partial^2 \ln \pi_1(\beta)}{\partial \beta^2} = -\frac{2}{2\,\sigma^2} < 0 \tag{4.27}$$

which implies that $\pi_1(\beta)$ is a log-concave density.

Finally, the prior density function of the censoring parameter γ is assumed to be truncated exponential and it is given by

$$\pi_1(\gamma) = \frac{\eta\, exp(-\eta\,\gamma)}{1 - exp(-\eta\, k)},\ 0 < \gamma \leq k < \infty. \tag{4.28}$$

Choices of a, b, c_0, d_0, σ, η and k will be based on the knowledge of previous similar studies. If we are conducting reliability studies then they could also be fixed based on engineering knowledge.

The posterior density function of (θ, β, γ) using (4.9), (4.24), (4.25) and (4.28) is given by

$$f(\theta,\ \beta,\ \gamma |\ \underline{z},\ \underline{\delta}) \propto \pi_1(\beta)\pi_1(\gamma)\pi_1(\theta|\beta) L(\theta,\ \beta,\ \gamma \mid \underline{z},\ \underline{\delta})$$

$$\propto \pi_1(\beta)\pi_1(\gamma)\beta \frac{b^a}{\Gamma a}\, exp(-\frac{b}{\theta^\beta}) \frac{1}{\theta^{\beta\, a+1}} \frac{\beta^n}{\theta^{n\,\beta}} \prod_{i=1}^{n} z_i^{\beta-1}\ \gamma^{n-\sum_{i=1}^n \delta_i}\ exp(-\frac{(1+\gamma)}{\theta^\beta} \sum_{i=1}^{n} z_i^\beta)$$

$$\propto \pi_1(\beta)\pi_1(\gamma)\ \frac{b^a}{\Gamma a}\ \frac{1}{\theta^{\beta(n+a)+1}} \beta^{n+1}\ \gamma^{n-\sum_{i=1}^n \delta_i}\ \prod_{i=1}^{n} z_i^{\beta-1}\ exp(-\frac{(1+\gamma)}{\theta^\beta} \sum_{i=1}^{n} z_i^\beta - \frac{b}{\theta^\beta}). \tag{4.29}$$

(ii) The other set of prior distributions that we consider for θ, β and γ are as follows. The prior pdf of θ given β is same as given in (4.24). Whereas the prior pdf of β is assumed to be truncated gamma with shape parameter g_a and scale parameter g_b with pdf

$$\pi_1(\beta) \propto \beta^{g_a-1}\, exp(-\beta\, g_b), \qquad 0 < c_0 \leq \beta \leq d_0 < \infty. \tag{4.30}$$

It can easily be shown that $\pi_1(\beta)$ given in (4.30) is log-concave.

Finally, the prior pdf of γ is assumed to be uniform with pdf

$$\pi_1(\gamma) = \frac{1}{l}, \qquad 0 < \gamma \leq l. \tag{4.31}$$

The posterior density function of (θ, β, γ) is of the same form as (4.29), except that $\pi_1(\beta)$ and $\pi_1(\gamma)$ are the pdfs given by (4.30) and (4.31).

4.3.2 The Gibbs Sampling Scheme

It is difficult to obtain the marginal posterior pdfs of θ, β and γ from the joint posterior pdf of θ, β and γ given in (4.29), by integration, in a closed form. Hence, we adopt the method of Gibbs sampler which enables us to obtain the marginal posterior moments, marginal posterior pdfs and the Bayes estimate of the survivor function.

The Gibbs sampler is a technique for generating random variables from a distribution indirectly without the necessity to calculate the density. The Gibbs sampler helps us to avoid difficult calculations. This sampling scheme became popular with the paper of Geman and Geman (1984). Gelfand and Smith (1990) proved several properties of the Gibbs sampler and showed its applications to a wide variety of conventional statistical problems. Casella and George (1992) give a simple account of the theory behind the Gibbs sampler. Most of the applications of the Gibbs sampler have been made in Bayesian models, but it can also be used in classical problems.

Suppose we are given a joint probability density function $f(z, y_1, y_2, \ldots, y_p)$ and we are interested in obtaining the characteristics of the marginal density

$$f(z) = \int\int \ldots \int f(z, y_1, y_2, \ldots, y_p)\, dy_1 \ldots dy_p,$$

such as the mean and variance. To do this we will have to first calculate $f(z)$ which may be difficult to obtain both analytically and numerically. In such cases the Gibbs sampler provides an alternative method for obtaining $f(z)$. The Gibbs sampler allows us to generate a sample $z_1, \ldots, z_m$ from $f(z)$ without really requiring $f(z)$. We can simulate a large enough sample to calculate the mean, variance or any other characteristic of $f(z)$.

To explain the Gibbs sampler in the most simple terms, let us consider the case of two variables. Suppose we have a pair of random variables (Y, Z), the Gibbs sampler generates a sample from the marginal densities $f(y)$ and $f(z)$ by sampling from the conditional densities $f(y|z)$ and $f(z|y)$ which are usually known and easy to obtain. We specify an initial value $Z = z_0$ and generate

$$Y_0 \sim f(y \mid Z = z_0),$$

$$Z_1 \sim f(z \mid Y = y_0),$$

$$Y_1 \sim f(y \mid Z = z_1),$$

$$Z_2 \sim f(z \mid Y = y_1)$$

and so on, iteratively. Suppose there are N such iterations giving us the Gibbs sequence of random variables

$$Z_0,\ Y_0,\ Z_1,\ Y_1, \ldots, Z_N,\ Y_N. \tag{4.32}$$

We can discard the first N_1 values of Y and Z and this is known as the burn in sample. For N large enough $Y_1, \ldots, Y_n$ where $n = N - N_1$ is a sample from $f(y)$ and $Z_1, \ldots, Z_n$ is a sample from $f(z)$. Hence the estimate of the mean of $f(y)$ is $\frac{1}{n}\sum_{i=1}^{n} Y_i$ and we can estimate the marginal density $f(y)$ itself with

$$\hat{f}(y) \;=\; \frac{1}{n}\sum_{i=1}^{n} f(y \mid z_i), \tag{4.33}$$

since

$$E(f(y \mid z)) = \int f(y \mid z)\, f(z)\, dz = f(y).$$

The end result of any calculations, although based on simulations, are the population quantities. For example, to estimate the mean of $f(y)$ we use $\frac{1}{n}\sum_{i=1}^{n} Y_i$ and

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} Y_i &= \int_{-\infty}^{\infty} y\, f(y)\, dy \\ &= E(Y). \end{aligned} \tag{4.34}$$

For a large enough N, any population characteristic, even the density itself, can be obtained to any degree of accuracy. Casella and George (1992) with a simple example of a 2×2 table with multinomial sampling show that the Gibbs sampler generates a Markov chain of random variables which converge to a random variable having the marginal distribution of interest.

The Gibbs sampler for estimating the characteristics of a posterior distribution under the framework of a Bayesian model works as follows. Suppose, we want to compute an integral of the form $\int \varphi(\xi) f(\xi \mid x)\, d\xi$ where φ is a function defined on $\Re^d$ and $f(\xi \mid x)$ is a general posterior pdf of the parameter $\xi = (\xi_1, \ldots, \xi_d) \in \Re^d$ given the data x. Suppose the full conditional pdfs $f(\xi_i \mid x, \xi_{-i})$ are available for sampling where $\xi_{-i} = \{\xi_j : j \neq i\}$. Then for the implementation of the Gibbs sampler we start with an initial value $\xi^{(0)} = (\xi_1^{(0)}, \ldots, \xi_d^{(0)})$ and generate

$$\xi_1^{(1)} \sim f(\xi_1 \mid x, \xi_2^{(0)}, \ldots, \xi_d^{(0)}),$$

$$\xi_2^{(1)} \sim f(\xi_2 \mid x, \xi_1^{(1)}, \xi_3^{(0)}, \ldots, \xi_d^{(0)}),$$

and so on, up to

$$\xi_d^{(1)} \sim f(\xi_d \mid x, \xi_1^{(1)}, \xi_2^{(1)}, \ldots, \xi_{d-1}^{(1)}),$$

giving $\xi^{(1)} = (\xi_1^{(1)}, \ldots, \xi_d^{(1)})$. Repeat the process with initial value $\xi^{(1)}$, continue iterating, ending with $\xi^{(N)} = (\xi_1^{(N)}, \ldots, \xi_d^{(N)})$ after N such iterations.

Under milder conditions, it has been shown that

$$\frac{1}{N-N_1}\sum_{j=N_1+1}^{N}\varphi(\xi^{(j)}) \to \int \varphi(\xi)\, f(\xi \mid x)\, d\xi \text{ as } N \to \infty$$

where N_1 is the discarded burn in sample size. Furthermore, the marginal posterior pdf of ξ_i can be estimated by

$$\hat{f}(\xi_i \mid x) = \frac{1}{N-N_1}\sum_{j=N_1+1}^{N} f(\xi_i \mid x,\ \xi_{-i}^{(j)})$$

where $\xi_{-i}^{(j)} = \{\xi_l^{(j)} : l \neq i\}$. See Gelfand and Smith (1990), Casella and George (1992), Ritter and Tanner (1992) and Roberts (1992) for further references.

4.3.3 Available Conditional Posterior Distributions

To implement the Gibbs sampler for our model, we need three conditional posterior distributions.

Conditional posterior distribution of θ given β, γ and the data

The conditional posterior pdf of θ given β, γ and the data from (4.29) is given by

$$f(\theta \mid \beta,\ \gamma,\ \underline{z},\ \underline{\delta}) \propto \frac{1}{\theta^{\beta(n+a)+1}} exp(-\frac{(1+\gamma)}{\theta^{\beta}}\sum_{i=1}^{n} z_i^{\beta} - \frac{b}{\theta^{\beta}}) \tag{4.35}$$

which is equivalent to $(\theta^{\beta} \mid \beta,\ \gamma,\ \underline{z},\ \underline{\delta}) \sim \text{IG}(n+a,\ b+(1+\gamma)\sum_{i=1}^{n} z_i^{\beta})$. To show this, let $(\theta^{\beta} \mid \beta,\ \gamma,\ \underline{z},\ \underline{\delta}) \sim \text{IG}(n+a,\ b+(1+\gamma)\sum_{i=1}^{n} z_i^{\beta})$ and $Y = \theta^{\beta}$. Then the pdf of $(Y \mid \beta,\ \gamma,\ \underline{z},\ \underline{\delta})$ is

$$f(y \mid \beta, \gamma, \underline{z}, \underline{\delta}) = \frac{(b+(1+\gamma)\sum_{i=1}^{n} z_i^{\beta})^{n+a}}{\Gamma(n+a)} \frac{1}{y^{(n+a)+1}} exp(-\frac{(b+(1+\gamma)\sum_{i=1}^{n} z_i^{\beta})}{y}). \tag{4.36}$$

Since $Y = \theta^{\beta}$, we have $dY = \beta\,\theta^{\beta-1}\,d\theta$.

The pdf of $(\theta \mid \beta,\ \gamma,\ \underline{z},\ \underline{\delta})$ is obtained by substituting θ^{β} for Y and multiplying by the Jacobian of transformation $\left|\frac{dy}{d\theta}\right|$ in (4.36). Hence we have

$$f(\theta \mid \beta, \gamma, \underline{z}, \underline{\delta}) = \frac{\beta(b+(1+\gamma)\sum_{i=1}^{n} z_i^{\beta})^{n+a}}{\Gamma(n+a)\ \theta^{\beta(n+a)+1}}\ exp(-\frac{1}{\theta^{\beta}}(b+(1+\gamma)\sum_{i=1}^{n} z_i^{\beta})). \quad (4.37)$$

Similarly, we can show the converse.

Conditional posterior distribution of β given θ, γ and the data

The conditional posterior pdf of β given θ, γ and the data from (4.29) is given by

$$f(\beta \mid \theta,\ \gamma,\ \underline{z},\ \underline{\delta}) \propto \pi_1(\beta)\ \frac{\beta^{n+1}}{\theta^{\beta(n+a)+1}}\ \prod_{i=1}^{n} z_i^{\beta-1} exp(-\frac{(1+\gamma)}{\theta^{\beta}}\sum_{i=1}^{n} z_i^{\beta} - \frac{b}{\theta^{\beta}}). \quad (4.38)$$

We have considered $\pi_1(\beta)$ to be (i) truncated normal distribution with pdf given in (4.25) and (ii) truncated gamma distribution with pdf given in (4.30) and as mentioned earlier, both are log-concave densities.

Now, we show that $f(\beta \mid \theta,\ \gamma,\ \underline{z},\ \underline{\delta})$ is a log-concave density. Taking natural logarithm of both sides of (4.38), we have

$$\begin{aligned} \ln f(\beta \mid \theta,\ \gamma,\ \underline{z},\ \underline{\delta}) \ &= \ \ln c^{*} + \ln \pi_1(\beta) + (n+1)\ln\beta - [\beta(n+a)+1]\ln\theta \\ &\quad +(\beta-1)\sum_{i=1}^{n}\ln z_i - \frac{(1+\gamma)}{\theta^{\beta}}\sum_{i=1}^{n} z_i^{\beta} - \frac{b}{\theta^{\beta}}. \end{aligned} \quad (4.39)$$

Upon differentiating (4.39) with respect to β, we get

$$\begin{aligned} \frac{\partial \ln f(\beta \mid \theta,\ \gamma,\ \underline{z},\ \underline{\delta})}{\partial\beta} \ &= \ \frac{\partial \ln \pi_1(\beta)}{\partial\beta} + \frac{n+1}{\beta} - (n+a)\ln\theta + \sum_{i=1}^{n}\ln z_i - \\ &\quad \frac{(1+\gamma)}{\theta^{\beta}}\sum_{i=1}^{n} z_i^{\beta}(\ln z_i - \ln\theta) + \frac{b}{\theta^{\beta}}\ln\theta \end{aligned} \quad (4.40)$$

and differentiating (4.40) with respect to β again, we get

$$\frac{\partial^2 \ln f(\beta \mid \theta,\ \gamma,\ \underline{z},\ \underline{\delta})}{\partial\beta^2} \ = \ \frac{\partial^2 \ln \pi_1(\beta)}{\partial\beta^2} - \frac{n+1}{\beta^2} - \frac{(1+\gamma)}{\theta^{\beta}}\sum_{i=1}^{n} z_i^{\beta}(\ln\frac{z_i}{\theta})^2 - \frac{b}{\theta^{\beta}}(\ln\theta)^2. \quad (4.41)$$

Since we have assumed $\pi_1(\beta)$ to be log-concave, from (4.41), we have

$$\frac{\partial^2 \ln f(\beta \mid \theta,\ \gamma,\ \underline{z},\ \underline{\delta})}{\partial \beta^2} < 0.$$

Hence $f(\beta \mid \theta,\ \gamma,\ \underline{z},\ \underline{\delta})$ is also a log-concave density. We employ the approach of Devroye (1986), Berger and Sun (1993) to sample from a log-concave density like (4.38). We discuss the method in Section 4.3.5.

Conditional posterior distribution of γ given θ, β and the data

The conditional posterior pdf of γ given θ, β and the data from (4.29) is given by

$$f(\gamma \mid \theta,\ \beta,\ \underline{z},\ \underline{\delta}) \propto \pi_1(\gamma)\ \gamma^{n-\sum_{i=1}^{n}\delta_i}\ exp(-\frac{(1+\gamma)}{\theta^\beta}\sum_{i=1}^{n} z_i^\beta). \tag{4.42}$$

(i) If $\pi_1(\gamma)$ is truncated exponential given by (4.28) then

$$f(\gamma \mid \theta,\ \beta,\ \underline{z},\ \underline{\delta}) \propto \frac{\eta\ exp(-\eta\ \gamma)}{1-exp(-\eta\ k)}\ \gamma^{n-\sum_{i=1}^{n}\delta_i}\ exp(-\frac{(1+\gamma)}{\theta^\beta}\sum_{i=1}^{n} z_i^\beta)$$

$$\propto \gamma^{n-\sum_{i=1}^{n}\delta_i}\ exp(-\gamma(\frac{\sum_{i=1}^{n} z_i^\beta}{\theta^\beta}+\eta)). \tag{4.43}$$

Hence $f(\gamma \mid \theta,\ \beta,\ \underline{z},\ \underline{\delta})$ is a truncated gamma distribution with shape parameter $(n-\sum_{i=1}^{n}\delta_i+1)$, scale parameter $(\frac{\sum_{i=1}^{n} z_i^\beta}{\theta^\beta}+\eta)$ and $0<\gamma\le k$.

(ii) If $\pi_1(\gamma)$ is uniform distribution given by (4.31) then

$$f(\gamma \mid \theta,\ \beta,\ \underline{z},\ \underline{\delta}) \propto \frac{1}{l}\ \gamma^{n-\sum_{i=1}^{n}\delta_i}\ exp(-\frac{(1+\gamma)}{\theta^\beta}\sum_{i=1}^{n} z_i^\beta)$$

$$\propto \gamma^{n-\sum_{i=1}^{n}\delta_i}\ exp(-\frac{\gamma}{\theta^\beta}\sum_{i=1}^{n} z_i^\beta). \tag{4.44}$$

From (4.44), we can see that $f(\gamma \mid \theta,\ \beta,\ \underline{z},\ \underline{\delta})$ is again truncated gamma with shape parameter $(n-\sum_{i=1}^{n}\delta_i+1)$, scale parameter $(\frac{\sum_{i=1}^{n} z_i^\beta}{\theta^\beta})$ and $0<\gamma\le l$.

4.3.4 Sampling from the Conditional Posterior Pdfs of θ and γ

Sampling from $f(\theta \mid \beta, \gamma, \underline{z}, \underline{\delta})$

To generate θ from $f(\theta \mid \beta, \gamma, \underline{z}, \underline{\delta})$ given in (4.35), we generate a random variable X from Gamma$(n + a,\ b + (1 + \gamma)\sum_{i=1}^{n} z_i^{\beta})$ and let $\theta = (\frac{1}{X})^{\frac{1}{\beta}}$. This amounts to generating θ from $f(\theta \mid \beta, \gamma, \underline{z}, \underline{\delta})$.

Sampling from $f(\gamma \mid \theta, \beta, \underline{z}, \underline{\delta})$

Let X be a continuous random variable with pdf $f(x)$ and cumulative distribution function (cdf) $F(x)$. If we truncate X between a_0 and a_1 then the pdf of the truncated random variable X is

$$g(x) = \frac{f(x)}{\int_{a_0}^{a_1} f(x)dx}, \quad a_o \leq x \leq a_1$$

and the cdf of the truncated random variable is

$$G(x) = \frac{\int_{a_0}^{x} f(x)dx}{F(a_1) - F(a_0)} = \frac{F(x) - F(a_0)}{F(a_1) - F(a_0)}. \tag{4.45}$$

To generate a random variable X from a truncated pdf, we first generate a random variable U from Uniform(0, 1) distribution and let

$$X = F^{-1}[U(F(a_1) - F(a_0)) + F(a_0)]. \tag{4.46}$$

Now, it follows from (4.46) that

$$\begin{aligned} P(X \leq x) &= P(F^{-1}[U(F(a_1) - F(a_0)) + F(a_0)] \leq x) \\ &= P(U(F(a_1) - F(a_0)) + F(a_0) \leq F(x)) \\ &= P(U \leq \frac{F(x) - F(a_0)}{F(a_1) - F(a_0)}) \\ &= \frac{F(x) - F(a_0)}{F(a_1) - F(a_0)} \\ &= G(x). \end{aligned} \tag{4.47}$$

Hence to sample from $f(\gamma \mid \theta, \beta, \underline{z}, \underline{\delta})$, we first generate a random variable U from Uniform(0, 1) distribution and let

$$\gamma = F^{-1}[U(F(k) - F(0)) + F(0)]. \tag{4.48}$$

Where (i) F is the cumulative gamma distribution function with shape parameter $(n - \sum_{i=1}^{n} \delta_i + 1)$ and scale parameter $(\frac{\sum_{i=1}^{n} z_i^{\beta}}{\theta^{\beta}} + \eta)$; and in the other case (ii) F is the cumulative gamma distribution function with shape parameter $(n - \sum_{i=1}^{n} \delta_i + 1)$ and scale parameter $(\frac{\sum_{i=1}^{n} z_i^{\beta}}{\theta^{\beta}})$.

4.3.5 Sampling from a Log-Concave Density

In this section, we present an algorithm for sampling from the conditional posterior density $f(\beta \mid \theta, \gamma, \underline{z}, \underline{\delta})$ given in (4.38). The algorithm we use is an explicit version of the accept-reject algorithm for sampling from a log-concave density given in Devroye (1986). Also see Berger and Sun (1993).

Acceptance-Rejection Algorithm

To simulate X from $f(x)$ may be difficult whenever $f(x)$ is a complicated distribution. So one way would be to find a simple function $g(x)$ which follows $f(x)$ closely, and satisfies the following:

$$f(x) \leq g(x) \ \forall x.$$

The density function $q(x)$ corresponding to $g(x)$ is

$$q(x) = \frac{g(x)}{\int_{-\infty}^{\infty} g(x)dx}.$$

The next step would be to generate X^* from $q(x)$ and also generate another random variable U from Uniform(0, 1). If $U \leq \frac{f(x^*)}{g(x^*)}$ then X^* will have distribution $f(x)$, otherwise we reject X^* and generate another X^* from $q(x)$ and U from Uniform(0, 1) and we continue this process till we accept X^*.

Since $f(x)$ may be a complicated function, calculation of $f(x^*)$ might take a long time and to avoid this to a certain extent we could find another simple function

$h(x)$ such that

$$h(x) \leq f(x) \leq g(x) \; \forall x.$$

Now generate X^* from $q(x)$ and generate U from Uniform$(0, 1)$.
If $U \leq \dfrac{h(x^*)}{g(x^*)}$ then $X^* \sim f(x)$, otherwise compute $f(x^*)$ and
if $U \leq \dfrac{f(x^*)}{g(x^*)}$ then $X^* \sim f(x)$. If this fails, we reject X^* and we again generate X^* from $q(x)$ and U from Uniform$(0, 1)$ and repeat the process till we accept X^*. One advantage of the acceptance-rejection method is that we can apply this method even if we know $f(x)$ only up to a constant.

We apply the above accept-reject algorithm to our problem of generating random variables from the conditional posterior distribution $f(\beta \mid \theta, \gamma, \underline{\xi}, \underline{\delta})$ given by (4.38), which is known up to a constant and is log-concave with support on the interval $[c_0, d_0]$. The algorithm proceeds as follows.
Let $\ln f(\beta \mid \theta, \gamma, \underline{\xi}, \underline{\delta}) = h(\beta)$. Our first aim is to find two functions $h_l(\beta)$ and $h_u(\beta)$ such that

$$h_l(\beta) \leq h(\beta) \leq h_u(\beta)$$

where h_l is a squeezing function and h_u is an envelope function of $h(\beta)$.

Step 1. Choose s_1 and s_3 (refer Figure 4.1) such that s_1 is close to the lower end point and s_3 is close to the upper end point of the support interval $[c_0, d_0]$.
Step 2. To find the envelope function h_u given by $ABCDEFG$ (see Figure 4.1), we have to find the equations of lines AC, CE and EG.
The equation of line AC is

$$y = h(s_1) + h'(s_1)(s - s_1) \tag{4.49}$$

and the equation of line EG is

$$y = h(s_3) + h'(s_3)(s - s_3). \tag{4.50}$$

In order to find the equation of CE we need to know s_2 which is the intersection point of AC and EG. It is obtained by solving equations (4.49) and (4.50) simultaneously. giving

$$s_2 = \frac{h(s_3) - h(s_1) - h'(s_3)s_3 + h'(s_1)s_1}{h'(s_1) - h'(s_3)}$$

and hence the equation of CE is

$$y = h(s_2) + h'(s_2)(s - s_2). \tag{4.51}$$

We also need to know the points u_1 and u_2 which are the intersection points of AC and CE, CE and EG, respectively. Hence we have

$$u_1 = \frac{h(s_2) - h(s_1) + h'(s_1)s_1 - h'(s_2)s_2}{h'(s_1) - h'(s_2)} \tag{4.52}$$

and

$$u_2 = \frac{h(s_3) - h(s_2) - h'(s_3)s_3 + h'(s_2)s_2}{h'(s_2) - h'(s_3)}. \tag{4.53}$$

Therefore, the envelope function h_u is given by

$$\begin{aligned} h_u(s) &= h(s_1) + h'(s_1)(s - s_1) \quad \text{if} \quad u_0 \le s < u_1 \\ &= h(s_2) + h'(s_2)(s - s_2) \quad \text{if} \quad u_1 \le s < u_2 \\ &= h(s_3) + h'(s_3)(s - s_3) \quad \text{if} \quad u_2 \le s < u_3. \end{aligned} \tag{4.54}$$

The result in (4.54) is obtained from (4.49), (4.50) and (4.51).

Step 3. To find the squeezing function h_l given by $u_0\, B\, D\, F\, u_3$, we have to find the equations of the lines u_0B, BD, DF and Fu_3.

The equation of u_0B is

$$y = h(u_0) + \frac{h(s_1) - h(u_0)}{(s_1 - u_0)}(s - u_0). \tag{4.55}$$

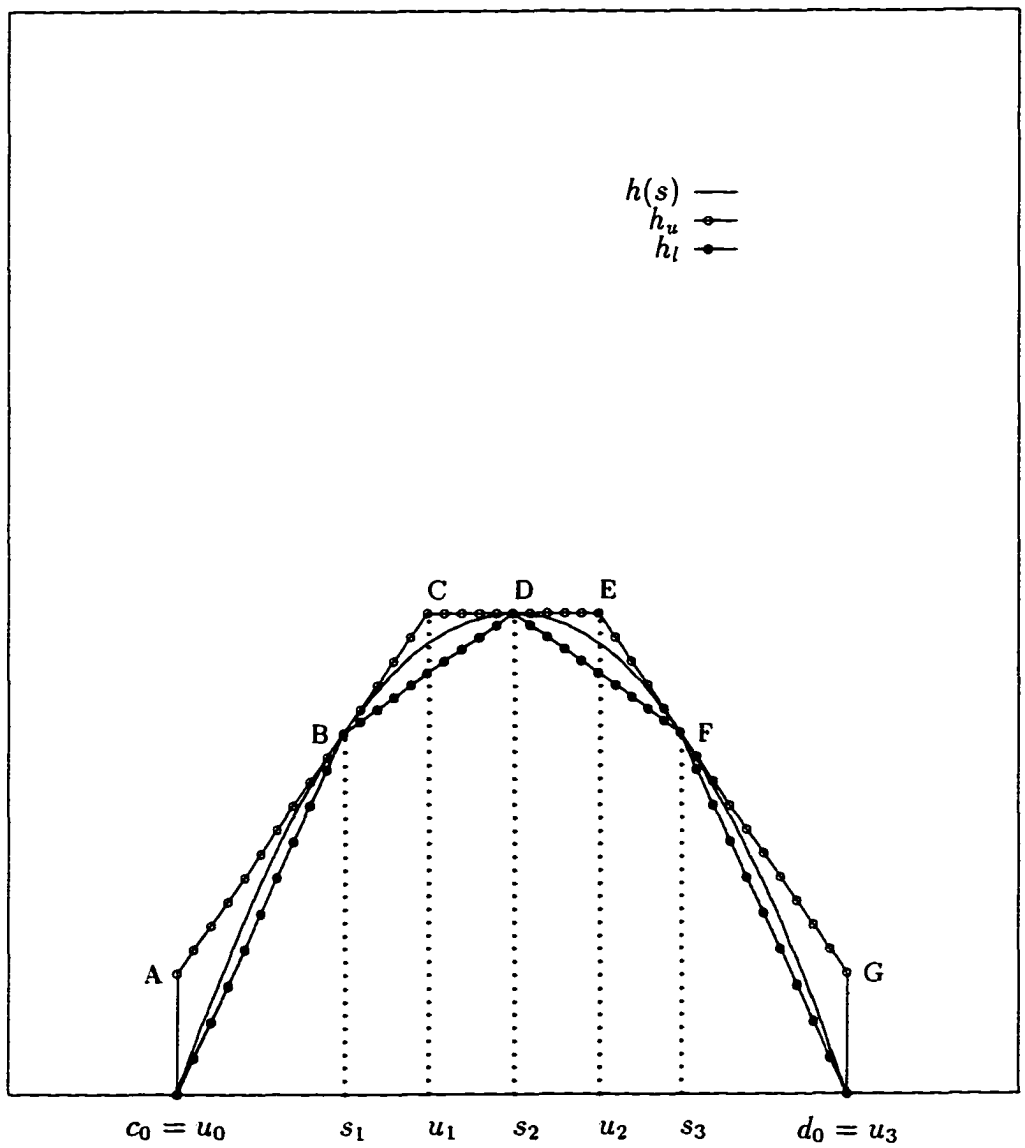

Figure 4.1: An Example of the Envelope (h_u) and Squeezing (h_l) Function of a Concave Function ($h(s)$).

The equation of BD is

$$y = h(s_1) + \frac{h(s_2) - h(s_1)}{(s_2 - s_1)}(s - s_1). \tag{4.56}$$

The equation of DF is

$$y = h(s_2) + \frac{h(s_3) - h(s_2)}{(s_3 - s_2)}(s - s_2) \tag{4.57}$$

and the equation of Fu_3 is

$$y = h(s_3) + \frac{h(u_3) - h(s_3)}{(u_3 - s_3)}(s - s_3). \tag{4.58}$$

Therefore, the squeezing function $h_l(s)$ is given by

$$\begin{aligned} h_l(s) &= h(u_0) + \frac{h(s_1) - h(u_0)}{(s_1 - u_0)}(s - u_0) \quad \text{if} \quad u_0 \leq s < s_1 \\ &= h(s_1) + \frac{h(s_2) - h(s_1)}{(s_2 - s_1)}(s - s_1) \quad \text{if} \quad s_1 \leq s < s_2 \\ &= h(s_2) + \frac{h(s_3) - h(s_2)}{(s_3 - s_2)}(s - s_2) \quad \text{if} \quad s_2 \leq s < s_3 \\ &= h(s_3) + \frac{h(u_3) - h(s_3)}{(u_3 - s_3)}(s - s_3) \quad \text{if} \quad s_3 \leq s < u_3. \end{aligned} \tag{4.59}$$

The result in (4.59) is obtained from (4.55), (4.56), (4.57) and (4.58).

Hence, we have determined $h_l(s)$ and $h_u(s)$, such that

$$h_l(s) \leq h(s) \leq h_u(s)$$

$$\Rightarrow exp(h_l(s)) \leq exp(h(s)) \leq exp(h_u(s))$$

where $c_0 \leq s \leq d_0$.

Step 4. To obtain the density function corresponding to $exp(h_u(s))$, it is necessary to find the area under $exp(h_u(s))$ between u_0 and u_3 where $u_0 = c_0$ and $u_3 = d_0$. Let A_1 be the area under $exp(h_u(s))$ between u_0 and u_1 then

$$\begin{aligned} A_1 &= \int_{u_0}^{u_1} exp(h(s_1) + h'(s_1)(s - s_1))ds \\ &= exp(h(s_1) - s_1h'(s_1)) \int_{u_0}^{u_1} \frac{h'(s_1)exp(sh'(s_1))}{h'(s_1)} ds \\ &= \frac{exp(h(s_1) - s_1h'(s_1))}{h'(s_1)} [exp(h'(s_1)u_1) - exp(h'(s_1)u_0)]. \end{aligned} \tag{4.60}$$

Let A_2 be the area under $exp(h_u(s))$ between u_1 and u_2 then

$$\begin{aligned} A_2 &= \int_{u_1}^{u_2} exp(h(s_2) + h'(s_2)(s - s_2))ds \\ &= \frac{exp(h(s_2) - s_2h'(s_2))}{h'(s_2)} [exp(h'(s_2)u_2) - exp(h'(s_1)u_1)]. \end{aligned} \tag{4.61}$$

Similarly, the area A_3 under $exp(h_u(s))$ between u_2 and u_3 is

$$A_3 = \frac{exp(h(s_3) - s_3h'(s_3))}{h'(s_3)} [exp(h'(s_3)u_3) - exp(h'(s_3)u_2)]. \tag{4.62}$$

Therefore, the density function $p(x)$ corresponding to $exp(h_u(s))$ is

$$\begin{aligned} p(x) &= \frac{exp(h(s_1) + h'(s_1)(s - s_1))}{A_1 + A_2 + A_3}, \quad u_0 \leq s < u_1 \\ &= \frac{exp(h(s_2) + h'(s_2)(s - s_2))}{A_1 + A_2 + A_3}, \quad u_1 \leq s < u_2 \\ &= \frac{exp(h(s_3) + h'(s_3)(s - s_3))}{A_1 + A_2 + A_3}, \quad u_2 \leq s < u_3. \end{aligned} \tag{4.63}$$

We have used (4.60), (4.61) and (4.62) to obtain (4.63).

Step 5. To generate X^* from the piecewise exponential density $p(x)$ given in (4.63), we first sample Z from

$$p(Z = k) = \frac{A_k}{A_1 + A_2 + A_3} \tag{4.64}$$

where $k = 1, 2, 3$. If $Z = k$ then generate X^* from

$$p_k(s) = \frac{h'(s_k)}{exp(u_k h'(s_k)) - exp(u_{k-1}h'(s_k))} exp(s\ h'(s_k)) \tag{4.65}$$

where $u_{k-1} \leq s < u_k$. We can generate X^* from $p_k(s)$ by generating a random variable U^* from Uniform(0, 1) and letting

$$X^* = \frac{\ln[U^* exp(u_k h'(s_k)) + (1 - U^*) exp(u_{k-1} h'(s_k)]}{h'(s_k)}. \tag{4.66}$$

It can be shown that $X^* \sim p(x)$ (see Berger and Sun (1993)).

Step 6. Next we generate U from Uniform(0, 1).

If $U \leq \frac{exp(h_l(X^*))}{exp(h_u(X^*))}$ then $X^* \sim f(\beta \mid \theta, \gamma, \underline{z}, \underline{\delta})$. Otherwise, compute $h(X^*)$ and

if $U \leq \frac{exp(h(X^*))}{exp(h_u(X^*))}$ then $X^* \sim f(\beta \mid \theta, \gamma, \underline{z}, \underline{\delta})$. If this fails then return to Step 5.

4.3.6 Estimation Using the Gibbs Sample

The estimates of the various posterior moments can be obtained, once we have generated the Gibbs sample. Suppose that $(\theta^{(j)}, \beta^{(j)}, \gamma^{(j)}; j = 1, \ldots, N)$ is a sample generated from the Gibbs sampler, then the posterior mean and variance of θ are estimated by

$$\hat{E}(\theta \mid \underline{z}, \underline{\delta}) = \frac{1}{N - N_1} \sum_{j=N_1+1}^{N} \theta^{(j)} \tag{4.67}$$

and

$$\hat{V}(\theta \mid \underline{z}, \underline{\delta}) = \frac{1}{N - N_1} \sum_{j=N_1+1}^{N} (\theta^{(j)})^2 - [\hat{E}(\theta \mid \underline{z}, \underline{\delta})]^2. \tag{4.68}$$

Also, the posterior marginal pdf of θ is estimated by

$$\hat{f}(\theta \mid \underline{z}, \underline{\delta}) = \frac{1}{N - N_1} \sum_{j=N_1+1}^{N} f(\theta \mid \beta^{(j)}, \gamma^{(j)} \underline{z}, \underline{\delta}), \tag{4.69}$$

where $f(\theta \mid \beta^{(j)}, \gamma^{(j)}, \underline{z}, \underline{\delta})$ is obtained from (4.35).

Similarly, estimates of the posterior mean, variance and marginal density of β are given by

$$\hat{E}(\beta \mid \underline{z}, \underline{\delta}) = \frac{1}{N - N_1} \sum_{j=N_1+1}^{N} \beta^{(j)}. \tag{4.70}$$

$$\hat{V}(\beta \mid \underline{z}, \underline{\delta}) = \frac{1}{N - N_1} \sum_{j=N_1+1}^{N} (\beta^{(j)})^2 - [\hat{E}(\beta \mid \underline{z}, \underline{\delta})]^2 \tag{4.71}$$

and

$$\hat{f}(\beta \mid \underline{z}, \underline{\delta}) = \frac{1}{N - N_1} \sum_{j=N_1+1}^{N} f(\beta \mid \theta^{(j)}, \gamma^{(j)}, \underline{z}, \underline{\delta}), \tag{4.72}$$

respectively where $f(\beta \mid \theta^{(j)}, \gamma^{(j)}, \underline{z}, \underline{\delta})$ is obtained from (4.38).

The predictive survivor function is estimated by

$$\hat{S}(t \mid \underline{z}, \underline{\delta}) = \frac{1}{N - N_1} \sum_{j=N_1+1}^{N} S(t \mid \theta^{(j)}, \beta^{(j)}, \underline{z}, \underline{\delta}). \tag{4.73}$$

4.3.7 Numerical Example

In this section, we show the implementation and efficiency of Gibbs sampler by a simulation study. We obtain the various posterior moments of θ, β and γ from the Gibbs sample.

We generate samples of size 20 and 60 from the Weibull distribution with $\theta = 0.5$ and $\beta = 1.5$ and with 20% ($\gamma = 0.25$), 33.3% ($\gamma = 0.5$) and 50% ($\gamma = 1$) censoring under the model assumptions given by (4.1) and (4.2). The data was generated in the following manner. The lifetimes $T_i's$ are generated from Weibull(θ_1, β) with $\theta_1 = 0.5$ and $\beta = 1.5$. Under the assumption of proportional hazards model, the corresponding censoring times $C_i's$ are generated from Weibull(θ_2, β) where $\theta_2 = \frac{\theta_1}{\gamma^{1/\beta}}$. If $T_i \leq C_i$ then take $Z_i = T_i$ and $\delta_i = 1$, otherwise we take $Z_i = C_i$ and $\delta_i = 0$. We repeat this n times to obtain a sample of size n.

We consider the following set of prior distributions of θ, β and γ.

(i) We first fix the prior of β. The prior of β is assumed to be truncated normal with mean 0 and standard deviation σ and the range of β is $[c_0, d_0]$, the pdf is as given in (4.25). The prior parameters σ, c_0 and d_0 are fixed based on prior information. One approach would be to consider the mean, variance and possibly another moment of the prior distribution and then choose σ, c and d (see Lee 1989, page 53). The

expected value and the variance of β (prior distribution) can be computed using the result of Johnson et al. 1994, see page 156. If $X \sim N(\mu, \sigma^2)$ and $a \leq X \leq b$ then

$$E(X) = \mu + \sigma \frac{Z(\frac{a-\mu}{\sigma}) - Z(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})}, \tag{4.74}$$

and

$$\begin{aligned} Var(X) &= \left[1 + \frac{(\frac{a-\mu}{\sigma})Z(\frac{a-\mu}{\sigma}) - (\frac{b-\mu}{\sigma})Z(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})}\right]\sigma^2 \\ &\quad - \left[\left(\frac{Z(\frac{a-\mu}{\sigma}) - Z(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})}\right)^2\right]\sigma^2 \end{aligned} \tag{4.75}$$

where $Z(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

We choose $\sigma = 1$, $c_0 = 1$ and $d_0 = 4$ such that the expected value of β is close to 1.5.

(ii) Once the prior distribution of β is fixed, we can fix the prior distribution of θ. We assume $\theta^\beta|\beta$ to be inverted gamma with parameters a and b. We fix b the scale parameter to be 1 and a is chosen such that the mean

$$E(\theta \mid \beta) = \frac{b^{\frac{1}{\beta}}\Gamma(a - \frac{1}{\beta})}{\Gamma(a)} \tag{4.76}$$

and the variance

$$V(\theta \mid \beta) = \frac{b^{\frac{2}{\beta}}}{(\Gamma(a))^2}(\Gamma(a)\Gamma(a - \frac{2}{\beta}) - (\Gamma(a - \frac{1}{\beta}))^2) \tag{4.77}$$

of the prior distribution of $(\theta \mid \beta)$ exist. $E(\theta \mid \beta)$ is defined if only if $a > \frac{1}{\beta}$ and $V(\theta \mid \beta)$ is defined if and only if $a > \frac{2}{\beta}$. Since we have fixed $1 \leq \beta \leq 4$, we choose $a = 11$ such that the above mentioned conditions are satisfied.

(iii) We assume the prior distribution of the censoring parameter γ to be truncated exponential with pdf given in (4.28). We can choose η and k by considering the expected value

$$E(\gamma) = \frac{1}{\eta} - \frac{kexp(-\eta k)}{1 - exp(-\eta k)} \tag{4.78}$$

and the variance of γ

$$V(\gamma) = \frac{1}{\eta^2} - \frac{k^2 exp(-\eta k)}{(1 - exp(-\eta k))^2}. \tag{4.79}$$

Here we choose $\eta = 2$ and $k = 3$ such that the expected value of γ is approximately 0.5.

Figures 4.2, 4.3 and 4.4 show the plots of the Gibbs sequence of θ, β and γ for 20%, 33.3% and 50% censoring, respectively. From the plots of the Gibbs sequence we see that 1500 iterations of the Gibbs sampler was quite satisfactory to achieve convergence of the posterior quantities.

The posterior mean, variance and marginal posterior density of θ are estimated by (4.67), (4.68) and (4.69), respectively and that of β are estimated by (4.70), (4.71) and (4.72), respectively. The prior and posterior moments of θ for 20%, 33.3% and 50% censoring are given in Table 4.1 and Table 4.2 for sample sizes of 20 and 60, respectively. The corresponding marginal prior and posterior densities are compared in Figure 4.5. Table 4.3 and Table 4.4 give the prior and posterior moments of β for 20%, 33.3% and 50% censoring and for sample sizes 20 and 60, respectively. The corresponding marginal prior and posterior densities are compared in Figure 4.6. Similarly, the prior and posterior moments of γ are given in Table 4.5 and Table 4.6 for sample sizes of 20 and 60, respectively. The corresponding marginal prior and posterior densities are compared in Figure 4.7.

Usually, we expect that data with a smaller percentage of censored observations gives better estimates of the parameters as compared to data with a higher percentage of censored observation. We can see from the tables of posterior moments of θ, β and γ that it is not always true. One possible reason would be that under our model assumptions same parameters are involved both in the lifetime distribution and censoring time distribution.

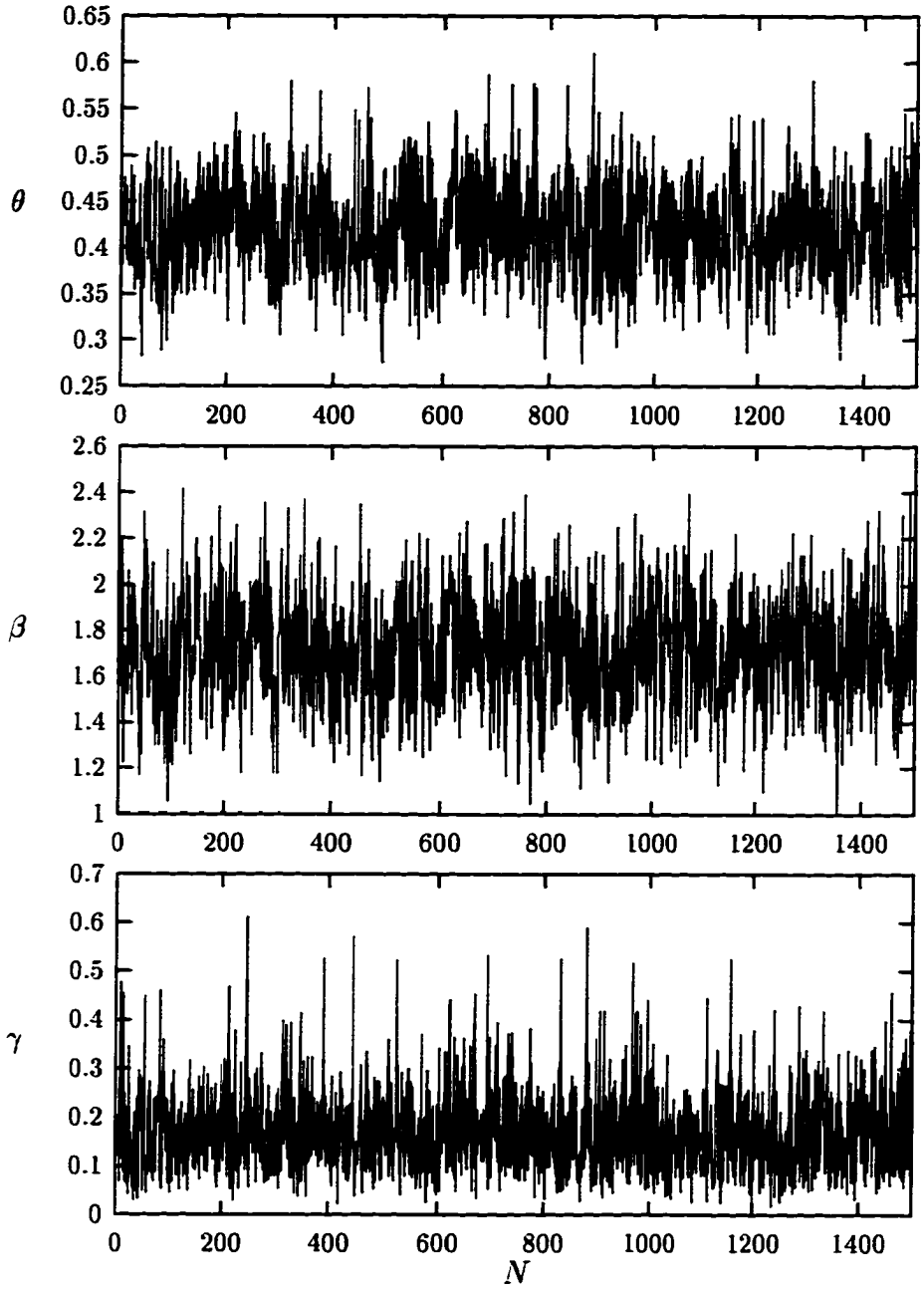

Figure 4.2: Plots of Gibbs Sequence of θ, β and γ for 20% Censoring.

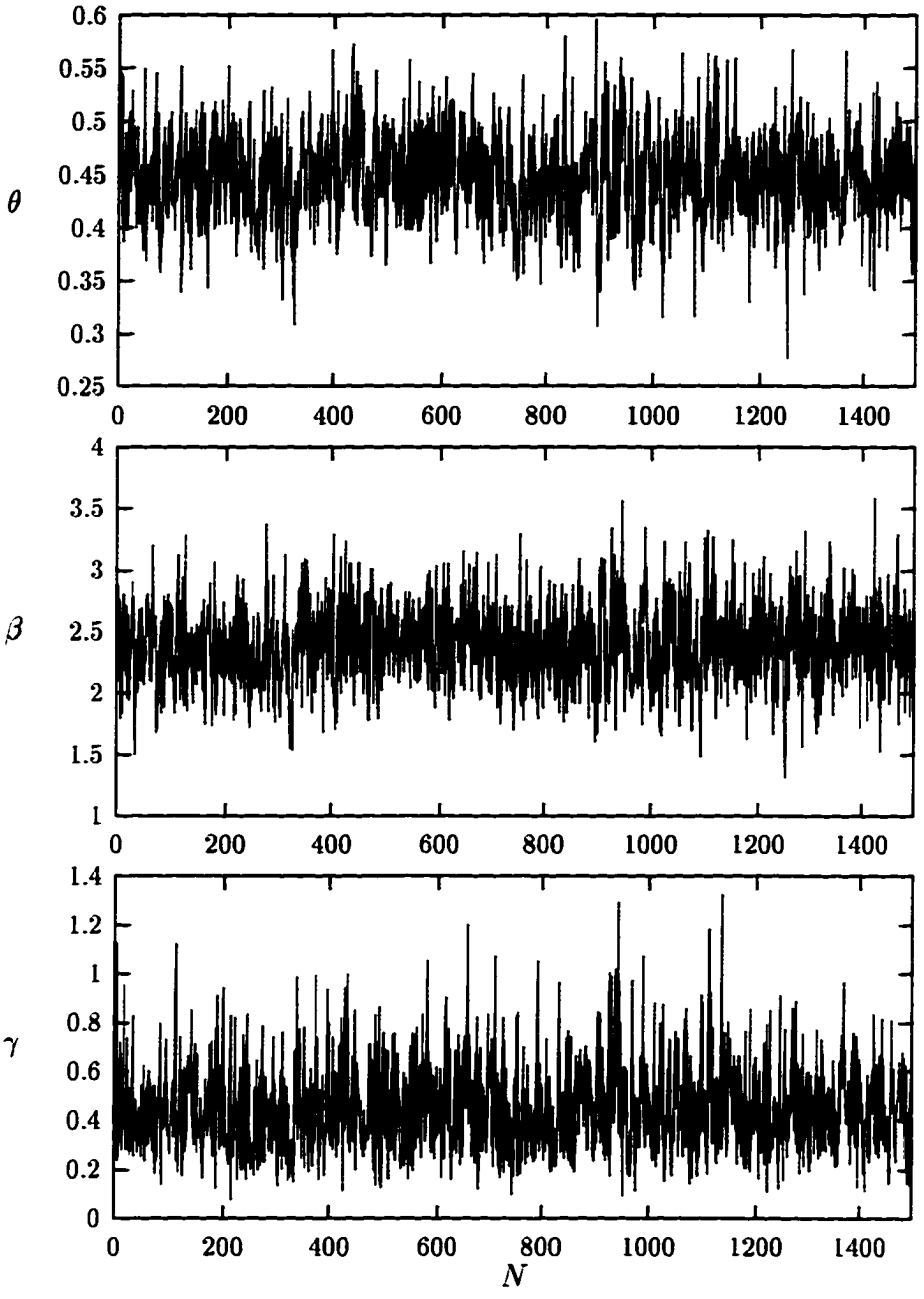

Figure 4.3: Plots of Gibbs Sequence of θ, β and γ for 33.3% Censoring.

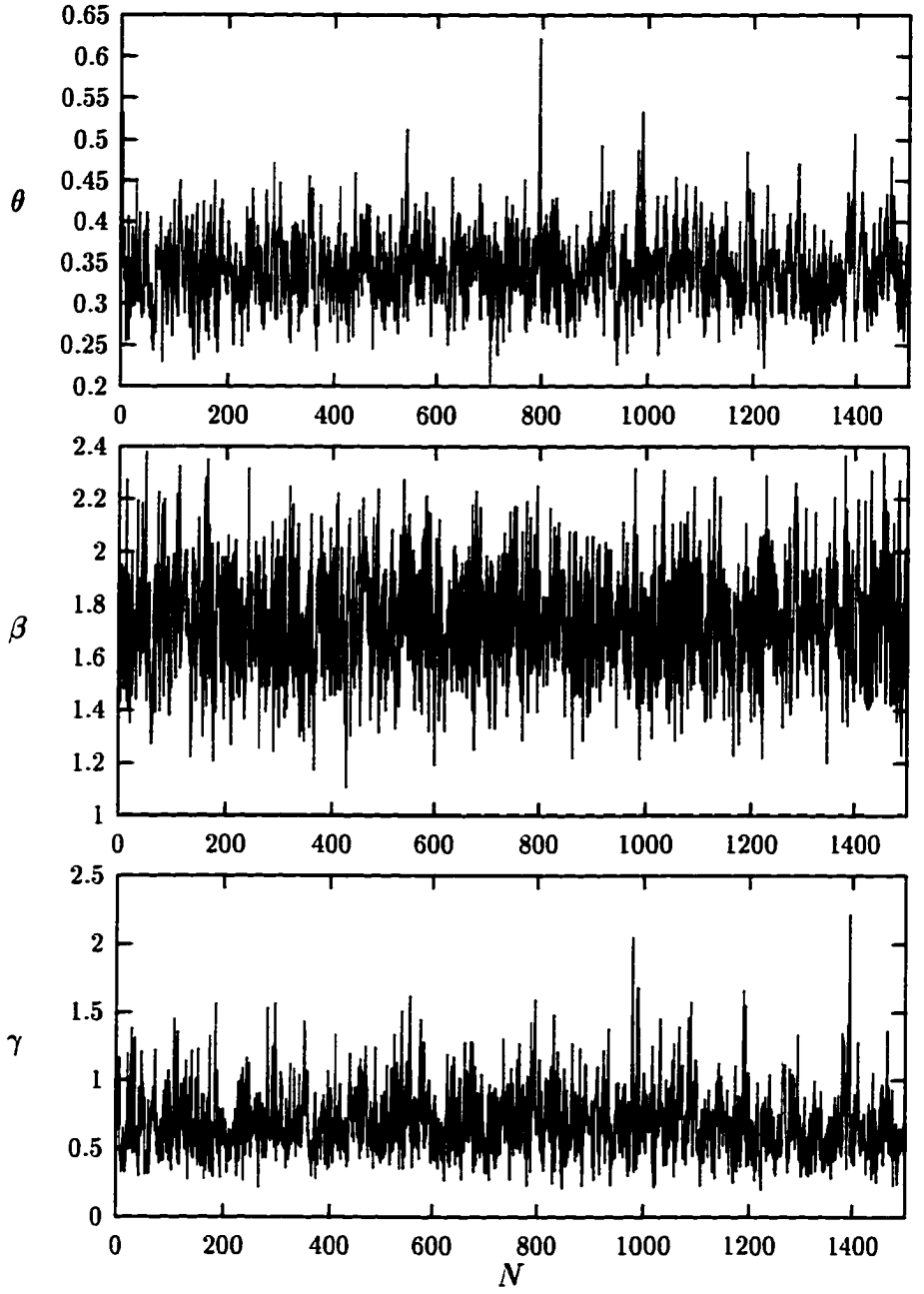

Figure 4.4: Plots of Gibbs Sequence of θ, β and γ for 50% Censoring.

Table 4.1: Prior and Posterior Moments of θ for Sample Size 20.

Prior distribution	Posterior distribution			
		20% censoring	33.3% censoring	50% censoring
$E(\theta) = 0.2117$	$E(\theta \mid data)$	0.4145	0.4459	0.3399
$\sqrt{V(\theta)} = 0.09662$	$\sqrt{V(\theta \mid data)}$	0.0506	0.0426	0.0457

Table 4.2: Prior and Posterior Moments of θ for Sample Size 60.

Prior distribution	Posterior distribution			
		20% censoring	33.3% censoring	50% censoring
$E(\theta) = 0.2117$	$E(\theta \mid data)$	0.5040	0.5242	0.4205
$\sqrt{V(\theta)}$	$\sqrt{V(\theta \mid data)}$	0.0472	0.0548	0.0431

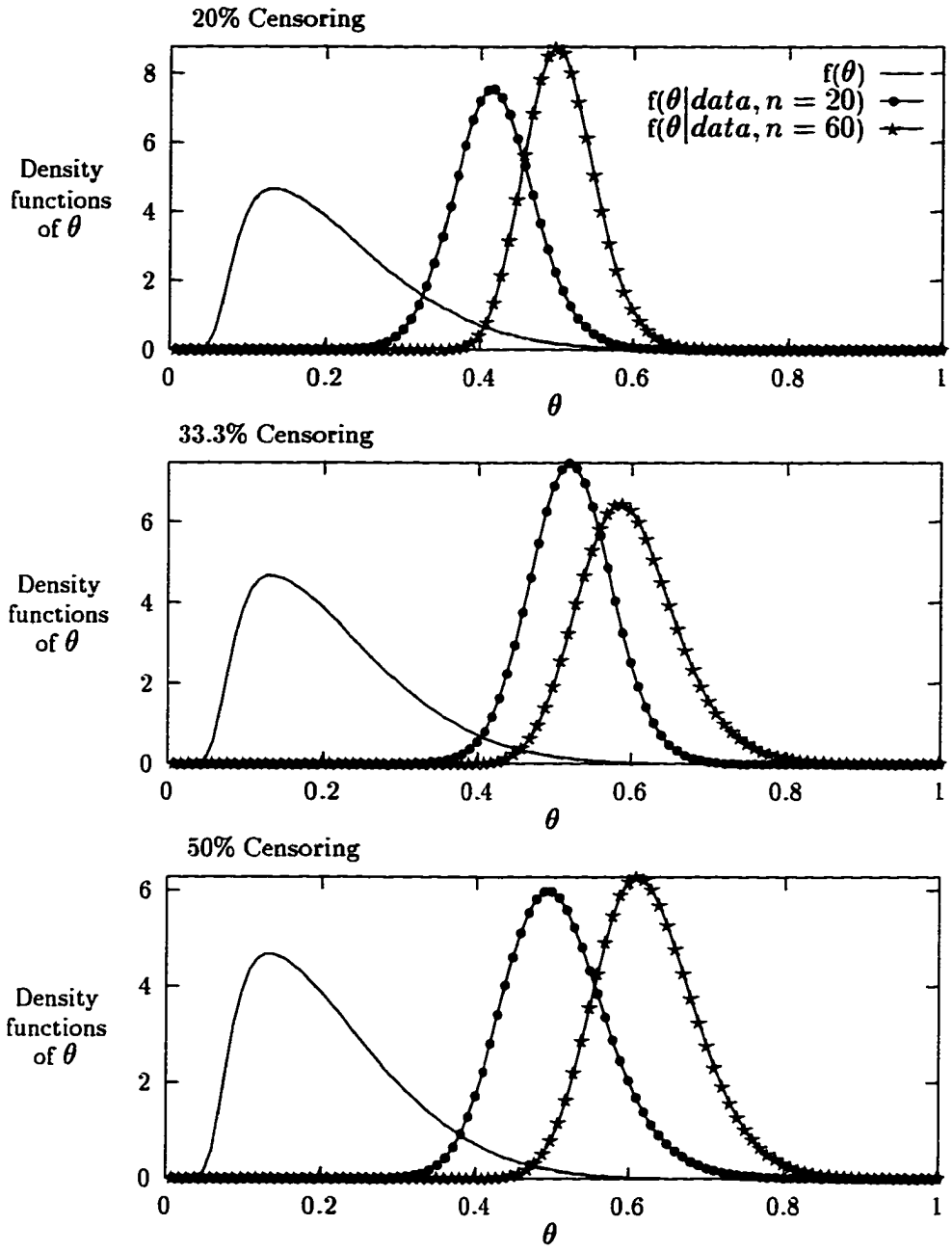

Figure 4.5: Marginal Prior and Posterior Densities of θ.

Table 4.3: Prior and Posterior Moments of β for Sample Size 20.

Prior distribution	Posterior distribution			
		20% censoring	33.3% censoring	50% censoring
$E(\beta) = 1.5229$	$E(\beta \mid data)$	1.7093	2.394	1.7437
$\sqrt{V(\beta)} = 0.4428$	$\sqrt{V(\beta \mid data)}$	0.2402	0.3419	0.2233

Table 4.4: Prior and Posterior Moments of β for Sample Size 60.

Prior distribution	Posterior distribution			
		20% censoring	33.3% censoring	50% censoring
$E(\beta) = 1.5229$	$E(\beta \mid data)$	1.5610	1.4433	1.4812
$\sqrt{V(\beta)} = 0.4428$	$\sqrt{V(\beta \mid data)}$	0.1433	0.1414	0.1353

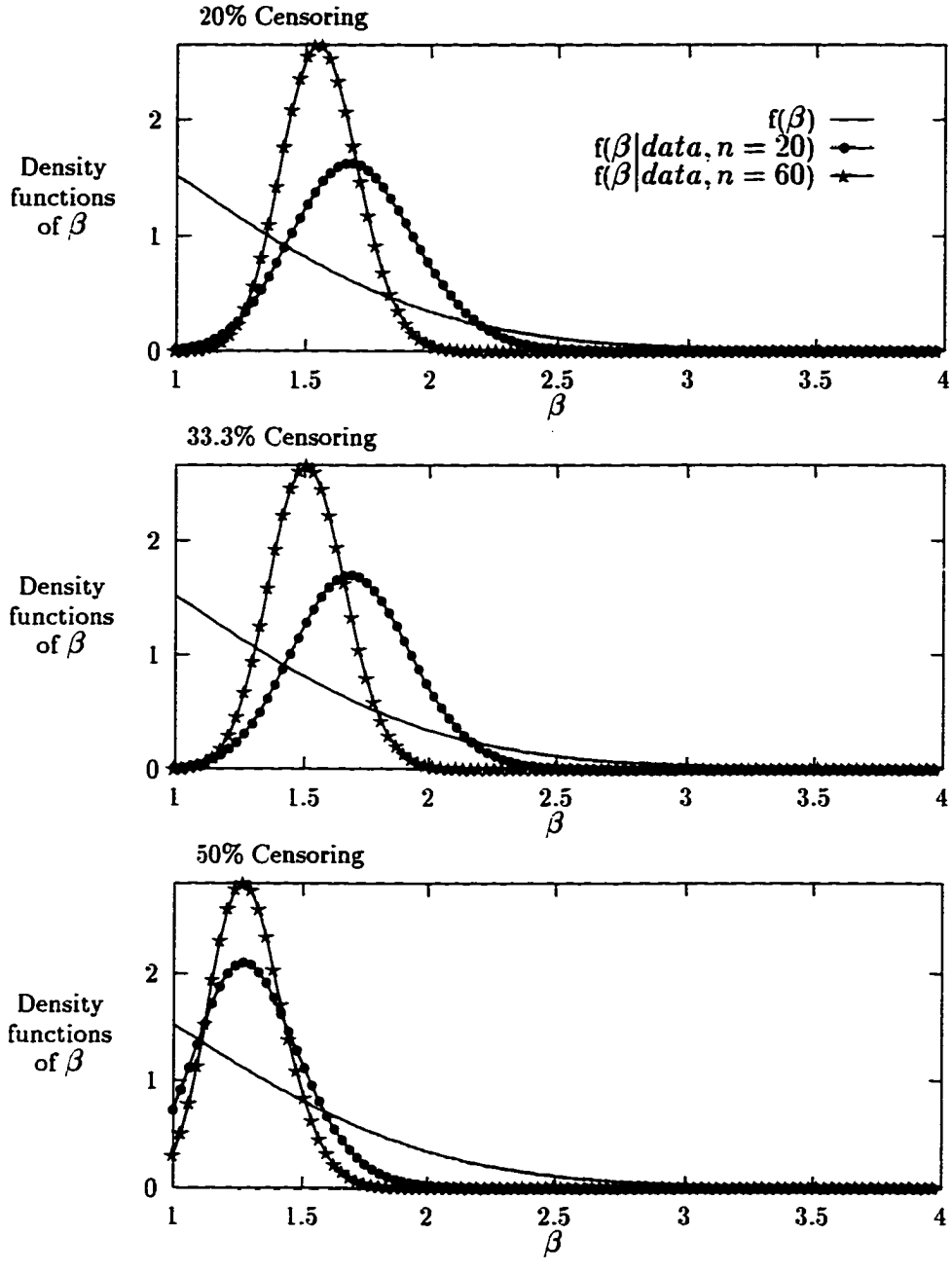

Figure 4.6: Marginal Prior and Posterior Densities of β.

Table 4.5: Prior and Posterior Moments of γ for Sample Size 20.

Prior distribution	Posterior distribution			
		20% censoring	33.3% censoring	50% censoring
$E(\gamma) = 0.4925$	$E(\gamma \mid data)$	0.1629	0.4513	0.6757
$\sqrt{V(\gamma)} = 0.4771$	$\sqrt{V(\gamma \mid data)}$	0.0830	0.1773	0.2564

Table 4.6: Prior and Posterior Moments of γ for Sample Size 60.

Prior distribution	Posterior distribution			
		20% censoring	33.3% censoring	50% censoring
$E(\gamma) = 0.4925$	$E(\gamma \mid data)$	0.3187	0.7271	0.8462
$\sqrt{V(\gamma)} = 0.4771$	$\sqrt{V(\gamma \mid data)}$	0.0981	0.1758	0.1877

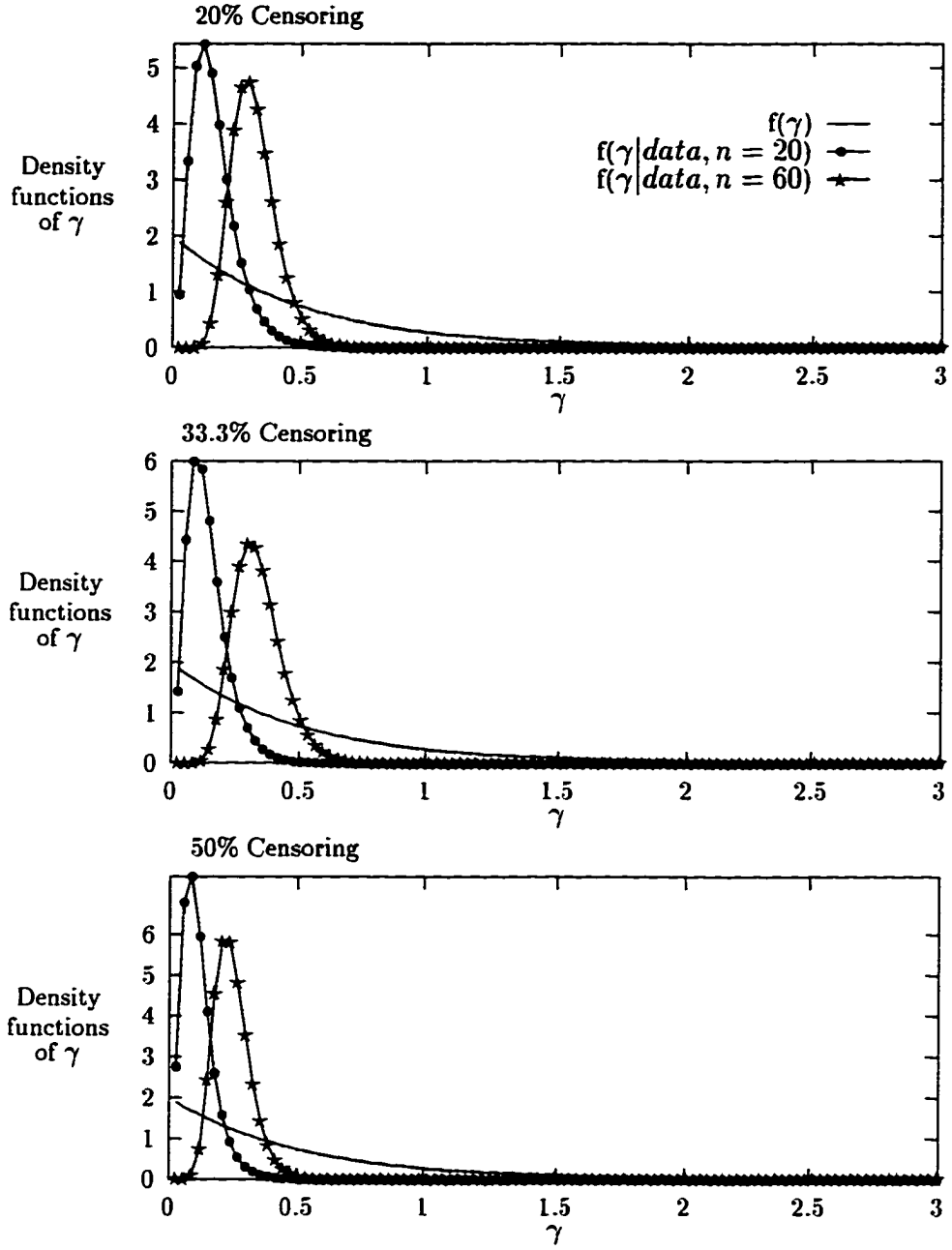

Figure 4.7: Marginal Prior and Posterior Densities of γ.

4.3.8 Comparison of MLE with the Bayes Estimator of the Survivor Function

In this section, we compare the MLE with the Bayes estimator of the survivor function for the Koziol-Green model given by (4.1) and (4.2) with Weibull lifetimes given by (4.4). We make the comparison by a simulation study.

Let $\tilde{S}(t)$ be the MLE of the survivor function in (4.5) given by (4.16) and $\hat{S}(t)$ be the Bayes estimate of the survivor function in (4.5) given by (4.73). For the simulation study, we generate 100 samples of size 20 each from Weibull distributions under the model assumptions (4.1) and (4.2) with $\theta = 0.5$ and $\beta = 1.5$ for 20%, 33.3% and 50% censoring. We make the comparisons by considering four sets of prior distributions of (θ, β, γ) to obtain the Bayes estimate $\hat{S}(t)$ of the survivor function.

Set 1. The prior distribution of β is assumed to be truncated normal with mean 0, standard deviation $\sigma = 2$ and pdf as given in (4.25). We assume $c_0 = 0.1$ and $d_0 = 6$. In this case the expected value of β works out to be greater than 1.5. For the prior distribution of θ, we assume $(\theta^{\beta}|\beta)$ to be IG($a = 11$, $b = 1$); the pdf of $(\theta|\beta)$ is given in (4.24). We assume the prior distribution of the censoring parameter γ to be truncated exponential with $\eta = 2$, $K = 3$ and pdf given in (4.28); the expected value of γ being close to 0.5.

Set 2. The prior distribution of β is considered to be truncated gamma with shape parameter $g_a = 3$, scale parameter $g_b = 1$ and pdf given in (4.30); the range of β is fixed again as $c_0 = 0.1$ and $d_0 = 6$. The expected value of β, in this case, is 2.72. The prior distribution of θ is considered to be the same as in Set 1. The prior distribution of γ is assumed to be Uniform(0, 3) with pdf given by (4.31).

Set 3. The prior distribution of β is assumed to be truncated normal with mean 0, standard deviation $\sigma = 1$, $c_0 = 1$ and $d_0 = 4$. So that the expected value of β is close to 1.5. The prior distributions of θ and γ are considered to be the same as in Set 1.

Set 4. The prior distribution of β is considered to be truncated gamma with shape parameter $g_a = 3$, scale parameter $g_b = 3$, $c_0 = 1$ and $d_0 = 4$. So that the expected

value of β is close to 1.53. The prior distribution of θ is considered to be the same as in Set 1. The prior distribution of γ is assumed to be Uniform(0, 3).

Note that, for fixing the parameters of the prior distributions and the ranges of θ, β and γ, we follow the same principles discussed in Section 4.3.7.

To compare the MLE and the Bayes estimator of the survivor function, we compute the mean square error (MSE) of $\tilde{S}(t)$ and the MSE of $\hat{S}(t)$ which are given by

$$MSE(\tilde{S}(t)) = \frac{1}{M}\sum_{i=1}^{M}(\tilde{S}(t) - S_T(t))^2$$

and

$$MSE(\hat{S}(t)) = \frac{1}{M}\sum_{i=1}^{M}(\hat{S}(t) - S_T(t))^2,$$

respectively where $M = 100$.

Figures 4.8 to 4.11 show the ratios of the MSEs of the MLE and Bayes estimator of the survivor function with 20%, 33.3% and 50% censoring for Set 1, Set 2, Set 3 and Set 4 of prior distributions, respectively. Figures 4.12 to 4.14 show the ratios of the MSEs of the MLE and Bayes estimator of the survivor function with the four different sets of prior distributions for 20%, 33.3% and 50% censoring, respectively. From all the figures we see that the Bayes estimator $\hat{S}(t)$ does better than the MLE $\tilde{S}(t)$ at the tails (left and right tails) of the distribution of the lifetimes for all the four sets of prior distributions and all the three percentages of censoring considered. This indicates that prior information plays an important role at the tails because of the lack of availability of information from observed data at the tails of the distribution. As expected, the MLE $\tilde{S}(t)$ does better than the Bayes estimator $\hat{S}(t)$ in the middle of the lifetime distribution.

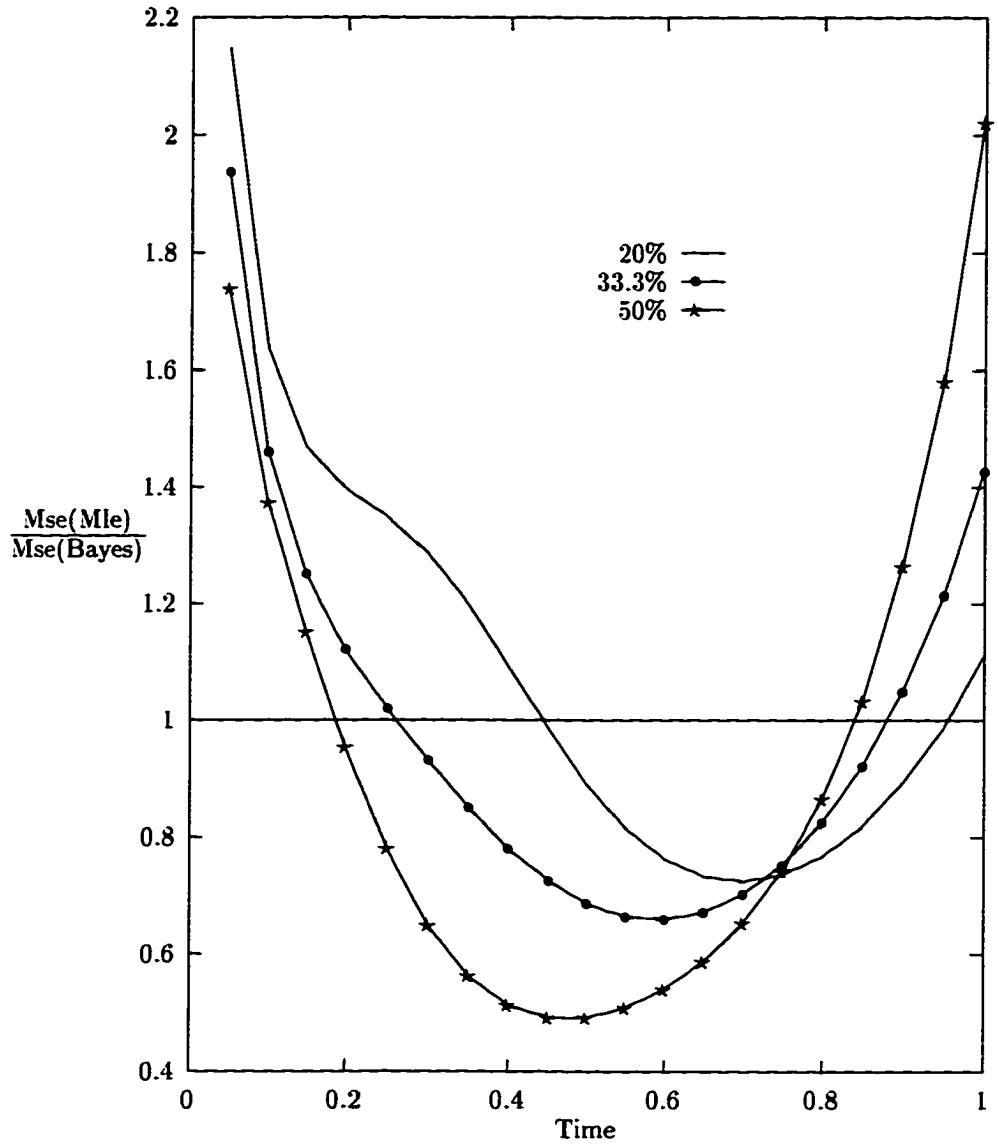

Figure 4.8: Ratios of MSEs of MLE and Bayes Estimator of Survivor Function for Randomly Censored Weibull Lifetime Data with Set 1 of Priors for Sample Size 20.

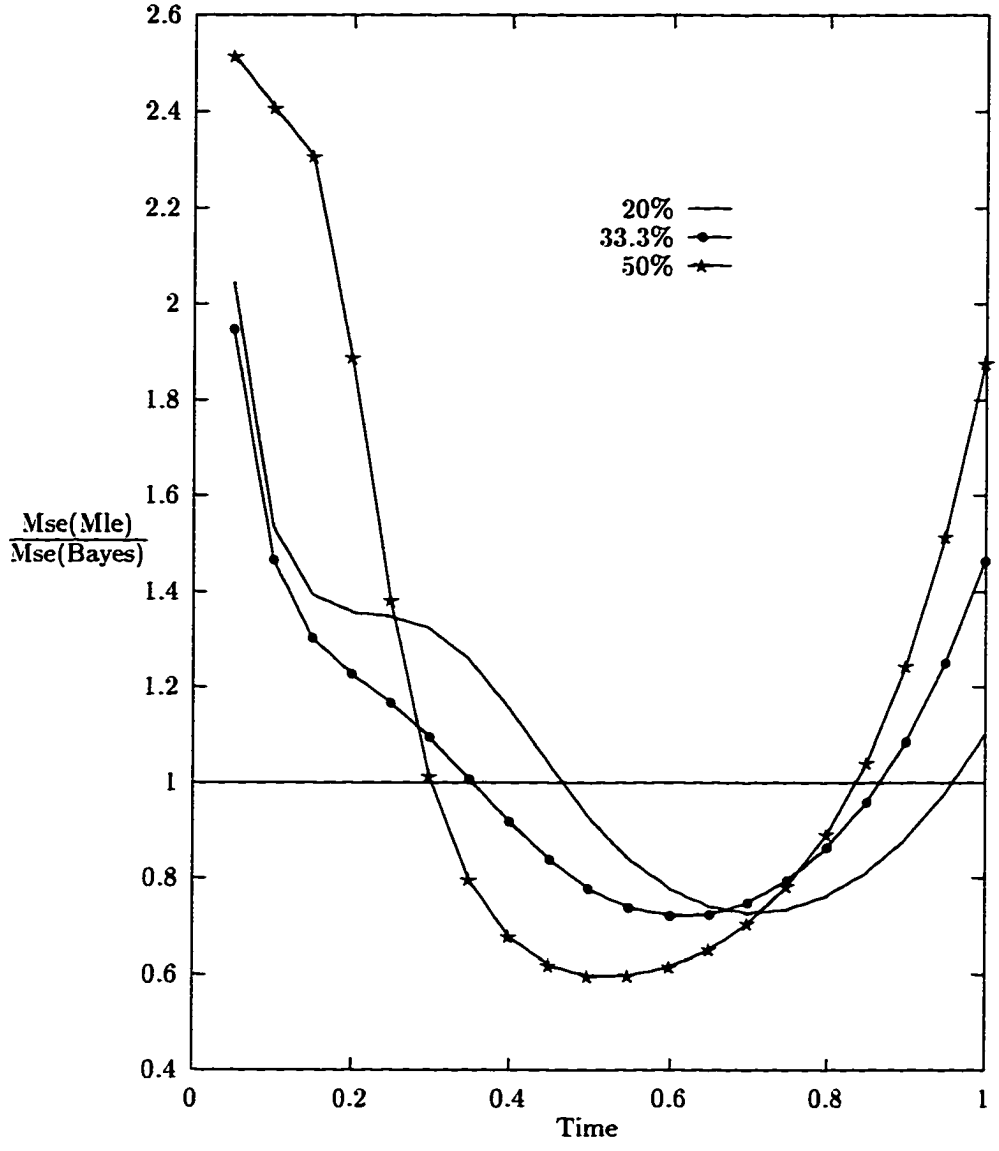

Figure 4.9: Ratios of MSEs of MLE and Bayes Estimator of Survivor Function for Randomly Censored Weibull Lifetime Data with Set 2 of Priors for Sample Size 20.

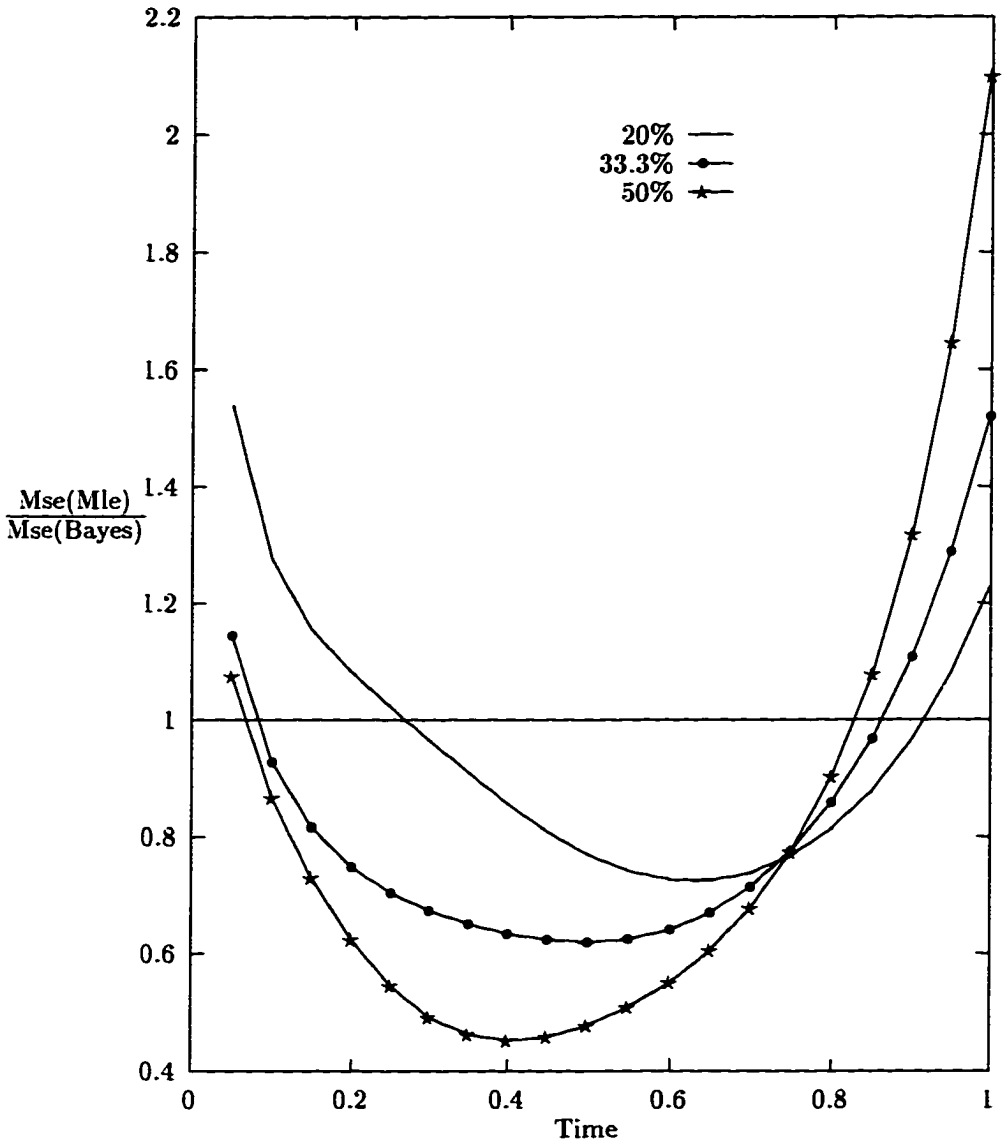

Figure 4.10: Ratios of MSEs of MLE and Bayes Estimator of Survivor Function for Randomly Censored Weibull Lifetime Data with Set 3 of Priors for Sample Size 20.

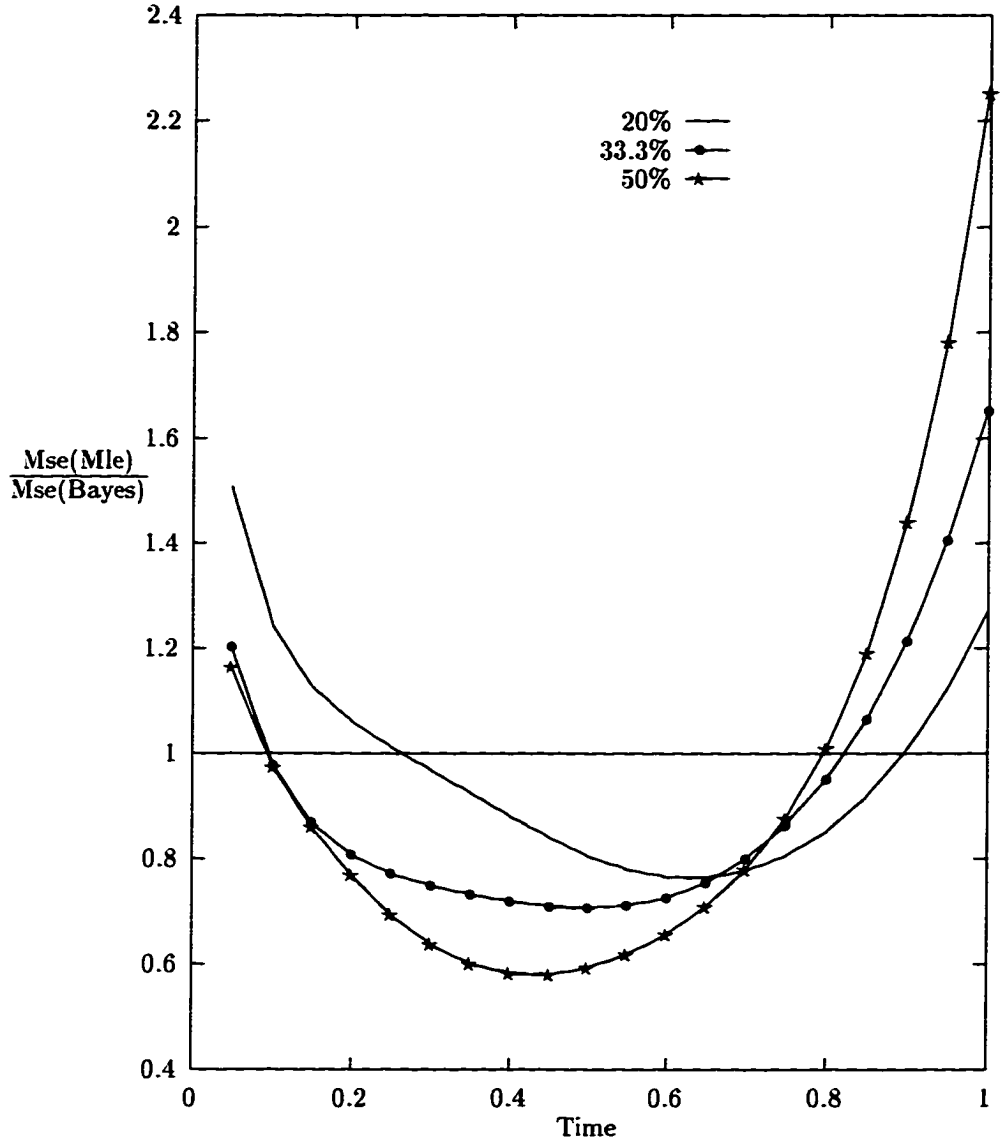

Figure 4.11: Ratios of MSEs of MLE and Bayes Estimator of Survivor Function for Randomly Censored Weibull Lifetime Data with Set 4 of Priors for Sample Size 20.

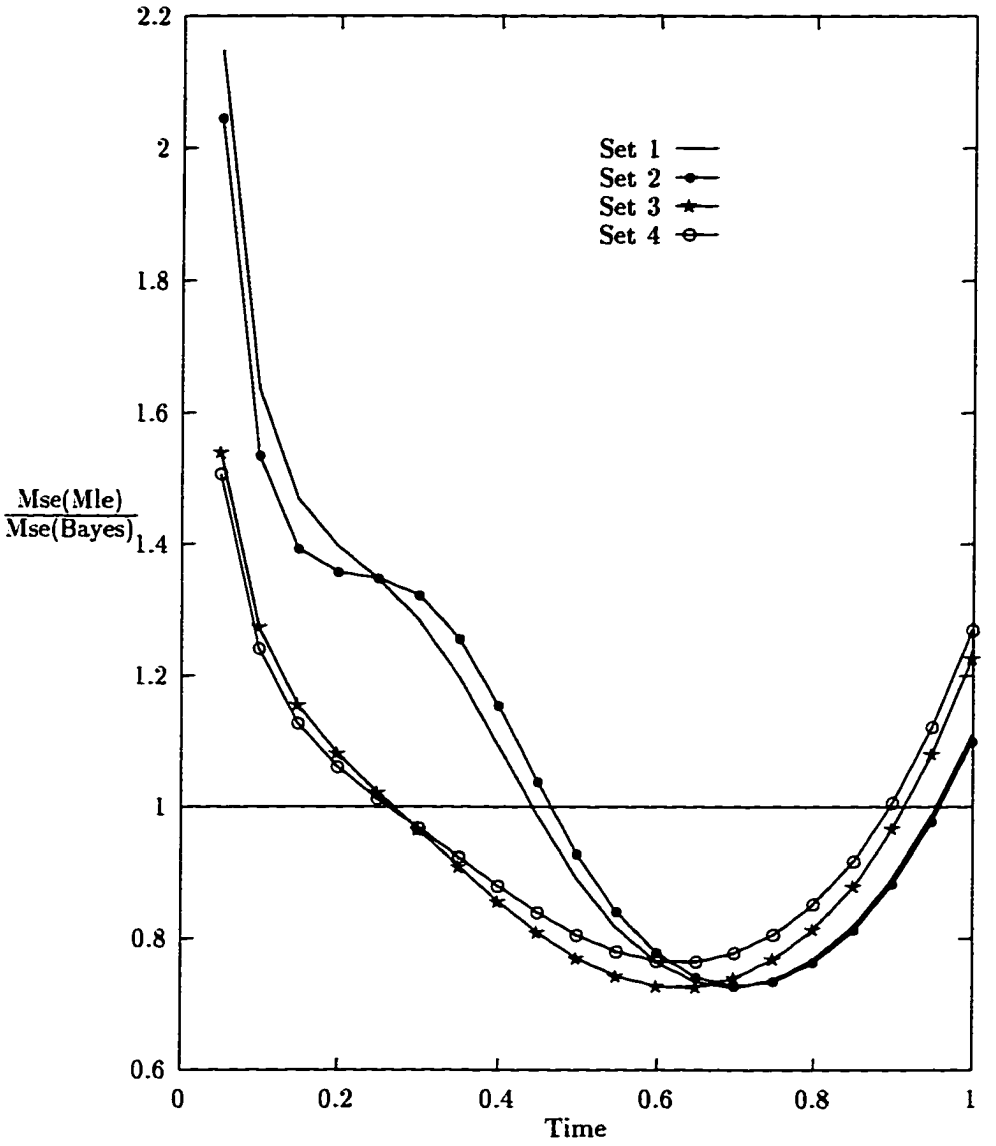

Figure 4.12: Ratios of MSEs of MLE and Bayes Estimator of Survivor Function for 20% Censored Weibull Lifetime Data with Different Sets of Priors for Sample Size 20.

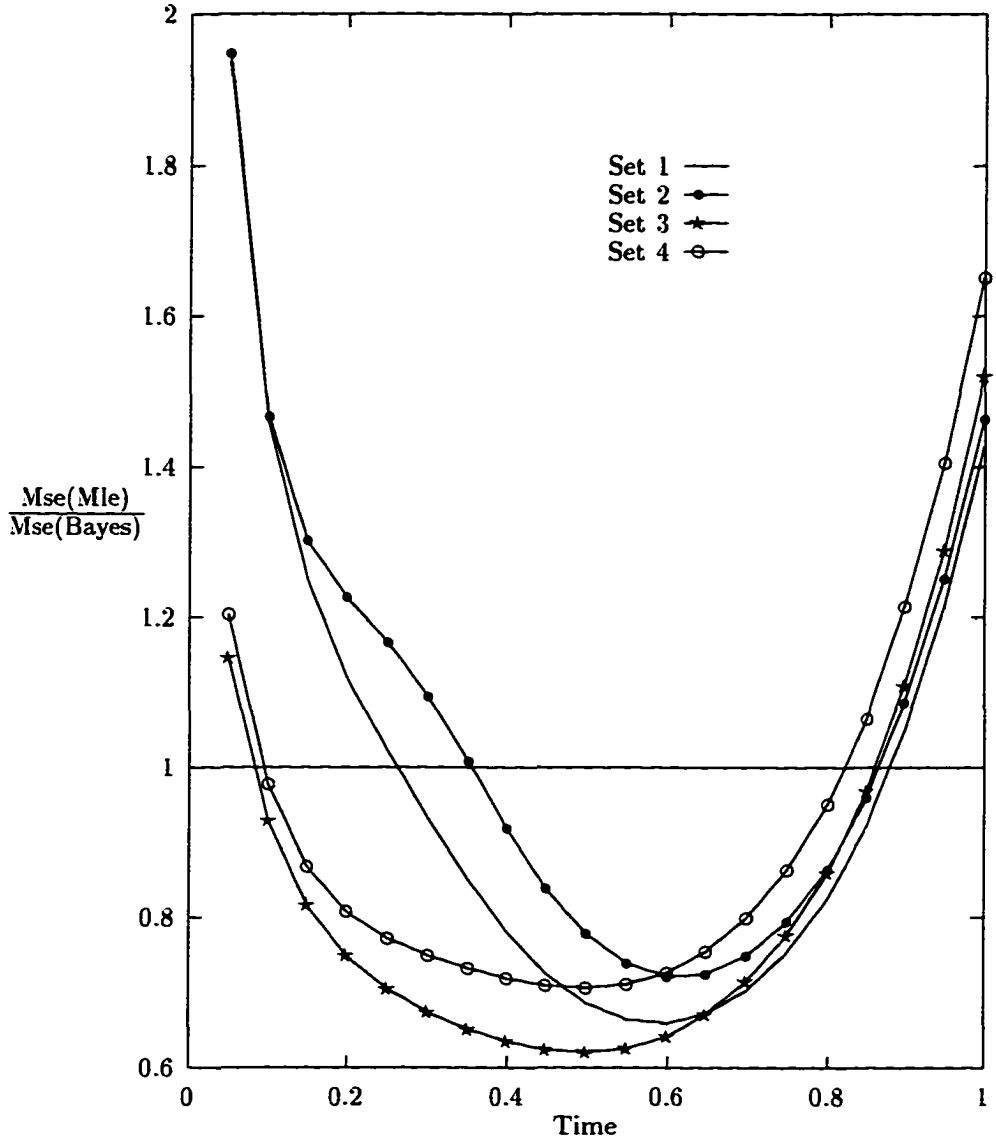

Figure 4.13: Ratios of MSEs of MLE and Bayes Estimator of Survivor Function for 33.3% Censored Weibull Lifetime Data with Different Sets of Priors for Sample Size 20.

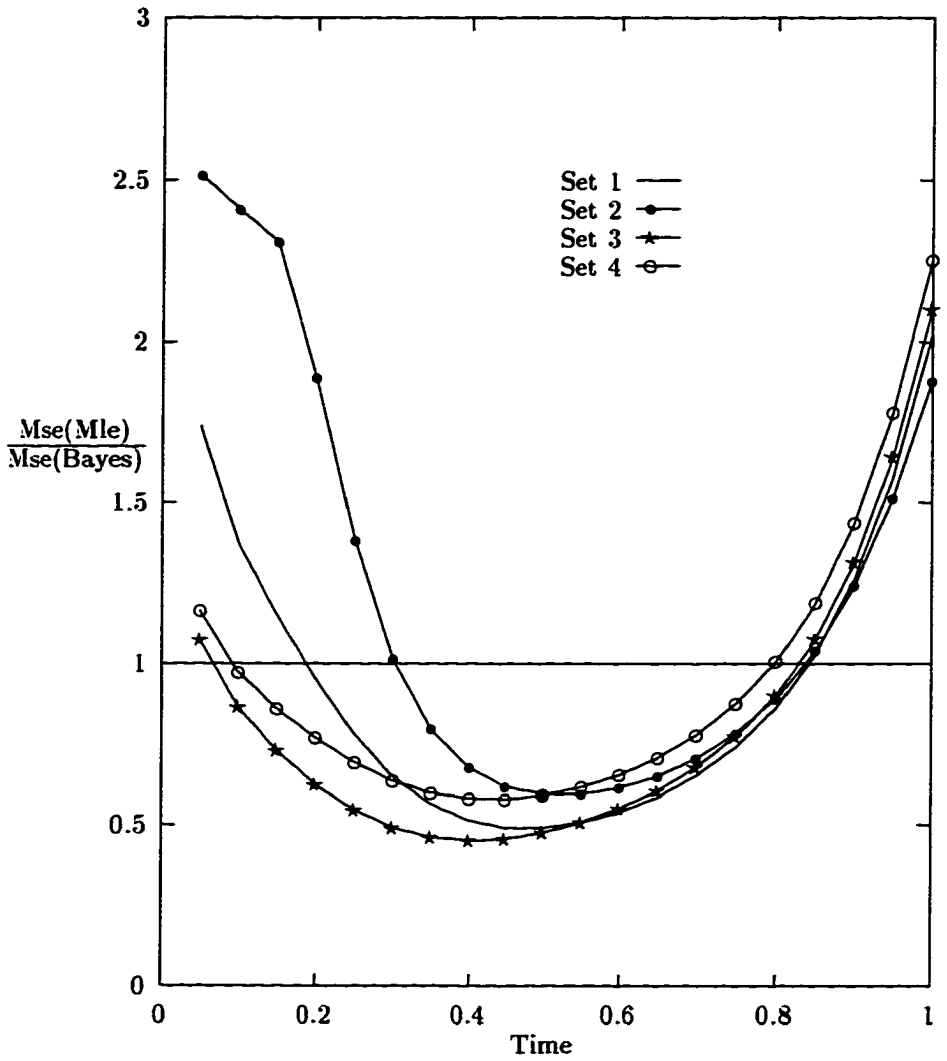

Figure 4.14: Ratios of MSEs of MLE and Bayes Estimator of Survivor Function for 50% Censored Weibull Lifetime Data with Different Sets of Priors for Sample Size 20.

4.3.9 An Example

In this section, we analyze a data set using MLE and Bayes estimation under the assumption of the Koziol-Green model as discussed in the preceding sections of this chapter.

The data given in Table 4.7 is obtained from Collett (1994, page 9). The data relates to the survival times and values of certain explanatory variables of 48 patients with multiple myeloma aged between 50 and 80 years. Multiple myeloma is a malignant disease characterized by the accumulation of abnormal plasma cells in the bone marrow.

The response variable here is time in months from diagnosis until death from multiple myeloma. The survival status of a patient is coded such that zero denotes a censored observation and unity denotes death from the above mentioned disease. Values of seven covariates are listed for each of the patients in the table. In Chapter 5, we consider the effect of these covariates on the survival times of the patients suffering from multiple myeloma. For now, we consider only the survival times and the survival status of the 48 multiple myeloma patients (i.e., the data in the second and third column of Table 4.7). The observations in column two correspond to $Z_i's$ for $i = 1,\ldots,48$. Here $Z_i's$ denote observed death times or censored times of the 48 patients. The observations in column three correspond to $\delta_i's$ for $i = 1,\ldots,48$, indicating the survival status of the patients.

In order to test whether the data in Table 4.7 is a good fit for the Koziol-Green model with Weibull lifetimes described by (4.1) and (4.2), we test for the fact that under the assumptions of the Koziol-Green model, the conditional distribution of Z given $\delta = 0$ is identical to the conditional distribution of Z given $\delta = 1$, which are in turn identical to the marginal distribution of Z. We found sufficient evidence to support the above fact for the data. The fit of the conditional distributions of Z given δ and the marginal distribution of Z are tested using the Kolmogrov's D test.

We obtain the MLE and Bayes estimates of the parameters and the standard errors of the corresponding estimators. We also obtain the MLE, Bayes estimate

Table 4.7: Survival Times of Patients in a Study on Multiple Myeloma.

Patient no.	Survival time	Status	AGE	SEX	BUN	CA	HB	PC	BJ
1	13	1	66	1	25	10	14.6	18	1
2	52	0	66	1	13	11	12	100	0
3	6	1	53	2	15	13	11.4	33	1
4	40	1	69	1	10	10	10.2	30	1
5	10	1	65	1	20	10	13.2	66	0
6	7	0	57	2	12	8	9.9	45	0
7	66	1	52	1	21	10	12.8	11	1
8	10	0	60	1	41	9	14	70	1
9	10	1	70	1	37	12	7.5	47	0
10	14	1	70	1	40	11	10.6	27	0
11	16	1	68	1	39	10	11.2	41	0
12	4	1	50	2	172	9	10.1	46	1
13	65	1	59	1	28	9	6.6	66	0
14	5	1	60	1	13	10	9.7	25	0
15	11	0	66	2	25	9	8.8	23	0
16	10	1	51	2	12	9	9.6	80	0
17	15	0	55	1	14	9	13	8	0
18	5	1	67	2	26	8	10.4	49	0
19	76	0	60	1	12	12	14	9	0
20	56	0	66	1	18	11	12.5	90	0
21	88	1	63	1	21	9	14	42	1
22	24	1	67	1	10	10	12.4	44	0
23	51	1	60	2	10	10	10.1	45	1
24	4	1	74	1	48	9	6.5	54	0
25	40	0	72	1	57	9	12.8	28	1

Table 4.7(continued): Survival Times of Patients in a Study on Multiple Myeloma.

Patient no.	Survival time	Status	AGE	SEX	BUN	CA	HB	PC	BJ
26	8	1	55	1	53	12	8.2	55	0
27	18	1	51	1	12	15	14.4	100	0
28	5	1	70	2	130	8	10.2	23	0
29	16	1	53	1	17	9	10	28	0
30	50	1	74	1	37	13	7.7	11	1
31	40	1	70	2	14	9	5	22	0
32	1	1	67	1	165	10	9.4	90	0
33	36	1	63	1	40	9	11	16	1
34	5	1	77	1	23	8	9	29	0
35	10	1	61	1	13	10	14	19	0
36	91	1	58	2	27	11	11	26	1
37	18	0	69	2	21	10	10.8	33	0
38	1	1	57	1	20	9	5.1	100	1
39	18	0	59	2	21	10	13	100	0
40	6	1	61	2	11	10	5.1	100	0
41	1	1	75	1	56	12	11.3	18	0
42	23	1	56	2	20	9	14.6	3	0
43	15	1	62	2	21	10	8.8	5	0
44	18	1	60	2	18	9	7.5	85	1
45	12	0	71	2	46	9	4.9	62	0
46	12	1	60	2	6	10	5.5	25	0
47	17	1	65	2	28	8	7.5	8	0
48	3	0	59	1	90	10	10.2	6	1

Table 4.8: Estimates of the Parameters and Standard Errors on Fitting Koziol-Green Model with Weibull Lifetimes to the Multiple Myeloma Data.

Parameter	MLE		$Bayes_a$		$Bayes_b$	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
θ	2.6054	0.4279	2.1097708	0.3188656	2.203005	0.3023263
β	1.0208	0.1130	0.9071306	0.0706174	0.9629012	0.054311
γ	0.3333	0.1111	0.2937878	0.094342	0.300202	0.0902365

of the survivor function and compare them with the PLE of the survivor function. Note that, the product limit estimator is the empirical survivor function for censored data. For the Bayesian analysis we consider two sets of prior distributions of θ, β and γ. They are as follows:

Set(a): The prior distribution of β is assumed to be truncated normal with mean 0, standard deviation $\sigma = 0.55$, $c_0 = 0.2$ and $d_0 = 2$. Here σ, c_0 and d_0 are chosen so that the expected value of β is close to the MLE of β and the prior variance of β is close to the estimate of the variance of MLE of β. For the prior distribution of θ, we assume $(\theta^{\beta}|\beta)$ to be IG($a = 12$, $b = 0.02$). The parameters a and b are chosen so that the prior mean and variance of $(\theta \mid \beta)$ exist and the expected value of θ is close to the MLE of θ. We assume the prior distribution of the censoring parameter γ to be truncated exponential with $\eta = 1.12$ and $K = 0.77$ so that the expected value of γ is close to the MLE of γ and the prior variance of γ is close to the estimate of the variance of MLE of γ.

Set(b): The prior distribution of β is considered to be truncated gamma with shape parameter $g_a = 2$, scale parameter $g_b = 9.2$, $c_0 = 0.9$ and $d_0 = 3$. For the prior distribution of θ we assume $(\theta^{\beta}|\beta)$ to be IG($a = 10$, $b = 1$) and the prior distribution of γ is assumed to be Uniform(0, 0.66). The parameters of the prior distributions are fixed in the same manner as the parameters in Set(a) are fixed.

In Tables 4.8 and 4.9. $Bayes_a$ corresponds to the Bayes estimate with Set(a)

of prior distributions, $Bayes_b$ corresponds to the Bayes estimate with Set(b) of prior distributions and S.E. stands for the standard error of the corresponding estimators. Table 4.8 gives the MLE, the Bayes estimates and corresponding standard errors of the MLE and Bayes estimators of θ, β and γ. Table 4.9 gives the PLE, MLE. $Bayes_a$, $Bayes_b$ and standard errors of the corresponding estimators of the survival probabilities of the multiple myeloma patients at 23 distinct death times. Figures 4.15 and 4.16 give the plots of the PLE, MLE and the Bayes estimate of the survivor function with Set(a) and Set(b) of prior distributions, respectively. All the estimates are obtained by converting the time in months to time in years i.e., the observations in column two of Table 4.7 are divided by 12 and the observations in column one of Table 4.9 represent the death times in years.

We see from Table 4.9 that the PLE is closer to Bayes estimates of the survival probabilities with both Set(a) and Set(b) of prior distributions than the MLE in the beginning, i.e., till two years of survival time. After which the PLE is closer to the MLE than the Bayes estimates with both Set(a) and Set(b) of prior distributions. Though, towards the very end the PLE is closer to Bayes estimates with both the sets of prior distributions than MLE. The PLE is closer to Bayes estimates with Set(b) of prior distributions than the Bayes estimates with Set(a) of prior distributions and this can be ascertained from Table 4.9, Figure 4.15 and Figure 4.16.

Table 4.9: Estimates of the Survival Probabilities and Standard Errors on Fitting Koziol-Green Model with Weibull Lifetimes to the Multiple Myeloma Data.

Time	PLE	MLE	$Bayes_a$	$Bayes_b$	S.E. PLE	S.E. MLE	S.E. $Bayes_a$	S.E. $Bayes_b$
0.083	0.9375	0.9707	0.9460	0.9568	0.0349	0.0540	0.01304	0.0087
0.33	0.8949	0.8846	0.8254	0.8469	0.0445	0.0617	0.02840	0.0229
0.42	0.8097	0.8573	0.7910	0.8141	0.0571	0.0607	0.0314	0.0263
0.5	0.767	0.8307	0.7586	0.7828	0.0616	0.0558	0.0340	0.0292
0.67	0.7451	0.7798	0.6992	0.7242	0.0636	0.0489	0.0379	0.0340
0.83	0.6575	0.7317	0.6457	0.6706	0.0696	0.0413	0.0407	0.0378
1	0.634	0.6864	0.5972	0.6214	0.071	0.0338	0.04290	0.0408
1.08	0.6096	0.6648	0.5746	0.5983	0.0723	0.0270	0.04377	0.0420
1.17	0.5852	0.6438	0.5531	0.5761	0.0734	0.0212	0.0445	0.0431
1.25	0.5608	0.6234	0.5325	0.555	0.0743	0.0163	0.0452	0.0441
1.33	0.5098	0.6037	0.5127	0.5344	0.0758	0.0124	0.04576	0.0449
1.42	0.4844	0.5846	0.4939	0.5147	0.0762	0.0093	0.0463	0.0456
1.5	0.4334	0.5660	0.4758	0.4959	0.0762	0.0069	0.0467	0.0462
1.92	0.4045	0.4814	0.3960	0.4120	0.0764	0.0051	0.0479	0.0480
2	0.3756	0.4661	0.3820	0.3971	0.0762	0.0037	0.0479	0.0482
3	0.3467	0.3151	0.2503	0.2567	0.0756	0.0027	0.0463	0.0461
3.33	0.2889	0.2764	0.2183	0.2224	0.0732	0.0020	0.0449	0.0444
4.17	0.2568	0.1989	0.1562	0.1561	0.0718	0.0014	0.0406	0.0391
4.25	0.2247	0.1925	0.1511	0.1507	0.0696	0.0010	0.0402	0.0386
5.42	0.1798	0.1211	0.0962	0.0926	0.0687	0.00072	0.0333	0.0306
5.5	0.1348	0.1172	0.0932	0.0895	0.0646	0.0005	0.0328	0.0301
7.33	0.0674	0.0564	0.0474	0.0425	0.0576	0.0004	0.0228	0.0195
7.58	0	0.0510	0.0434	0.0385	0	0.0003	0.0217	0.0183

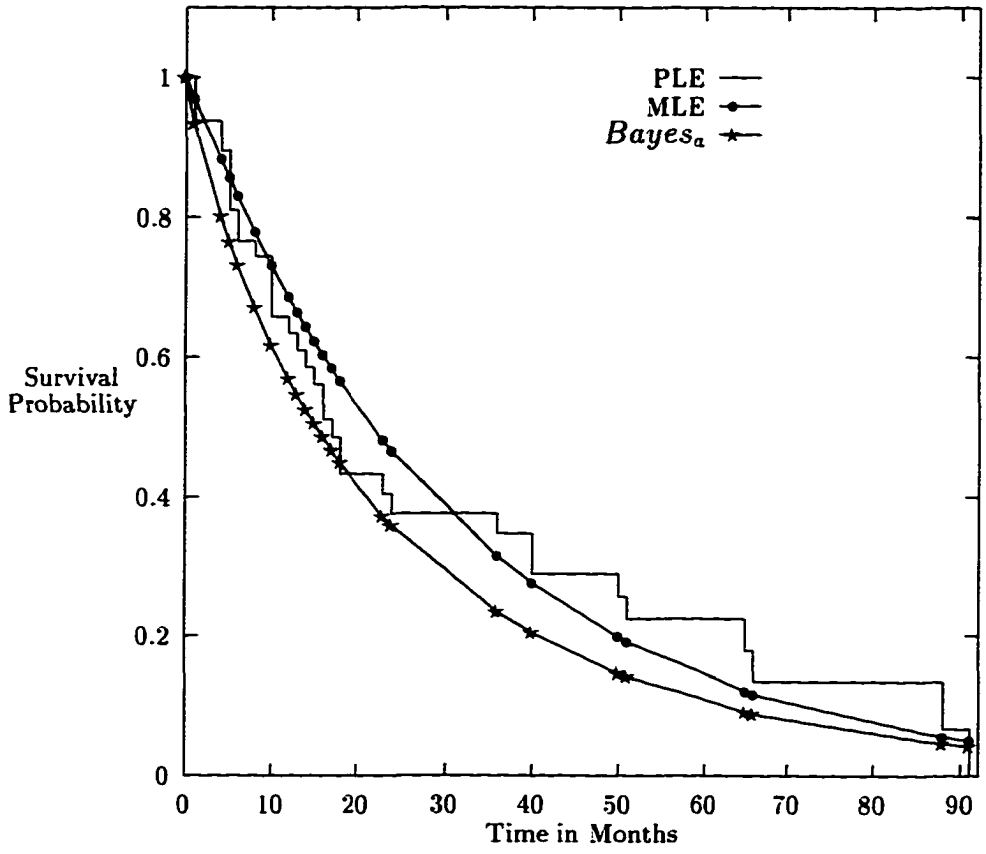

Figure 4.15: Plots of PLE, MLE and Bayes Estimate of Survival Probabilities with Set(a) of Prior Distributions.

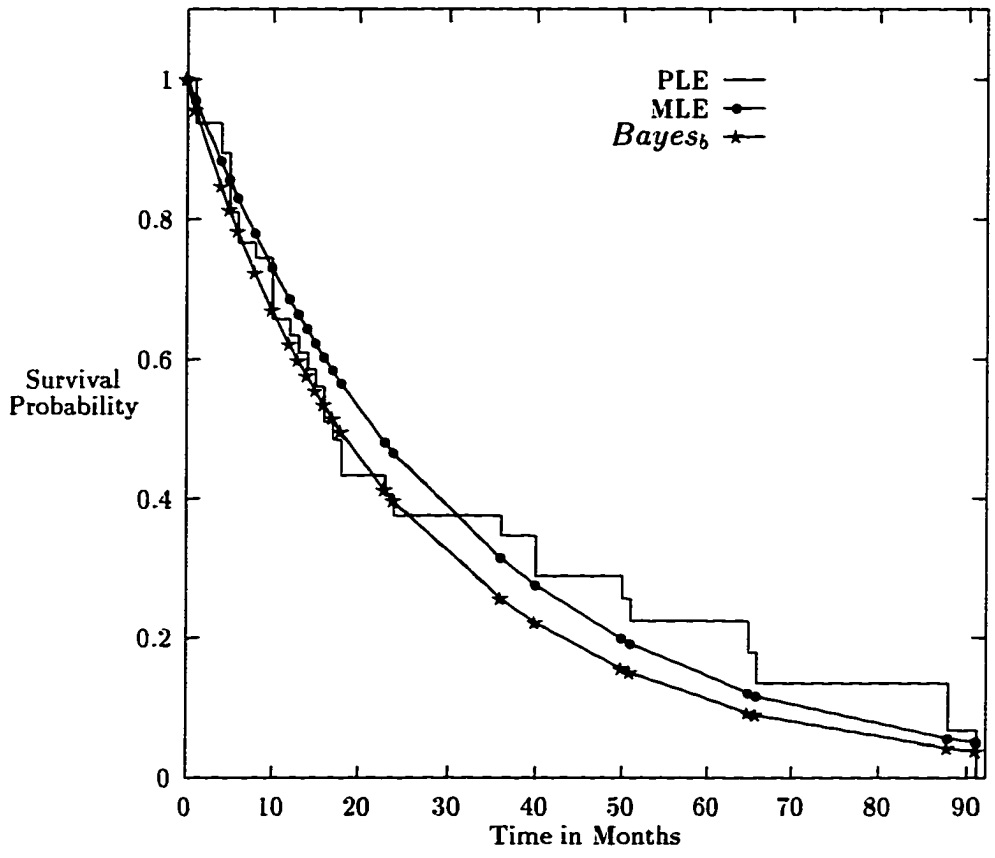

Figure 4.16: Plots of PLE, MLE and Bayes Estimate of Survival Probabilities with Set(b) of Prior Distributions.

Chapter 5

The Koziol-Green Survival Model with Covariates

5.1 Introduction

In this chapter, we consider the Koziol-Green survival model by incorporating covariates in the model and study their effect on the lifetimes.

Let $T_1, T_2, \ldots, T_n$ be independent and identically distributed random variables denoting failure times or lifetimes with a continuous distribution function say F and $C_1, C_2, \ldots, C_n$ be the corresponding independent and identically distributed random variables denoting censoring times with a continuous distribution function say G. Let

$$Z_i = min(T_i, C_i) \text{ and } \delta_i = I(T_i \leq C_i) \tag{5.1}$$

and (Z_i, δ_i), $i = 1, \ldots, n$ be the observed data. Let $S_T(t) = P(T > t)$ be the survivor function of the lifetimes and $S_C(t) = P(C > t)$ be the survivor function of the censoring times

Suppose there are p explanatory variables or covariates $X_1, X_2, \ldots, X_p$ associated with the lifetimes and x_{ij} denotes the value of the jth covariate recorded for the ith individual. Let the survivor function for the ith individual be given by

$$S_{T_i}(t) = [S_{C_i}(t)]^{\gamma_i} \tag{5.2}$$

where γ_i is the censoring parameter associated with the ith individual and it is given by

$$\gamma_i = exp(\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}) \tag{5.3}$$

for $i = 1, \ldots, n$. If $x_{ij} = 0$ for all j then the model reduces to the usual Koziol-Green survival model without covariates considered in Chapter 4.

Let $C_1, C_2, \ldots, C_n$ follow a Weibull distribution with pdf given by

$$f_C(t) = \frac{\alpha}{\theta^\alpha}\, t^{\alpha-1} exp(-(\frac{t}{\theta})^\alpha) \tag{5.4}$$

and the corresponding survivor function is given by

$$S_C(t) = exp(-(\frac{t}{\theta})^\alpha) \tag{5.5}$$

where $\theta > 0$ is the scale, $\alpha > 0$ is the shape parameter and $t > 0$.

Hence, under the model assumptions (5.1) and (5.2) for randomly censored data, the survivor function, hazard function and probability density function of life-times $T_1, T_2, \ldots, T_n$ are

$$S_{T_i}(t) = exp(-(\frac{t}{\theta})^\alpha)^{\gamma_i} = exp(-\gamma_i\,(\frac{t}{\theta})^\alpha), \tag{5.6}$$

$$h_{T_i}(t) = \frac{\gamma_i\, \alpha\, t^{\alpha-1}}{\theta^\alpha} \tag{5.7}$$

and

$$f_{T_i}(t) = \frac{\gamma_i\, \alpha}{\theta^\alpha}\, t^{\alpha-1} exp(-\gamma_i\,(\frac{t}{\theta})^\alpha), \tag{5.8}$$

respectively.

In Section 5.2, we consider the maximum likelihood estimation of the parameters of the model described by (5.1) and (5.2). Section 5.2.1, gives the asymptotic variances and covariances of the estimators. In Section 5.3, we briefly discuss covariate selection procedures and in Section 5.4, we consider a real life example.

5.2 Maximum Likelihood Estimation

The likelihood function of $(\theta, \alpha, \beta_0, \beta_1, \ldots, \beta_p)$ under the model assumptions (5.1), (5.2), (5.3) and (5.4) is given by

$$\begin{aligned} L(\theta, \alpha, \beta_0, \beta_1, \ldots, \beta_p \mid \underline{z}, \underline{\delta}) &= \prod_{i=1}^{n} [S_C(z_i)\, f_T(z_i)]^{\delta_i}\, [S_T(z_i)\, f_C(z_i)]^{1-\delta_i} \\ &= \frac{\alpha^n}{\theta^{n\alpha}} \prod_{i=1}^{n} z_i^{\alpha-1} \prod_{i=1}^{n} \gamma_i^{\delta_i} exp(-\sum_{i=1}^{n} \frac{(1+\gamma_i)}{\theta^\alpha} z_i^\alpha) \quad (5.9) \end{aligned}$$

where $\underline{z} = (z_1, z_2, \ldots, z_n)$, $\underline{\delta} = (\delta_1, \delta_2, \ldots, \delta_n)$ and γ_i is given by (5.3).
The natural logarithm of the likelihood in (5.9) is

$$\ln L = n \ln \alpha - n\alpha \ln \theta + (\alpha - 1) \sum_{i=1}^{n} \ln z_i + \sum_{i=1}^{n} \delta_i \ln \gamma_i - \sum_{i=1}^{n} \frac{(1+\gamma_i)}{\theta^\alpha} z_i^\alpha \qquad (5.10)$$

where $L = L(\theta, \alpha, \beta_0, \beta_1, \ldots, \beta_p \mid \underline{z}, \underline{\delta})$.

To get the MLEs of θ, α, $\beta_0, \beta_1, \ldots, \beta_p$, we partially differentiate the log-likelihood (5.10) with respect to θ, α, $\beta_0, \beta_1, \ldots, \beta_p$ respectively and equate the partial derivatives to zero. Hence, the estimating equations are

$$\frac{\partial \ln L}{\partial \theta} = 0$$

$$\Rightarrow \theta^\alpha = \frac{\sum_{i=1}^{n} (1+\gamma_i) z_i^\alpha}{n}, \qquad (5.11)$$

$$\frac{\partial \ln L}{\partial \alpha} = 0$$

$$\Rightarrow \frac{n}{\alpha} - n\ln\theta + \sum_{i=1}^{n}\ln z_i - \sum_{i=1}^{n}(1+\gamma_i)(\ln z_i - \ln\theta)(\frac{z_i}{\theta})^{\alpha} = 0. \qquad (5.12)$$

$$\frac{\partial \ln L}{\partial \beta_0} = 0$$

$$\Rightarrow \sum_{i=1}^{n}\delta_i - \frac{\sum_{i=1}^{n}\gamma_i z_i^{\alpha}}{\theta^{\alpha}} = 0 \qquad (5.13)$$

and

$$\frac{\partial \ln L}{\partial \beta_j} = 0, \quad j = 1,\ldots,p$$

$$\Rightarrow \sum_{i=1}^{n}\delta_i\, x_{ij} - \frac{\sum_{i=1}^{n}\gamma_i\, x_{ij}\, z_i^{\alpha}}{\theta^{\alpha}} = 0. \qquad (5.14)$$

Substituting for θ^{α} given by (5.11) in (5.12), (5.13) and (5.14), we get (5.15), (5.16) and (5.17), respectively.

$$(n+\alpha\sum_{i=1}^{n}\ln z_i)\sum_{i=1}^{n}(1+\gamma_i)z_i^{\alpha} - n\alpha\sum_{i=1}^{n}(1+\gamma_i)z_i^{\alpha}\ln z_i = 0, \qquad (5.15)$$

$$(\sum_{i=1}^{n}\delta_i)(\sum_{i=1}^{n}(1+\gamma_i)z_i^{\alpha}) - n\sum_{i=1}^{n}\gamma_i z_i^{\alpha} = 0 \qquad (5.16)$$

and

$$(\sum_{i=1}^{n}\delta_i\, x_{ij})(\sum_{i=1}^{n}(1+\gamma_i)z_i^{\alpha}) - n\sum_{i=1}^{n}\gamma_i\, x_{ij}\, z_i^{\alpha} = 0. \qquad (5.17)$$

We simultaneously solve the equations (5.15), (5.16) and (5.17), using Newton-Raphson method, to obtain the MLEs of α, β_0 and β_j for $j = 1,\ldots,p$. The MLE of θ can be obtained by substituting the MLEs of α, β_0 and β_j, $j = 1,\ldots,p$ in (5.11).

5.2.1 Variances and Covariances of Estimators

The asymptotic variance-covariance matrix of the MLEs of the parameters θ, α, β_0 and β_j for $j = 1,\ldots,p$ is obtained by inverting the information matrix whose elements are the negative expected values of the second order derivatives of the log-likelihood given by (5.10). For sufficiently large samples we can estimate the expected values by their MLEs.

Let $\hat{\phi} = (\hat{\theta},\ \hat{\alpha},\ \hat{\beta}_0,\ \hat{\beta}_j,\ j = 1,\ldots,p)$ be the MLE of $\phi = (\theta, \alpha, \beta_0, \beta_j,\ j = 1,\ldots,p)$. The estimate of the asymptotic variance-covariance is given by the inverse of the observed information matrix of order $(p+3) \times (p+3)$. The elements of the observed information matrix are the following:

$$-\frac{\partial^2 \ln L}{\partial \theta^2}\bigg|_{\hat{\phi}} = -\frac{n\alpha}{\theta^2} + \frac{\alpha(\alpha+1)}{\theta^{\alpha+2}} \sum_{i=1}^{n} (1+\gamma_i)\, z_i^{\alpha}\bigg|_{\hat{\phi}}, \tag{5.18}$$

$$-\frac{\partial^2 \ln L}{\partial \theta \partial \alpha}\bigg|_{\hat{\phi}} = \frac{n}{\theta} - \frac{\alpha}{\theta} \sum_{i=1}^{n} (1+\gamma_i) \left(\frac{z_i}{\theta}\right)^{\alpha} \ln \frac{z_i}{\theta} - \frac{1}{\theta} \sum_{i=1}^{n} (1+\gamma_i) \left(\frac{z_i}{\theta}\right)^{\alpha}\bigg|_{\hat{\phi}}, \tag{5.19}$$

$$-\frac{\partial^2 \ln L}{\partial \theta \partial \beta_0}\bigg|_{\hat{\phi}} = -\frac{\alpha}{\theta^{\alpha+1}} \sum_{i=1}^{n} \gamma_i\, z_i^{\alpha}\bigg|_{\hat{\phi}}, \tag{5.20}$$

$$-\frac{\partial^2 \ln L}{\partial \theta \partial \beta_j}\bigg|_{\hat{\phi}} = -\frac{\alpha}{\theta^{\alpha+1}} \sum_{i=1}^{n} \gamma_i\, x_{ij}\, z_i^{\alpha}\bigg|_{\hat{\phi}},\ j = 1,\ldots,p, \tag{5.21}$$

$$-\frac{\partial^2 \ln L}{\partial \alpha^2}\bigg|_{\hat{\phi}} = \frac{n}{\alpha^2} + \sum_{i=1}^{n} (1+\gamma_i) (\ln z_i - \ln \theta)^2 \left(\frac{z_i}{\theta}\right)^{\alpha}\bigg|_{\hat{\phi}}, \tag{5.22}$$

$$-\frac{\partial^2 \ln L}{\partial \alpha \partial \beta_0}\bigg|_{\hat{\phi}} = \sum_{i=1}^{n} \gamma_i (\ln z_i - \ln \theta) \left(\frac{z_i}{\theta}\right)^{\alpha}\bigg|_{\hat{\phi}}, \tag{5.23}$$

$$-\frac{\partial^2 \ln L}{\partial \alpha \partial \beta_j}\bigg|_{\hat{\phi}} = \sum_{i=1}^{n} \gamma_i\, x_{ij} (\ln z_i - \ln \theta) \left(\frac{z_i}{\theta}\right)^{\alpha}\bigg|_{\hat{\phi}},\ j = 1,\ldots,p, \tag{5.24}$$

$$-\frac{\partial^2 \ln L}{\partial \beta_0^2}\bigg|_{\hat{\phi}} = \sum_{i=1}^{n} \gamma_i \left(\frac{z_i}{\theta}\right)^{\alpha}\bigg|_{\hat{\phi}}, \tag{5.25}$$

$$-\frac{\partial^2 \ln L}{\partial \beta_0 \partial \beta_j}\bigg|_{\hat{\phi}} = \sum_{i=1}^{n} \gamma_i\, x_{ij} \left(\frac{z_i}{\theta}\right)^{\alpha}\bigg|_{\hat{\phi}},\ j = 1,\ldots,p, \tag{5.26}$$

$$-\frac{\partial^2 \ln L}{\partial \beta_j \partial \beta_s}\bigg|_{\hat{\phi}} = \sum_{i=1}^{n} \gamma_i\, x_{ij}\, x_{is} \left(\frac{z_i}{\theta}\right)^{\alpha}\bigg|_{\hat{\phi}},\ j,s = 1,\ldots,p. \tag{5.27}$$

5.3 Variable Selection Procedures

In this section, we describe the procedure used to select important covariates which have a considerable influence on the survival or the hazard rates from a list of potential covariates. Variable selection procedures have been discussed in detail by Collett (1994, Chapter 3).

In order to compare alternative models fitted to an observed set of survival data, we need a statistic which measures the extent to which the data are fitted by a particular model. Since the likelihood function summarizes the information that the data contains about the unknown parameters in a given model, a suitable summary statistic is the value of the likelihood function when the parameters are replaced by their MLEs. For a given data set, a large value of the maximized likelihood indicates an agreement between the model and observed data. If we denote the maximized likelihood for a given model by $\hat{L}$, then $-2\log\hat{L}$ is considered as the summary measure of agreement between the model and the data. The value of $-2\log\hat{L}$ will always be positive for a given set of data and a smaller value of $-2\log\hat{L}$ indicates that the model fits the data well. The value of $-2\log\hat{L}$ is only useful when comparing models fitted to the same data.

Let us consider two models, Model(1) and Model(2) say. Suppose that p covariates $X_1, X_2, \ldots, X_p$ are fitted in Model(1), then the survivor function of Model(1) can be written as

$$S_T(t) = [S_C(t)]^{exp(\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p)} \tag{5.28}$$

Suppose Model(2) is fitted with additional covariates $X_{p+1}, X_{p+2}, \ldots, X_{p+q}$, then the survivor function of Model(2) can be written as

$$S_T(t) = [S_C(t)]^{exp(\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \beta_{p+1} X_{p+1} + \cdots + \beta_{p+q} X_{p+q})} \tag{5.29}$$

Let the value of the maximized log-likelihood function for each model be denoted by $\hat{L}(1)$ and $\hat{L}(2)$ respectively. Then the statistic given by

$$\hat{\lambda} = -2 \log \frac{\hat{L}(1)}{\hat{L}(2)}, \tag{5.30}$$

is the log-likelihood ratio statistic for testing the null hypothesis that the q parameters $\beta_{p+1}, \ldots, \beta_{p+q}$ in Model(2) are all zero. The statistic $\hat{\lambda}$ has an asymptotic chi-square distribution under the null hypothesis that the coefficients of the additional variables are zero. The degrees of freedom of this chi-square distribution is equal to the difference between the number of parameters being fitted under the two models. Here $\hat{\lambda}$ has a chi-square distribution with q degrees of freedom under the null hypothesis that $\beta_{p+1}, \ldots, \beta_{p+q}$ are all zero.

If the observed value of $\hat{\lambda}$ is not significantly large then we conclude that the additional covariates $X_{p+1}, X_{p+2}, \ldots, X_{p+q}$ do not significantly affect the survival or hazard rates and Model(1) would be preferred to Model(2). If the value of $\hat{\lambda}$ is significant then we deem it necessary to include the additional covariates in the model. Hence the general strategy for model selection is as follows:

(1) The first step would be to identify a set of explanatory variables that have the potential for being included in the model.

(2) The second step is to fit models that contain each of the covariates one at a time. The values $-2 \log \hat{L}$ for these models are then compared with that of the null model (without any covariates) to determine which covariates on their own significantly reduce the value of this statistic.

(3) The covariates which are determined as significant in step 2, are then fitted together. Those covariates which do not significantly increase the value of $-2 \log \hat{L}$. when they are omitted from the model can now be discarded. Only those that lead to a significant increase in the value of $-2 \log \hat{L}$ are retained in the model.

(4) Covariates which were not important on their own and were not under consideration in step 3, may become important in the presence of others. These variables are therefore added to the model from step 3 one at a time. The covariates that reduce $-2 \log \hat{L}$ significantly are retained in the model.

(5) Finally, a check is made to ensure that no term in the model can be omitted with-

out significantly increasing the value of $-2\log\hat{L}$ and no term that is not included in the model significantly reduces $-2\log\hat{L}$.

Some times it may be necessary to include interaction and other higher order terms such as powers of certain covariates. Such terms would be added to the model identified in step 4 above. Also there may be some variables which may not appear to be statistically significant in the modeling of a given data set but they may be important on medical grounds or other non-statistical considerations, so, it would be sensible to include these variables also in the model.

5.4 An Example

In this section, we analyze the data on the survival times of 48 patients suffering from multiple myeloma give in Table 4.7. This data set was also considered in Chapter 4 without covariates. Here our aim is to study the association between the values of certain explanatory variables or covariates and the survival times of the patients under our model assumptions given by (5.1), (5.2), (5.3) and (5.4). In Chapter 4, we have mentioned that the data in Table 4.7, is a good fit for the Koziol-Green model with the lifetime distribution being Weibull. Hence, the data is appropriate for the model considered in this chapter.

The multiple myeloma data in Table 4.7 contains values of seven covariates which were recorded for each patient along with their survival times and survival status. The covariates are as follows:

AGE: age of the patient

SEX: sex of the patient (0 =male and 1 =female)

BUN: blood urea nitrogen

CA: serum calcium

HB: serum hemoglobin

PC: percentages of plasma cells

BJ: Bence-Jones protein (0 =absent and 1 =present)

The sex of the patient and the variable associated with Bence-Jones protein are factors with two levels. These terms are fitted using the indicator variable SEX and BJ. It is not necessary to include all the seven covariates in our model. We have to determine the most appropriate subset of these variables, in doing so, we adopt the stepwise procedure discussed in Section 5.3.

The first step is to fit the null model i.e., the model without any covariates and then we fit models with each of the seven covariates one at a time. We consider the natural logarithm of the variable BUN, since the values of BUN range from 6 to 172 and the distribution of the values across the 48 individuals is positively skewed. This would prevent the extreme values from influencing the estimate of the coefficient of the variable BUN. Let LBUN denotes the natural logarithm of BUN.

A summary of the values of $-2\log\hat{L}$ for all the models considered is given in Table 5.1. We can see that LBUN and HB lead to significant reduction in $-2\log\hat{L}$. The reduction in $-2\log\hat{L}$ on adding LBUN to the null model is 3.7919 which is significant at 10% level when compared with the percentage point of a chi-square distribution with 1 degree of freedom. The reduction in $-2\log\hat{L}$ on adding HB to the null model is 4.9974 which is significant at 5% level when compared with the percentage point of a chi-square distribution with 1 degree of freedom. The only other variable which seems to have some explanatory power on its own is BJ. The other covariates do not bring any significant reduction in $-2\log\hat{L}$.

The next step would be to fit the model that contains LBUN, HB and BJ. When we fit all the three variables in the model, the value of $-2\log\hat{L}$ is 202.9933. If LBUN is omitted from this model, the increase in $-2\log\hat{L}$ is 4.4816 which is significant at 5% level. If HB is omitted from the model the increase in $-2\log\hat{L}$ is 1.9843 which is not significant even at 10% level. Similarly, omitting BJ from the model leads to an increase of 3.00968 in $-2\log\hat{L}$ which is significant at 10% level. Even though HB is not significant in the presence of LBUN and BJ, we decide to keep it in the model based on medical grounds. We also decide to keep BJ in the model since it is significant at 10% level in presence of LBUN and HB.

Finally, we examine if any of the covariates AGE, SEX, CA and PC could be

Table 5.1: Values of $-2\log\hat{L}$ for Models Fitted to the Multiple Myeloma Data.

Variables in model	$-2\log\hat{L}$
None	213.9591
AGE	213.9264
SEX	213.3509
LBUN	210.1671
CA	213.3310
HB	208.9616
PC	213.8323
BJ	211.3106
LBUN + HB	206.0030
LBUN + BJ	204.9777
HB + BJ	207.4749
LBUN + HB + BJ	202.9933
LBUN + HB + BJ + AGE	202.2227
LBUN + HB + BJ + SEX	202.6607
LBUN + HB + BJ + CA	202.9840
LBUN + HB + BJ + PC	202.9806

Table 5.2: Parameter Estimates and their Standard Errors on Fitting the Model Given by (5.31) to the Multiple Myeloma Data.

Parameter	Estimate	S.E.
θ	7.0436	1.9028
α	1.1248	0.1234
β_0	0.1517	1.2994
β_1	0.6840	0.3204
β_2	-0.0865	0.0609
β_3	-0.6687	0.3942

included in the model that contains LBUN, HB and BJ. We can see from Table 5.1 that when any of these four variables are added to the model containing LBUN, HB and BJ, the reduction in $-2\log\hat{L}$ is less than 0.8. Hence, we do not need to include these variables in the model and we consider the most satisfactory model to be that containing LBUN, HB and BJ. The survivor function under the model containing LBUN, HB and BJ for the ith individual is given by

$$S_{T_i}(t) = [S_{C_i}(t)]^{exp(\beta_0 + \beta_1 LBUN_i + \beta_2 HB_i + \beta_3 BJ_i)} \qquad (5.31)$$

for $i = 1, \ldots, n$ where $LBUN_i$, HB_i and BJ_i are the values of the covariates LBUN, HB and BJ, respectively for the ith individual. Table 5.2 gives the MLEs of the parameters including the estimates of the coefficients of the covariates and also their standard errors (S.E.) for the model containing LBUN, HB and BJ.

The Koziol-Green model with covariates would be an appropriate model to study the association of a set of covariates with lifetimes for randomly censored data when the lifetime survivor function is a power of the censored time survivor function. This model would be especially useful when we have a large percentage of censored observations in the data set.

Chapter 6

Summary

In this dissertation we have concentrated on the analysis of mark-recapture and survival data. In Chapters 2 and 3, we considered the analysis of mark-recapture data. We have introduced a model based approach to the estimation of the exploitation rate of a fish population by combining mark-recapture procedures with a creel survey. Our model is simple, useful and it is for a closed population. Also it does not rely on voluntary tag returns or rewards.

We obtained the maximum likelihood estimator and the moment estimator of the exploitation rate. A simulation study was performed to compare the two estimators and it was found that the moment estimator performed slightly better than the maximum likelihood estimator when we sampled a large number of units and the probability of a marked fish getting captured was reasonably high, otherwise both the estimators performed equally well. One can use the moment estimator to obtain point and interval estimates of the exploitation rate, since it is easier to derive the distributional properties of the moment estimator. The model based procedures of Chapter 2 were extended to the case where we stratify the space-time units of the fisheries according to fishing areas or seasons. We obtained the maximum likelihood estimator and two moment estimators of the exploitation rate and we compared the three estimators through a simulation study. We found that the second moment estimator $\hat{u}_2$ performed better that the other two estimators, hence we concentrated

on $\hat{u}_2$ to construct a confidence interval of the exploitation rate.

In Chapters 4 and 5, we considered the analysis of a randomly censored survival data. We considered a parametric survival model where the lifetime survivor function is a power of the censored time survivor function with the lifetimes following a Weibull distribution. We considered the Bayesian analysis of this model by specifying parametric prior distributions of the parameters of the model. We showed the implementation of the Gibbs sampler and obtained the Bayes estimates of the parameters of the model. We also compared the maximum likelihood estimator with the Bayes estimator of the survivor function. We found that the Bayes estimator performed better than the maximum likelihood estimator at the tails of the distribution of the data which seems to be in concert with the fact that there is less information at the tails. In Chapter 5 we incorporated covariates in the Koziol-Green survival model to study the effect of covariates on the lifetimes. This proportional hazards model would be appropriate to study the association of covariates with lifetimes when the data is randomly censored and it shows evidence that the lifetime survivor function is a power of the censored time survivor function. This model would be particularly useful when we have a large percentage of censored observations in the data set. We used the model to analyze a real life data set related to 48 patients suffering from multiple myeloma. We considered maximum likelihood estimation of the parameters of the model and adopted a stepwise procedure to select the most important subset of covariates from the available set of covariates.

Bibliography

Allen, W. R. (1963). A note on conditional probability of failures when hazards are proportional. *Operations Research*, **11**, 658-659.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. New York: Springer-Verlag.

Berger, J. O. and Sun, D. (1993). Bayesian analysis for the Poly-Weibull distribution. *Journal of the American Statistical Association*, **88**, 1412-1418.

Box, G. E. P. and Tiao, G. (1973). *Bayesian Inference in Statistical Analysis*. London: Addison-Wesley.

Brownie, C., Anderson, D. R., Burnham, K. P., and Robson, D. S. (1985). *Statistical Inference From Band-Recovery Data: A Handbook*, 2nd ed, Resource Publication 156. Washington, DC: Fish and Wildlife Service, U.S. Department of Interior.

Casella, G. and George, E. I. (1992). Explaining the Gibbs Sampler. *The American Statistician*, **46**, 167-174.

Chen, C. H. (1984). A correlation goodness of fit test for randomly censored data. *Biometrika*, **71 (2)**, 315-322.

Chen, Y. Y., Hollander, M., and Langberg, N. A. (1982). Small-sample results for the Kaplan-Meier estimator. *Journal of the American Statistical Association*. **77**. 141-144.

Cochran, W. G. (1977). *Sampling Techniques*. New York: J.Wiley and Sons.

Cohen, A. C. (1965). Maximum likelihood estimation in Weibull distribution based on complete and on censored samples. *Technometrics*, **7**, 579-588.

Collet, D. (1994). *Modelling Survival Data in Medical Research*. London: Chapman & Hall.

Conroy, M. J. and Blandin, W. W. (1984). Geographic and temporal differences in band reporting rates for American black ducks. *Journal of Wildlife Management*. **48**, 23-27.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society,* **B, 74**, 187-220.

Dahiya, R. C. (1981). An improved method of estimating an integer-parameter by maximum likelihood. *The American Statistician,* **35,** 34-37.

Devroye, L. (1986). *Nonuniform Random Variate Generation.* New York: Springer-Verlag.

Efron, B. (1967). The two sample problem with censored data. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability,* **4**, 831-852.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association,* **85.** 398-409.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **6,** 721-740.

Henny C. J. and Burnham, K. P. (1976). A Mallard reward band study to estimate band reporting rates. *Journal of Wildlife Management,* **40,** 1-14.

Jagielo, T. (1991). Synthesis of mark-recapture and fishery data to estimate open population parameters. *American Fisheries Society Symposium,* **12**, 492-506.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions.* Volume 1, 2nd ed. New York: J.Wiley and Sons.

Koziol, J. A. and Green, S. B. (1976). A Cramer-von Mises statistic for randomly censored data. *Biometrika,* **63,** 465-474.

Lee, P. M. (1989). *Bayesian Statistics: An Introduction.* New York: Oxford University Press.

Lemon, G. H. (1975). Maximum likelihood estimation for the three parameter Weibull distribution based on censored samples. *Technometrics,* **17**, 247-254.

Nelson, L. J., Anderson, D. R., and Burnham, K. P. (1980). The effect of band loss on estimates of annual survival. *Journal of Field Ornithology*, **51**, 30-38.

Nichols, J. D., Stokes, S. L., Hines, J. E., and Conroy, J. J. (1982). Additional comments on the assumption of homogeneous survival rates in modern bird banding estimation models. *Journal of Wildlife Management*, **46**, 953-962.

Pollock, K. H., Hines, J. E., and Nichols, J. D. (1984). The use of auxiliary variables in capture-recapture and removal experiments. *Biometrics*, **40**, 329-340.

Pollock, K. H., Hoenig, J. M., and Jones, C. M. (1991). Estimation of fishing and natural mortality when a tagging study is combined with a creel survey or port sampling. *American Fisheries Society Symposium*, **12**, 423-434.

Pollock, K. H. and Raveling, D. G. (1982). Assumptions of modern band-recovery models, with emphasis on heterogeneous survival rates. *Journal of Wildlife Management*, **46**, 88-98.

Pollock, K. H., Winterstein, S. R., and Conroy, M. J. (1989). Estimation and analysis of of survival distributions for radio-tagged animals. *Biometrics*, **45**, 99-109.

Pierce, R. B., Tomcko, C. M., and Schupp, D. H. (1995). Exploitation of northern pike in seven small north-central Minnesota Lakes. *North American Journal of Fisheries Management*, **15**, 601-609.

Ritter, C. and Tanner, M. A. (1992). Facilitating the Gibbs sampler: The Gibbs stopper and the Griddy-Gibbs sampler. *Journal of the American Statistical Association*, **87**, 861-868.

Roberts, G. O. (1992). Convergence diagnostics for the Gibbs sampler. *Bayesian Statistics*, **4**, eds. J.M. Bernardo et al. London: Oxford University Press, 775-782.

Seber, G. A. F. (1982). *The Estimation of Animal Abundance and Related Parameters*. 2nd ed. London: Charles W. Griffin.

White, G. C. (1983). Numerical estimation of survival rates from band recovery and biotelemetry data. *Journal of Wildlife Management*, **47**, 716-728.

Vita

Name: Shampa Saha

Education: 1987 M.Sc. in Statistics, Bangalore University
Bangalore, India.

1985 B.Sc., Bangalore University
Bangalore, India.

Awards: 1987 Central college gold medal
awarded by Bangalore University
Bangalore, India.

1987 Professor Umarji Rao gold medal
awarded by Bangalore University
Bangalore, India.

Experience: 1991-1997 Graduate Teaching Assistant
Department of Mathematics and Statistics
Old Dominion University
Norfolk, VA.

1988-1991 Lecturer, Department of Statistics
M.E.S College of Arts, Commerce and Science
Bangalore, India.