Bayes' Estimators of Finite Population Size

ALFRED A. NNADOZIE

A Thesis

in

The Department

of

Mathematics and Statistics

Presented in Partial Fulfillment of Requirements

for the Degree of Master of Science at

Concordia University

Montreal, Quebec, Canada

March 1996

ISBN 0-612-18424-2

ABSTRACT

Bayes' Estimators of Finite Population Size

Alfred A. Nnadozie

The estimation of a finite population is tackled from the Bayesian point of view Truncated priors from the class of power series distributions are considered and the resulting Bayes' estimators compared based on their sensitivity to changes in prior distribution. Three numerical examples corresponding to situations when the prior mean is less than. equal to. or greater than the prior variance indicate that the Bayes' estimator resulting from the truncated poisson is preferable. A truncated "poisson difference" prior is proposed and extension of the study to situations where the observations can be viewed as time dependent data is suggested for future research.

ACKNOWLEDGEMENT

My deepest gratitude goes to my supervisor Professor Y. Chaubey for his strong guidance and inspiration during the course of this research. It was the result of his work with Professors M. Ahmad and S. Alalouf (both at Université du Québec à Montréal) that inspired this thesis. I very much appreciate his unflichingly invaluable direction that helped bring it to a conclusion. My sincere gratitude goes also to Professor M. Ahmad.

To God, "them all", and all persons of good will.

Contents

LIST of TABLES

Chapter 1 Introduction

1.1 Introduction

A well known and important problem in finite population inference is that of ascertaining the size of the population, N when it is unknown. This problem is encountered in many areas of application ranging from simple surveys to more general problems related to species counts in ecological and environmental studies especially in surveys of biological populations where some sampling techniques have gained wide audience within the context of 'capture-recapture' sampling theory which essentially prescribes obtaining information by the procedure of capturing-marking or tagging and then - releasing back into the population, some of its members. Among some remarkable treatise and applications of this theory to the problem of estimating ecological populations are the works of Petersen(1896), Lincoln(1930), Schnabel(1938), Chapman (1951,1954), Craig(1953), Darroch(1958), Seber(1965,1973), Cormack(1968), Robson (1969), Otis et al. (1978), etc. Also, Nayak(1988) used a recapture debugging design to get extra information in estimating the number of faults in a reliability system such as a computer software with an unknown number of errors. N Estimate of N is also required in estimating population total, which is an important quantity in some applications [see Ahmad et al. (1995a)].

In most cases, stochastic models are constructed so that N is estimated as a parameter in the probability distribution of some appropriate statistic(s). A sufficient statistic that has been found very useful in estimating N is the number of distinct units, D in a sample. Feller (1957) gives the distribution of D and Basu(1958) observed that in simple random sampling with replacement, the sample mean based on D is a sufficient statistic for the population mean. Harris(1968) considered estimation of N using the distribution of D. Recently, Ahmad et al. (1995) introduced two variants of this scheme involving random sample sizes. They remarked that some prior knowledge of how large the population may be, could be useful for reducing the variance of the estimators. In their work, absence of prior information is either presumed or implied. However, prior knowledge is frequently available and it is rarely the case that an experiment is planned without any prior information on the statistical properties of the interplaying variables. Prior knowledge could be available from past experience, technical considerations, theoretically objective reasoning or even from subjective imagination. Unfortunately, the classical methods of maximum likelihood and moment estimation do not provide the means to incorporate such prior knowledge. An approach which provides a general method of incorporating prior information with observed data is the Bayesian approach which we will adopt in this thesis. First we present the basic tool which forms the bedrock of Bayesian statistics called the Bayes theorem. This theorem contained in the post-humous paper, Bayes(1763), of the English clergyman Thomas Bayes who died in 1761, can be rephrased here as follows:

Theorem 1.1.1 Bayes' Theorem (Discrete Case)

Let $\{X_i\}$ be a sequence of mutually exclusive events and $X_1 \cup X_2 \cup \ldots = \aleph$. Let d be another event for which the sequence of conditional probabilities $P(X_i \mid d)$, $i = 1, 2, \ldots$ are defined. Then

$$P(X_i \mid d) = \frac{P(d \mid X_i)\,P(X_i)}{\sum_{i=1} P(d \mid X_i)\,P(X_i)} \quad \text{provided} \quad P(d) > 0 \tag{1.1}$$

The above is the Bayes' theorem for the discrete case with which we are concerned in this thesis. In the formula, $P(X)$ which tells us what is known about X without knowlege of d, is called the **prior probability** of X because it represents the distribution of one's degree of belief about X prior to observing the event d, i.e., prior to carrying out any experiment that might bear on the value of X. Correspondingly, $P(X \mid d)$ which tells us what is known about X given knowledge of d, is called the **posterior**

probability of X given d because it is a distribution determined subsequent to having observed the outcome of an experiment bearing on the value of X. Thus it can be said from a subjective probabilistic point of view that Bayes' theorem provides a medium for changing or updating the degree of belief about X in the light of more recent information coming from d. Actually Bayesian theory is a normative theory for learning from experience in the sense that it provides a formal procedure for merging knowledge obtained from experience, or theoretical understanding of what X can be, with observational data.

1.2 Outline of the thesis

The next chapter is a review of the essential elements of the theory used in the sequel. In chapter 3 we apply the results of chapter 2 to the problem of estimating X (here considered a random variable) using D and prior information (modelled as a prior probability distribution of X). The sensitivity of the resulting Bayes estimators to change in prior distribution is then investigated in chapter 4 where some computational results are used to illustrate the performance of the different estimators when the prior mean is less, equal or greater than the prior variance. In chapter 5, further types of prior distributions are proposed and introduced and some limit properties of the Bayesian estimators are considered. Finally, the thesis is concluded with a discussion.

Chapter 2 Essential Elements of Bayesian Estimation

2.1 Introduction

The development of Bayes' estimators(rules) can be approached in two ways: as a **traditional statistical inference** problem or as a **statistical decision problem**[see Berger (1980)]. The traditional approach consists of simply finding the posterior distribution and then obtaining estimators from the posterior in any way one thinks reasonable. No formal consideration is given to the loss that might be associated with an estimator. On the other hand, the Bayesian decision-theoretic approach makes use of a loss function and seeks to minimize the **Bayes risk** which is a measure of the average risk associated with a decision rule(estimator). Essentially, both approaches require sample information and prior information in order to arrive at a posterior distribution

2.2 Sample information and likelihood function

Generally, a statistical method for estimating Λ should be directed towards the use of sample information in making inferences about Λ considered as the unknown quantity or parameter involved in the probability distribution of an appropiate statistic. In what follows we will use the distribution of the sufficient statistic D. Ideally, the sample observation d will be postulated to be the value taken by the random variable, D which follows a probability distribution indexed by the quantity Λ, where Λ takes values in a set $\aleph$ so that $\Lambda \in \aleph$. Technically, we will speak in terms of a sample space, Ω of elements D (the sample), endowed with an appropiate Borel field (σ-field) of sets over which is given a family of probability measures indexed by the quantity Λ (the parameter) belonging to $\aleph$ (the parameter space). Also, it will be supposed that these probability measures are dominated by a σ-finite measure, $\mu(d)$ (which in this case is the counting measure) so that they may be described through their probability mass functions, $P(d|\Lambda)$, with respect to this measure $\mu(d)$.

The quantity in the denominator of (1.1) is the marginal distribution of D, $P(d)$ which can be written as

$$P(d) = \sum_{i=1}^{\infty} p(N_i)p(d|N_i) = C^{-1}.$$

This marginal distribution is merely a "normalizing" constant necessary to ensure that the posterior distribution $P(N|d)$ sums to unity. Thus (1.1) can be written as

$$P(N|d) = C\,P(d|N)P(N) \tag{2.1}$$

so that given the data, $P(d|N)$ in (2.1) may be regarded as a function not of d, but of N. When it is so regarded, in accordance with Fisher(1922), it is called the likelihood function of N for given $D = d$ and it can be written as $l(N|d)$.

In summary therefore, Bayes theorem tells us that the posterior distribution of $\boldsymbol{N}$ given $D = d$, is proportional to the product of the prior distribution of N and the likelihood function of N given $D = d$. That is

$$P(N|d) \propto P(N)l(N|d) \tag{2.2}$$

This highlights the important role of the likelihood function in the Bayes' formula. The likelihood function, $l(N|d)$ can be interpreted as the function through which the sample data, D modifies prior knowledge about N. Hence it can be understood as representing the information about N contained in the sample data. And because the likelihood function is defined up to a multiplicative constant (that is, multiplication by a constant does not change the likelihood), multiplication by an arbitrary constant has no effect on the posterior distribution of N. The constant cancels upon normalizing the product on the right hand side of (2.2). Therefore, the important thing is only the relative value of the likelihood This points to a quite general feature of Bayesian methodology [see Lehmann

(1983) pg. 243] which we record here as a remark for future reference:

Remark 2.2.1

The posterior distribution does not depend on the sampling method used but on the likelihood of the obser.ed results.

2.3 **Prior information modelling and posterior distribution**

Prior information is information about N coming from sources other than the statistical investigation and as such it may not necessarily be precise. Therefore a convenient way to quantify such information is in terms of a prior probability distribution $\pi(N)$ on the set $\aleph$. The essential idea of the Bayesian method is that of using Bayes' theorem to modify the prior distribution. which we shall hereafter denote by $\pi(N)$. in the light of the sample data. d to determine a posterior distribution. hereafter denoted by $\pi(N|d)$. (the conditional distribution of N given d) from which all decisions and inferences about N are made. That is, noting that N and D have a joint distribution

$$f(d,N) = \pi(N)\, p(d|N) \tag{2.3}$$

and that D has marginal density

$$m(d) = \sum \pi(N)\, p(d|N) \tag{2.4}$$

it is obvious that for $m(d) \neq 0$, the posterior distribution is given by

$$\pi(N|d) = \frac{f(d,N)}{m(d)} \tag{2.5}$$

Note that the range of N in $\pi(N|d)$ is the union of $\{N > d\}$ and any restriction implied by the prior, $\pi(N)$. In general, $m(d)$ and $\pi(N|d)$ may not be easily calculable. For this reason, a number of concepts and ways have been worked out to facilitate the calculation of $\pi(N|d)$. Among some of the methods and problems involved in the construction of such prior probability distributions are the following;

(i) Conjugate prior distributions

A large part of the Bayesian literature is devoted to finding families of prior distributions for which $\pi(N|d)$ can easily be calculated. Such types of priors are called conjugate priors.

Definition: *Let denote the class of probability functions $P(d|N)$ (indexed by N). A class Γ of prior distributions is said to be a conjugate family for if $\pi(N\ |d)$ is in the class for all $p \in$ and $\pi \in \Gamma$.*

Although it is not necessary to choose a prior from a particular family, there is the advantage of obtaining a compact-form expression for the Bayes' estimator(rule). For instance, starting with a beta prior for the binomial, one ends up with a beta posterior in such a way that the updating of the prior takes the form of updating its parameters.

(ii) Improper Prior Distributions

When a probability density function does not integrate or sum over its admissible range to unity, it is said to be an improper distribution. Thus, when $\pi(N)$ is an improper distribution, we call it an improper prior. The analysis leading to the posterior distribution can still be formally done even if $\pi(N)$ is an improper prior. Improper priors are frequently used in Bayesian analysis as sensibly practical approximations to proper prior distributions. This is done by supposing that to a sufficient approximation, the prior is improper only over the range of appreciable likelihood and that it suitably tails off to zero outside that range thereby ensuring that the priors actually used are proper.

(iii) Non-informative prior distributions

These are priors that supposedly reflect a complete lack of information about $\boldsymbol{N}$, and can hence be considered objective. The drawback to the use of such priors is that usually a clearcut non-informative prior does not exist. Moreover they are usually improper, thereby making the interpretation of the posterior distribution unclear.

(iv) Subjective and objective prior distributions

When a distribution is arbitrarily chosen for N apriori, it is called a subjective prior distribution otherwise it is called an objective prior. This concept is very important in situations when N may not objectively be considered to be a random variable. The essential idea of subjective probability [see DeGroot(1970)] is to let the probability of an event reflect the personal belief in the 'chance' of the occurence of the event thereby enabling one to talk about probabilities when the frequency viewpoint does not apply. The subjectivity(objectivity) of the prior distribution is a major source of the controversy over Bayesian analysis. In general, an objective prior, when available is the ideal since a subjective prior raises controversy as to the objectivity of the whole inference process. However, in this thesis, we shall not be as interested in this controversy as we shall be in the robustness of the prior which is the most serious [see Berger (1980) pg. 85] question concerning prior information. Thus we shall be looking for a plausibly convenient way of modelling prior information into prior distributions that are as non-informative as possible yet so robust that slight changes in the prior would not cause significant changes in the inference or decision procedure concerning N.

2.3.1 Methods of determining a prior

A number of methods for determining a prior distribution have been suggested in the literature [see for example Berger(1980) pg.63, Lehmann(1983) pg.241, Iversen(1984) pg. 64 etc]. In all the methods, the most common approach is that of matching a given functional form to prior information. The idea is essentially to assume that the prior is of a given functional form which most closely matches prior beliefs. Then prior parameters can easily be calculated from estimated prior moments. A drawback to this method is that the estimation of prior moments is often an extremely uncertain undertaking because the tails of a probability function can have effect on its moments. One way to circumvent such problem is to use non-informative prior distributions. In this thesis, we shall apply a method similar to that developed by Jaynes(1968) which tries to strike a balance between the methods mentioned above. Jaynes' method employs the concept of entropy which we define here as follows;

Definition: Entropy of a distribution

*The **entropy** of $\pi(N)$ which we shall denote by $\varepsilon_n(\pi)$ is defined as*

$$\varepsilon_n(\pi) = -\sum_{\aleph} \pi(N_i)\log\pi(N_i) \tag{2.6}$$

where the quantity $\pi(N_i)\log\pi(N_i)$ is defined to be zero when $\pi(N_i) = 0$.

Although Entropy is directly linked to information theory, it measures, in a sense, the amount of uncertainty inherent in a probability distribution [see for example Rosenkranzt(1977)].

For instance the prior mean can be specified so that among different prior distributions with this mean, we can seek the most noninformative distribution. Prior information can conveniently be considered in the form of restrictions on the prior distributions such as

$$E_\pi[g_k(N)] = \sum_i \pi(N_i)g_k(N_i) = \mu_k, \qquad k = 1,\ldots,m \tag{2.7}$$

Thus for $\aleph \subset \Re^1$, $g_1(N) = N$, and $g_k(N) = (N-\mu_1)^k$, $2 \le k \le m$, equation (2.7) corresponds to the specification of the first m central moments, μ_i, of π. Now it seems reasonable to seek the prior distribution which maximizes entropy among all those distributions which satisfy the given set of restrictions (i.e. given prior information). Intuitively this should result in a prior distribution which incorporates the available prior information, but otherwise is as non-informative as possible. Using calculus of variation techniques, [see for example Ewing(1969)] it can be shown that provided the distribution is proper, the solution to the maximization of the entropy $_n(\pi)$ defined in (2.6) subject to the restrictions in (2.7) and $\sum\pi(N_i) = 1$ is given by

$$\bar{\pi}(N) = \frac{\exp\{\sum_{k=1}^{m} \lambda_k g_i(N_i)\}}{\sum_i \exp\{\sum_{k=1}^{m} \lambda_k g_i(N_i)\}} \tag{2.8}$$

where the λ_k' s are constants to be determined from the constraints in (2.7).

Example

Assuming $\aleph = \{a, a+1, ..., b\}$, and it is thought that $E_\pi[N] = N_0$. This restriction is of the form (2.7) with $g_1(N) = N$, and $\mu_1 = N_0$. Then the restricted maximum entropy prior is therefore,

$$\begin{aligned} \bar{\pi}(N) &= \frac{\exp\{\sum \lambda_k g_k(N)}{\sum_i \exp\{\sum \lambda_k g_k(i)} \\ &= \frac{\exp\{\lambda_1 N)}{\sum_{i=a}^{b} \exp\{\lambda_1 i\}} \end{aligned} \tag{2.9}$$

For given values of a and b the mean of the prior in (2.9) can be set equal to N_0, and λ_1 can be obtained by solving the resulting equation. A drawback to this method is that often a maximum entropy prior may not exist although this can be overcome by truncating the parameter space. Another difficulty lies in the choice of the restrictions. Choosing moment restrictions is easiest analytically, but is generally inferior to the use of fractile restrictions from the point of view of robustness.

2.4 Point estimation under Bayesian inference approach

Under the Bayesian inference approach, the estimation of N can be done by applying some classical techniques to the posterior distribution. One such classical technique and perhaps the most popular is the Maximum Likelihood Estimation (MLE). The Bayesian analogue of MLE is the generalized MLE which we record here as follows:

Definition: Generalized MLE of N

The generalized MLE, GMLE of N is the largest mode of $\pi(N|d)$. i.e. the value of N which maximizes $\pi(N|d)$, considered as a function of N.

It is important to note here that the posterior mean and median are also commonly used as Bayesian estimates of N. Also it is recorded in the literature [see Berger(1980) pg.101] that the posterior mean and median are frequently better estimates of N than the mode.

In Classical Inference, it is customary to give, along with an estimate, an associated confidence region. A bayesian analogue to this is called a **Credible Region**.

A $100(1-\alpha)\%$ credible region for N is a subset C of $\aleph$ such that

$$1-\alpha \leq \Pr(C|d) = \sum \pi(N|d).$$

Because the posterior distribution is an actual probability distribution on $\aleph$, we can speak meaningfully (though sometimes subjectively) of the probability that N is in C. This is an advantage over the classical analogue which can only be interpreted in terms of long run coverage probability. A useful extension of this idea is the **Highest Posterior Density (HPD) Credible Region.** A $100(1-\alpha)\%$ HPD credible region for N is a subset, C of $\aleph$ of the form

$$C = \{N \in \aleph : \pi(N|d) \geq k(\alpha)\}$$

where $k(\alpha)$ is the largest constant such that

$$\Pr(C|d) \geq 1-\alpha.$$

The HPD credible region is the minimized C containing only those points with the most likely values. When $\pi(N|d)$ is unimodal, HPD credible region is very useful. However when dealing with bimodal distributions, it consists of two disjoint intervals. In such a case, the shortest interval having probability $1-\alpha$ is preferable unless it is worthwhile to have two disjoint intervals.

2.5 Loss function and Bayesian decision-theoretic approach

The Bayesian decision-theoretic analysis approach requires, in addition to a prior, a loss function which can be defined as follows;

Definition: ***Loss Function*** *$L(\theta,a)$*

A loss function $L(\theta,a)$ - is a real valued function which measures the loss incurred by taking action a ($a \in \mathcal{A}$ = set of all possible actions) when θ is the true state of nature

In our particular terminology, θ is the population size N, and the action, a to be taken corresponds to the estimate, *Nbayes* of N which we are looking for. Usually, it is assumed that $L(N, a)$ is defined for all $(N, a) \in (\aleph \times \quad)$. Technically, only $L(N, a) \geq -K > -\infty$ are considered where k is some positive real number. Three commonly used loss functions are:

1) **Squared-Error Loss**

$$L(N,a) = (N - a)^2 \tag{2.10}$$

2) **Absolute error Loss**

$$L(N,a) = |N - a| \tag{2.11}$$

3) **"0-1" Loss**

$$L(N,a) = \begin{cases} 0 & \text{if } a = N \quad \text{exactly,} \\ 1 & \text{otherwise} \end{cases} \tag{2.12}$$

Of these three, the squared-error is perhaps the most popular in practice. The reason is mainly because of its mathematical tractability. For instance, (2.10) is easily differentiated (an important requirement in minimization problems) while (2.11) and (2.12) may not be differentiated. Generally, under the decision-theoretic bayesian analysis, the posterior mean, median, and mode are respectively the optimal actions corresponding to the three loss functions. Here optimality is in the sense of minimizing the Bayes risk which is defined as follows;

Definition : Bayes Risk

The Bayes' risk, with respect to a prior distribution $\pi(N)$, is defined by

$$r(\pi,\delta) = E_N\,[R(N,\delta)] \qquad (2.13)$$

where

$$R(N,\delta) = E_{d|N}\,[\,L(N,\delta\,(d)] \qquad (2.14)$$

is the risk function in terms of which $\delta(d)$ is evaluated and

$$\delta(d):\ \Omega \rightarrow \mathcal{A}' \qquad (2.15)$$

is called the ***decision rule*** *(which in this thesis we shall equivalently refer to as the the Bayes' estimator Nbayes)*

Usually, admissibility of decision rules (estimators) is a very familiar and useful concept in mathematical statistics which can be used to select or eliminate certain obviously bad estimators. However, a basic feature of the Bayesian method is that Bayes' rules(estimators), when the exist, are in general admissible rules[see Ferguson(1967) or Lehmann(1983)]. Thus an additional principle must be introduced in order to select a specific rule for use. In classical statistics there are a number of these principles such as the maximum likelihood, unbiasedness, minimum variance and least-squares principles etc. In decision theory there are also several possible principles to go by; the three most common being the Bayes principle, the minimax principle and the invariance principle. In this thesis, we shall be concerned with the Bayes principle which we record here as follows;

The Bayes Principle:

A decision rule δ_1 *is preferred to a rule* δ_2 *if*

$$E_\pi[R(N,\delta_1)] < E_\pi[R(N,\delta_2)].$$

Thus a best decision rule (estimator) according to the Bayes principle is one which minimizes (over all $d \in \Omega$*) the Bayes risk,* $r(\pi,\delta)$ *defined in (2 13).*

Normally, one would seek to minimize the bayes risk directly. However due to theoretical difficulty involved in minimizing the bayes risk directly coupled with the practical difficulty in computing the bayes risk itself directly, there has been proposed in the literature an alternative way of going about the minimization of the Bayes' risk when these difficulties arise. These alternative views have given rise to two forms of analysis within the Bayesian decision-theoretic framework. These are: **Normal Form of Analysis**: - which consists of choosing a decision rule to minimize $r(\pi,\delta)$ directly and the alternative form of analysis [see Berger(1980)] called **Extensive Form of Analysis**: - which consists of choosing for each d, a decision rule which minimizes the posterior expected loss defined as

Definition: Posterior expected loss

The posterior expected loss is defined as the expected value of the loss function where expectation is taken under the posterior distribution. That is

$$\textit{Posterior Expected Loss} = E_{N|d}[L(N,\delta(d)]$$

The normal and the extensive forms of analysis lead to the same result. This can be shown by the following lemma.

Lemma : *The normal and extensive forms of analysis lead to the same Bayes rule.*

Proof: *Consider the Bayes' risk defined in (2.13). We have that,*

$$
\begin{aligned}
r(\pi,\delta) &= E_N[R(N,\delta)] \\
&= E_N\left\{\sum_d L(N,d(d))p(d|N)\right\} \\
&= \sum_N\left\{\sum_d L(N,\delta(d))p(d|N)\right\}\pi(N) \\
&= \sum_d\sum_N L(N,\delta(d))p(d|N)\pi(N) \qquad (2.16)
\end{aligned}
$$

Now to minimize the last expression in (2.16), $\delta(d)$ *should be chosen to minimize*

$$\sum_N L(N,\delta(d))p(d|N)\pi(N)$$

for each $d\in\Omega$. *Note that if a minimizes*

$$\sum_N L(N,a)p(d|N)\pi(N),$$

then a minimizes

$$[m(d)]^{-1}\sum_N L(N,a)p(d|N)\pi(N) = \sum_N L(N,a)\pi(N|d)$$

$$= E_{N|d}[L(N,a)].$$

The last quantity is the posterior expected loss of the action a, and is simply the expected loss with respect to $\pi(N|d)$, the posterior distribution of N given d.

2.5.1 Point estimation under Bayesian decision-theoretic approach

As pointed out earlier, the squared-error loss is popular for its mathematical tractability and it turns out to be the one mostly used for estimation problems. Under the extensive form of analysis, and using the squared-error loss function, the Bayes rule(estimator) for a given estimation problem is obtained as the action which minimizes the posterior expected loss. That is, we seek a $\delta(d)$ that minimizes $E_{N|d}[(N-\delta(D))^2]$ The posterior mean accomplishes this task. This result will be shown in chapter 3 where we shall use it to obtain the Bayes estimators resulting from different choices of prior distributions under the squared-error loss function.

Chapter 3 Application of Bayesian Theory to Estimation of N

3.1 Choice and specification of sampling strategy

In simple random sampling with replacement, the number of distinct units D in a sample has been found useful for making inference about N [see Harris(1968)] Feller(1957) gives the probability distribution of D as

$$P(d \mid N) = \binom{N}{d} \sum_{i=0}^{d} (-1)^i \binom{d}{i} \left(\frac{d-i}{N} \right)^n \tag{3.1}$$

where n is the **fixed** sample size. The expected value of D in this case is [see Basu(1958)] given by

$$E(D) = N[1-(1-\frac{1}{N})^n] \tag{3.2}$$

Harris (1968) considered estimation of N using the distribution of D given in (3.1). Recently, Ahmad et. al. (1995) introduced two variants of this sheme in which the sample sizes are **random**. One of these schemes is a strategy which considers random independent draws until one of the previously drawn units reappears, in which case, the probability distribution of D is given by

$$P(D = d \mid N) = \frac{dN!}{N^{d+1}(N-d)!} \tag{3.3}$$

where $d = 1, 2, \ldots, N$. The sample size (which in this case is random) is $D + 1$. They observed that the method of moments estimator coincides with the maximum likelihood estimator which is the same in (3.1) and (3.2), since both involve maximizing the same likelihood function.

$$L(N) = \frac{N!}{(N-d)!} N^{-n} \tag{3.4}$$

3.1.1 MLE and method of moments estimator of N

For completeness, we summarize here the results on the maximum likelihod and moment estimators of N based on the sampling distribution of D given in (3.3) which were obtained by Ahmad et. al. (1995). As in Ahmad et. al. (1995), one can equivalently maximize the function $\Psi(x)$ with respect to $x = 1/N$, where

$$\Psi(x) = x^r \prod_{i=1}^{d-1} (1 - ix) \tag{3.5}$$

and $r = n\text{-}d$. Since necessarily, $N > d\text{-}1$, we must have

$$x < (d-1)^{-1}.$$

Proposition 3.1.1 *Provided $r \geq 1$, $\Psi(x)$ does have a unique maximum in*

$$(0,(d-1)^{-1}).$$

Proof:

Setting the derivative of $\log[\Psi(x)]$ to zero, the unique value x_0 maximizing $\Psi(x)$ must satisfy $\Phi(x_0) = r$, where

$$\Phi(x) = \sum_{i=1}^{d-1} \frac{ix}{1-ix} \tag{3.6}$$

The function $\Phi(x)$ is an increasing function on

$[0,(d-1)^{-1})$, $\Phi(0) = 0$ *and* $\lim_{x \uparrow (d-1)^{-1}} \Phi(x) = \infty$.

Therefore, there is a unique solution to $\Phi(x) = r$ in $(0,(d-1)^{-1})$, except when $r = 0$, when the solution $x_0 = 0$, corresponds to $\hat{N} = \infty$ is not admissible. Since any solution x can take only values of the form $1/k$ for an integer $k > d\text{-}1$, the ML estimator $\hat{N}$ is not necessarily the value $1/x_0$, but it is one of the two integers $\hat{N}$ and $\hat{N}+1$ satisfying

$$\hat{N} \leq \frac{1}{x_0} < \hat{N} + 1$$

and therefore we can take the integer part of $1/x_0$ to be the ML estimator

Proposition 3.1.2 *The Maximum Likelihood and the Method of Moments estimators coincide.*

Proof:

In (3.1), the moment estimator is obtainable by using the equality in (3.2). That is, *by solving for X the equation given by*

$$X\left(1-\left(1-\frac{1}{X}\right)^{n}\right)=d \qquad (3.7)$$

or, equivalently

$$\left(1-\frac{1}{X}\right)^{n}=1-\frac{d}{X} \qquad (3.8)$$

Again by setting $x = 1/X$, it can be shown by drawing graphs of $(1-x)^n$ and that of $1-dx$ the equation (3.7) has a unique solution X such that

$$\hat{X} \leq \frac{1}{x_1} < \hat{X}+1 \qquad (3.9)$$

where x_1 is the unique solution of $(1-x)^n = (1-dx)$. To show that $\hat{X}$ is also the MLE we show that

$$L(\hat{X}) \geq L(\hat{X}-1) \ \textit{and} \ L(\hat{X}) \geq L(\hat{X}+1) \qquad (3.10)$$

that is

$$\frac{\left(1-\frac{1}{\hat{X}}\right)^{n}}{\left(1-\frac{d}{\hat{X}}\right)} \geq 1 \ \textit{and} \ \left(1-\frac{d-1}{\hat{X}}\right)\left(1+\frac{1}{\hat{X}}\right)^{n-1} \geq 1$$

or, with $x = 1/\hat{X}$, we must have

$$\frac{(1-x)^{n}}{1-dx} \geq 1 \ \textit{and} \ (1-(d-1)x)(1+x)^{n-1} \geq 1$$

It can be observed that

$x < x_1$ *iff* $\frac{(1-x)^n}{1-dx} < 1$. *Since* $1/\hat{N} \geq x_1$, *we must have* $\frac{(1-x)^n}{1-dx} \geq 1$, *i.e.*, $L(\hat{N}) \geq L(\hat{N}-1)$. *Further, Since,* $1/(\hat{N}+1) < x_1$, $\frac{(1-x')^n}{1-dx'} < 1$, *for* $x' = 1/(\hat{N}+1)$ *i.e.*, $L(\hat{N}) \geq L(\hat{N}+1)$. *Therefore,* $\hat{N}$ *is a "local" maximum. Since,* $L(N)$ *has a single maximum for* $N > (d-1)$, $\hat{N}$ *is also an absolute maximum.*

Since both (3.1) and (3.2) lead to the same likelihood function, we shall be using the latter as our sampling distribution beause of its simpler form.

3.2 Choice and specification of prior distribution

Since $\boldsymbol{N}$ is necessarily a positive integer, it would be wise to look to the class of power series distribution as a natural candidate in the choice of prior distribution for $\boldsymbol{N}$.

Definition 3.2.1

The distribution of the discrete random variable N with probabilities

$$P(\boldsymbol{N} = N) = \begin{cases} \dfrac{a(N)\theta^N}{C(\theta)} & \text{for } N = 0,1,\ldots \\ 0 & \text{elsewhere} \end{cases} \qquad (3.11)$$

where $a(N) \geq 0$; $\theta > 0$ and $\sum a(N)\theta^N < \infty$ is called a power series distribution.

The power series distribution is an exponential family and as such (3.11) can be expressed in canonical form as

$$p(N;\theta) = \exp\{N\log(\theta) - C(\theta)\}\, a(N) \qquad (3.12)$$

where C in (3.11) and (3.12) is a real-valued function of θ.

From (3.12) or otherwise the moment generating function can be written as

$$M_N(t) = \frac{C(\theta e^t)}{C(\theta)} \qquad (3.13)$$

It can be shown that the binomial, the negative binomial, and Poisson distributions are special cases of the power series distribution for appropriate values of θ and $C(\theta)$.

Proposition 3.2.1: *The binomial distribution, b(N; r1, p1) can be obtained from (3.11).*
Proof: *By putting*

$a(N) = \binom{r1}{N}$ *for* $N = 0,1,\ldots,r1$, *and* $a(N) = 0$ *otherwise;* $\theta = p1/(1-p1)$ *and* $C(\theta) = (\theta+1)^{r1}$,
where r1 is a positive integer and $0 \leq p1 \leq 1$, *we obtain*

$$p(N) = \frac{\binom{r1}{N}\left(\frac{p1}{1-p1}\right)^{N}}{\left(\frac{p1}{1-p1}+1\right)^{r1}}$$

$$= \begin{cases} \binom{r1}{N}(p1)^{N}(1-p1)^{N-r1} & \text{for } N = 0,1,\ldots,r1 \\ 0 & \text{elsewhere} \end{cases}$$

$$\equiv b(N; r1, p1).$$

Proposition 3.2.2 *The negative binomial distribution, nb(N; r2, p2) can be obtained from (3 11)*

Proof: *By putting*

$a(N) = \binom{r2+N-1}{r2-1}$ *for* $N = 0,1,2,\ldots$ *and* $a(N) = 0$, *otherwise,* $\theta = (1-p2)$,

and $C(\theta) = (1-\theta)^{-r2}$
where r2 is a positive intege. and $0 \leq p2 < 1$, *we have that*

$$p(N) = \frac{\binom{r2+N-1}{N}(1-p2)^{r2}}{(1-(1-p2))^{N}}$$

$$= \begin{cases} \binom{r2+N-1}{N}(1-p2)^{r2}(p2)^{N}, & N = 0,1,2,\ldots \\ 0 & \text{elsewhere} \end{cases}$$

$$\equiv nb(N; r2, p2)$$

Proposition 3.2.3 *The Poisson distribution, p(N;λ) can be obtained from (3 11).*

Proof: *By putting*

$$a(N) = \frac{1}{N!}, \quad \theta = \lambda, \quad and \quad C(\theta) = e^{\lambda}$$

where $\lambda > 0$, *we have*

$$\begin{aligned}
p(\boldsymbol{N} = N) &= \frac{1/N!\,(\lambda)^N}{e^{\lambda}} \\
&= \begin{cases} \dfrac{e^{-\lambda}\lambda^N}{N!}, & N = 0,1,2\ldots \\ 0, & elsewhere \end{cases} \\
&\equiv p(N;\lambda).
\end{aligned}$$

Evidently truncation is in the spirit of Bayesian analysis since it does use prior information and notably in this case where it is obvious that N is finite. That is we have that $0 < N < \infty$. Thus we shall be using the doubly truncated form of the power series distribution which we record here as follows:

Definition 3.2.2: ***Doubly Truncated Power Series Distribution***

The distribution of N given in (3.11) is said to be doubly truncated to the left at a and to the right at b when

$$P(\boldsymbol{N} = N) = P(N \mid a \le N \le b) = \begin{cases} 0 & for \;\; N < a,\; N > b \\ P(N) \Big/ \sum\limits_{N=a}^{b} P(N) & for \;\; a \le N \le b \end{cases} \tag{3.14}$$

where a and b, $a \le b$, are positive integers.

3.2.1 Determination of prior parameters

After choosing a particular functional form, the next step in the specification of a prior distribution is the determination of the prior parameters. A number of ways for determining the parameters of a prior distribution are available in the literature. Iversen(1984) suggests two methods. The first method is by trial and error, preferably using a computer to plot $\pi(N)$ for various values of the parameters. Then the particular $\pi(N)$ that most closely represents our prior knowledge is chosen. The drawback to this method is that the curves for nearby values of the parameters would not be very different, and there could well be other curves that represent our limited prior knowledge equally well. The second method which we shall adopt here is similar to the maximum entropy method described in chapter 2 and it involves the specification of the expected value and variance of N. This may be guided by a target value or a preliminary estimate of N, say M, along with a measure of its precision V. Then the prior parameters can be computed from M and V by equating the first two moments. The truncation points a and b may also be obtained from M and V by setting

$$a = M - S\sqrt{V} \quad and \quad b = M + S\sqrt{V}$$

where S, $0 < S < \infty$ is some real number indicating the *spread* of N around M.

In line with the foregoing, interactive programming in MapleV3 [see appendix A] was used to determine, for particular situations, the prior parameters associated with the three main prior distributions considered in the sequel, namely:

i) Truncated binomial prior [$\pi_{tb}(N; r1,p1,a,b)$]

Definition 3.2.3

We shall call the prior distribution resulting from a truncated binomial distribution the truncated binomial prior [$\pi_{tb}(N; r1,p1,a,b)$] which is given by

$$\pi_{tb}(N;r1,p1,a,b) = \frac{\binom{r1}{N}\left(\frac{p1}{1-p1}\right)^N}{\sum_{i=a}^{b}\binom{r1}{i}\left(\frac{p1}{1-p1}\right)^i}, \quad for\ N = a, a+1, \ldots, b \qquad (3.15)$$

where $r1$, a, b, $a \le b$, are positive integers, $0 < p1 < 1$, and $\pi_{tb}(N; r1,p1,a,b) = 0$ elsewhere.

ii) Truncated negative binomial prior [$\pi_{tnb}(N;r2,p2,a,b)$]

Definition 3.2.4

We shall call the prior distribution resulting from a truncated negative binomial distribution the truncated negative binomial prior [$\pi_{tnb}(N;r2,p2,a,b)$] which is given by

$$\pi_{tnb}(N;r2,p2,a,b) = \frac{\binom{r2+N-1}{r2-1}(1-p2)^N}{\sum_{i=a}^{b}\binom{r2+i-1}{r2-1}(1-p2)^i}, \quad for \quad N = a, a+1, \ldots, b \qquad (3.16)$$

where $r2$, a, b, $a \leq b$, are positive integers, $0 < p2 < 1$, and $\pi_{tnb}(N;r2,p2,a,b) = 0$ elsewhere.

iii) **Truncated Poisson prior [$\pi_{tp}(N;\lambda,a,b)$]**

Definition 3.2.5

We shall call the prior distribution resulting from a truncated poisson distribution the truncated poisson prior [$\pi_{tp}(N;\lambda,a,b)$] which is given by

$$\pi_{tp}(N;\lambda,a,b) = \frac{\frac{\lambda^N}{N!}}{\sum_{i=a}^{b}\frac{\lambda^i}{i!}}, \quad for\ N = a, a+1, \ldots, b \qquad (3.17)$$

where $\lambda > 0$, a, b, $a \leq b$, are positive integers, and $\pi_{tp}(N;\lambda,a,b) = 0$ elsewhere.

3.3 Derivation of posterior distributions and Bayes' estimators

Using the formulations in (2.3) - (2.5) of chapter 2, and the sampling distribution of D given in (3.3), the posterior distributions of N corresponding to the prior distributions in (3.15) - (3.17) are derived as follows:

The joint probability mass function of D and N is $f(d,N) = \pi(N)P(d|N)$ for $d = 1, \ldots, N$, and $N = \max\{a,d\}, \ldots, b$, since $a \leq N \leq b$, and $N \geq d$. The marginal distribution of D is obtainable by summing out d from this joint function. Hence dividing the joint function by the marginal for D we obtain the following posterior distributions;

i) **Truncated binomial posterior [$\pi_{tb}(N|d; r1,p1,a,b)$]**

Definition 3.3.1

We shall call the posterior distribution resulting from a truncated binomial prior distribution the truncated binomial posterior [$\pi_{tb}(N|d;\, r1, p1, a, b)$] which is given by

$$\pi_{tb}(N|d;r1,p1,a,b) = \frac{\binom{r1}{N}\left(\frac{p1}{1-p1}\right)^{N} N!}{N^{(d+1)}(N-d)! \sum_{i=\max(a,d)}^{b} \binom{r1}{i}\left(\frac{p1}{1-p1}\right)^{i} i! \; i^{(d+1)}(i-d)!} \tag{3.18}$$

ii) **Truncated negative binomial posterior [$\pi_{tnb}(N|d;\, r2, p2, a, b)$]**

Definition 3.3.2

We shall call the posterior distribution resulting from a truncated negative binomial prior distribution, the truncated negative binomial posterior [$\pi_{tnb}(N|d;r2,p2,a,b)$] which is given by

$$\pi_{tnb}(N|d;r2,\,p2,a,b) = \frac{\binom{r2+N-1}{N}(p2)^{N}\, N!}{N^{(d+1)}(N-d)! \sum_{i=\max(a,d)}^{d} \frac{\binom{r2+i-1}{i}(p2)^{i}\, i!}{i^{(d+1)}(i-d)!}} \tag{3.19}$$

iii) **Truncated poisson posterior [$\pi_{tp}(N|d;\lambda, a, b)$]**

Definition 3.3.3

We shall call the posterior distribution resulting from a truncated poisson distribution the truncated poisson posterior [$\pi_{tp}(N|d;\lambda, a, b, d)$] which is given by

$$\pi_{tp}(N|d;\lambda,\, a,b) = \frac{\lambda^{N}}{N^{(d+1)}(N-d)! \sum_{i=\max(a,d)}^{b} \frac{\lambda^{N}}{i^{(d+1)}(i-d)!}} \tag{3.20}$$

In chapter 2, we pointed out that the development of Bayes's estimators can be approached in two ways; as a traditional statistical inference problem or as a statistical decision problem. By statistical inference, we mean [see Box and Tiao(1973)] inference about the state of nature made in terms of probability, and a statistical inference problem is regarded as solved as soon as one can make appropriate probability statements about the state of nature in question. Thus under the statistical inference approach, a solution to the estimation problem is supplied by the posterior distribution $\pi(N|d)$ which shows what can be inferred about N from the data d given a relevant prior state of knowledge modelled as $\pi(N)$. In section 2.3, we introduced the Generalized Maximum Likelihood Estimator, *GMLE* as a Bayesian analogue of the classical MLE. The *GMLE* is simply the largest mode of the posterior distribution [see definition 2.3.1]. Thus the posterior distribution can be examined to cast light on the precision of the estimate. This is usually done numerically for specific situations. Some numerical examples are given in tables 1 - 3 below.

Example (*i*) Prior mean, $M = 13$, Prior Variance, $V = 1/9$, $a = 13$, $b = 24$.

Table 1. Posterior Distributions and Generalized Maximum Likelihood Estimates of N for different values of d.

N	$\pi_{tb}(N\|d, r1, p1, a, b)$	$\pi_{tb}(N\|d; r1, p1, a, b)$	$\pi_{tp}(N\|d; \lambda, a, b)$
		$d = 1$	
13	0.9961	0.9834	0.9961
14	0.0039	0.0164	0.0039
15	0.0000	0.0003	0.0000
24	0.0000	0.0000	0.0000
		$d = 2$	
13	0.9960	0.9833	0.9960
14	0.0040	0.0165	0.0040
24	0.0000	0.0000	0.0000
		$d = 6$	
13	0.9956	0.9814	0.9956
14	0.0044	0.0183	0.0044
24	0.0000	0.0000	0.0000
		$d = 7$	
13	0.9953	0.9802	0.9953
14	0.0047	0.0194	0.0047
15	0.0000	0.0004	0.0000
24	0.0000	0.0000	0.0000
		$d = 13$	
13	0.9791	0.9146	0.9791
14	0.0207	0.0813	0.0207
15	0.0002	0.0040	0.0002
16	0.1444	0.0001	0.0000

24	0.0000	0.0000	0.0000
		$d = 15$	
15	0.9825	0.9099	0.9790
16	0.0174	0.0855	0.0208
17	0.0001	0.0044	0.0002
18	0.0000	0.0002	0.0000
24	0.0000	0.0000	0.0000
		$d = 24$	
24	1.0000	1.0000	1.0000

From table 1, it can be observed that the *GMLE* of N when d = 1, 2, 6, 7, 13, 15 and 24 are respectively 13, 13, 13, 13, 13, 15, and 24. When $d = 1$, all values of $N > a = 13$ are improbable except in the case of the truncated negative binomial. As d increases, some other values of N other than the mode tend to be probable.

Example (*ii*) $M = V = 10, a = 1, b = 30$.

Table 2. Posterior Distributions and Generalized Maximum Likelihood Estimates of N

N	$\pi_{tb}(N\|d; r1,p1,a,b)$	$\pi_{tb}(N,d, r1,p1,a,b)$	$\pi_{tp}(N\|d;\lambda,a,b,d)$
		$d = 1$	
1	0.0039	0.0042	0.0040
2	0.0097	0.0105	0.0100
3	0.0218	0.0231	0.0223
4	0.0412	0.0429	0.0418
5	0.0664	0.0682	0.0669
6	0.0926	0.0941	0.0930
7	0.1139	0.1145	0.1139
* 8	0.1249	0.1246	0.1245 *
9	0.1237	0.1225	0.1230
10	0.1114	0.1098	0.1107
11	0.0920	0.9043	0.0915
15	0.0204	0.0201	0.0205
20	0.0008	0.0008	0.0008
30	0.0000	0.0000	0.0000
		$d = 5$	
5	0.0108	0.0112	0.0109
6	0.0364	0.0373	0.0366
7	0.0724	0.0736	0.0726
8	0.1087	0.1095	0.1086
9	0.1343	0.1344	0.1339
* 10	0.1429	0.1429	0.1423 *

11	0.1344	0.1128	0.1339
20	0.0019	0.0020	0.0020
30	0.0000	0.0000	0.0000

From table 2, it can be observed that when $d = 1$, although $GMLE = 8$ with maximum posterior probability, both 7, 9, and 10 are not unlikely and none of the values of $N < 20$ is improbable(to four decimal places). The situation is the same as d increases. It was observed that from $d = 10$ upwards, $GMLE = d + 4$. except when $d = b = 30$, in which case $GMLE = 30$ obviously.

Example (*iii*) $M = 3$, $V = 4$, $a = 1$, $b = 20$.

Table 3. Posterior Distributions and Generalized Maximum Likelihood Estimates of N

N	$\pi_{tb}(N\|d;\, r1, p1, a, b)$	$\pi_{tb}(N\|d;\, r1, p1, a, b)$	$\pi_{tp}(N\|d; \lambda, a, b, d)$
	$d = 1$		
1	0.3234	0.5200	0.2334
2	0.2830	0.2366	0.2299
3	0.2044	0.1196	0.2013
4	0.1150	0.0612	0.1487
5	0.0506	0.0312	0.0937
6	0.0158	0.1577	0.0513
7	0.0084	0.0039	0.0248
8	0.0011	0.0039	0.0107
9	0.0002	0.0020	0.0042
10	0.0000	0.0010	0.0015
11	0.0000	0.0005	0.0005
12	0.0000	0.0002	0.0001
13	0.0000	0.0001	0.0000
14	0.0000	0.0000	0.0000
20	0.0000	0.0000	0.0000
	$d = 2$		
2	0.3334	0.4000	0.2234
3	0.3211	0.2697	0.2608
4	0.2032	0.1553	0.2167
5	0.0954	0.0844	0.1457
6	0.0345	0.0445	0.0831
7	0.0098	0.0229	0.0412
8	0.0022	0.0116	0.0181
9	0.0004	0.0058	0.0072
10	0.0000	0.0029	0.0026
11	0.0000	0.0014	0.0008
12	0.0000	0.0007	0.0003
13	0.0000	0.0000	0.0000
20	0.0000	0.0000	0.0000

	$d = 10$		
10	0.6530	0.0738	0.2286
11	0.2861	0.1531	0.3157
12	0.0549	0.1531	0.2388
13	0.0057	0.1766	0.1300
14	0.0031	0.1422	0.0567
15	0.0000	0.1030	0.0209
16	0.0000	0.0691	0.0068
17	0.0000	0.0438	0.0020
18	0.0000	0.0266	0.0005
19	0.0156	0.0156	0.0001
20	0.0000	0.0089	0.0000
	$d = 15$		
15	1.0000	0.0392	0.2320
16	0.0000	0.1144	0.3255
17	0.0000	0.1875	0.2431
18	0.0000	0.2278	0.1279
19	0.0000	0.2019	0.0184

From table 3, it can be observed that when d = 1, 2, $GMLE$ = 1, 2 respectively under all three priors. However the truncated binomial gives these values with higher probabilities. As d inreases, for example, when d = 10, 15, $GMLE$ differs for all three priors. The truncated binomial gives $GMLE = d$, until d =15 after which all other values of N are improbable. The truncated negative binomial gives $GMLE = d+2$ and d+4 respectively, while the truncated poisson gives $GMLE = d+1$.

Under the Bayesian decision-theoretic approach, a loss function is needed in order to arrive at a Bayesian estimator. The loss function was introduced in chapter 2. There, we remarked that the squared-error loss function is the most widely used because of its mathematical tractability. In what follows, we show that under squared-error loss function, the Bayes' estimators of N, $\tilde{N}$ resulting from the three priors given in (3.15)-(3.17) are respectively the posterior means of the corresponding posterior distributions given in (3.18) - (3.20).

Lemma 3.3.1 *The Bayes' estimator of N under squared-error loss is the mean of the posterior distribution of N.*

Proof: *Under the extensive form of analysis [see section 2.4], the Bayes' estimator is the decision rule which minimizes for each d, the posterior expected loss. Using definition (2.4.5), the posterior expected loss is*

$$E_{N|d}[L(N,\delta(d)] = \sum_{N=\max(a,d)}^{b} (N-\delta(d))^2 \pi(N|d).$$

Expanding the quadratic expression, differentiating with respect to δ *, and setting equal to zero, the value of* $\delta(d)$ *which minimizes the posterior expected loss can be obtained from the resulting equation*

$$0 = \frac{d}{d\delta}\left(\sum N^2\pi(N|d) - 2\delta\sum N\pi(N|d) + \delta^2\sum \pi(N|d)\right)$$

$$= -2E_{N|d}[N] + 2\delta$$

Now by solving for δ *the last expression, the desired result is obtained as*

$$\delta(d) = E_{N|d}[N].$$

Thus adopting a squared-error loss, lemma 3.3.1 leads to the following results;

Result 3.3.1.

The Bayes' estimator for N under squared-error loss function and a truncated binomial prior, which we shall denote by $\tilde{N}_{tb}(r1,p1,a,b,d)$, is the mean of the truncated binomial posterior distribution, $\pi_{tb}(N|d;\ r1,p1,a,b)$, defined in (3.18). That is

$$\tilde{N}_{tb}(r1,p1,a,b,d) = \sum_{N=a}^{b} N \times \pi_{tb}(N|d;r1,p1,a,b) \tag{3.21}$$

Result 3.3.2.

The Baye's estimator for N under squared-error loss function and a truncated negative binomial prior, which we shall denote by $\tilde{N}_{tnb}(r2,p2,a,b,d)$, is the mean of the truncated negative binomial posterior distribution, $\pi_{tnb}(N|d;r2,p2,a,b)$, defined in (3.19). That is

$$\tilde{N}_{tnb}(r2,p2,a,b,d)=\sum_{N=a}^{b} N\times\pi_{tnb}(N|d;r2,p2,a,b) \tag{3.22}$$

Result 3.3.3.

The Bayes' estimator for N under squared-error loss function and a truncated poisson prior, which we shall denote by $\tilde{N}_{tp}(\lambda,a,b)$, is the mean of the truncated poisson posterior distribution, $\pi_{tp}(N|d;\lambda,a,b)$, defined in (3.20). That is

$$\tilde{N}_{tp}(\lambda,a,b,d)=\sum_{N=a}^{b} N\times\pi_{tp}(N|d;a,b) \tag{3.23}$$

The following tables show the computed values of the estimators for different values of d, and different values of prior mean M and and prior varince V corresponding to the three different situations where M is less, equal or greater than V.

Note: In the following tables, we have reported the values of the Bayes estimators as computed. Although values should be rounded off to the nearest integer, they are not expected to make difference in computations. As a matter of fact the Bayes' estimator, $\tilde{N}$ is one of the two integers $\tilde{N}$ and $\tilde{N}+1$ satisfying $\tilde{N}\le\tilde{N}_*<\tilde{N}+1$ where $\tilde{N}_*$ is the reported value in the table. [see a similar argument in section 3.1.1]

Example (*i*) when $M = 13 > V = 1/9,\ a = 13,\ b = 24.$

Table 4. Bayes' estimates of N for different values of d

d	$\tilde{N}_{tb}(r1,p1,a,b,d)$	$\tilde{N}_{tnb}(N\|d;r2,p2,a,b)$	$\tilde{N}_{tp}(N\|d;\lambda,a,b)$
1	13.00396	13.00396	13.00399
5	13.00426	13.00426	13.00426
10	13.00662	13.00662	13.00662
12	13.01142	13.01143	13.01142
13	13.02112	13.02116	13.02115
14	14.01941	14.02177	14.02120
15	15.01768	15.02238	15.02125
23	23.00356	23.00264	23.00264
24	24.00000	24.00000	24.00000

It can be observed that table 4 suggests that the Bayes' estimators in all three cases are almost the same. For $d < a$, $\tilde{N} \cong a$, while for $d \geq a$, $\tilde{N} \cong d$. This suggests that the Bayes' rule could tend to be, choosing $\max\{a,d\}$.

Example (*ii*) when $M = 10 = V = 10,\ a = 1,\ b = 30.$

Table 5. Bayes' estimates of N for different values of d.

d	$\tilde{N}_{tb}(r1,p1,a,b,d)$	$\tilde{N}_{tnb}(N\|d;r2,p2,a,b)$	$\tilde{N}_{tp}(N\|d;\lambda,a,b)$
1	8.8597	8.8051	8.8479
5	10.7023	10.6870	10.7092
10	14.7869	14.8176	14.8181
15	19.4028	19.4707	19.4547
20	24.1378	24.2292	24.2023
30	30.0000	30.0000	30.0000

Example (*iii*) when $M = 3 < V = 4,\ a = 1,\ b = 20.$

Table 6. Bayes' estimates of N for different values of d.

d	$\tilde{N}_{tb}(r1,p1,a,b,d)$	$\tilde{N}_{tnb}(N\|d;r2,p2,a,b)$	$\tilde{N}_{tp}(N\|d;\lambda,a,b)$
1	2.3649	1.9722	2.9866
2	3.2167	3.2879	3.8607
10	10.4143	13.3045	11.5742
15	15.0000	18.0990	16.4997
16		18.6802	17.4314
17		19.1413	18.2853
18		19.5018	19.0140
19		19.7821	19.5855
20		20.0000	20.0000

Chapter 4 Sensitivity Analysis of the Bayes' Estimators to change in Prior Distribution

4.1 Introduction

In chapter 2, we pointed out that a Bayesian decision problem model contains three basic elements; the sample distribution, the prior distribution, and the loss function. Assumptions in the model are usually made regarding aspects of these basic elements. The sensitivity of a decision rule(estimator) to assumptions in the model is called the robustness of the rule and it is usually investigated with respect to any one of these three basic elements. The specification of the sample distribution is often less subjective than that of the loss function and the prior. As such, attention is often devoted more to robustness of losses and priors. The squared-error loss function is frequently the choice among other types of loss function because of mathematical convenience (especially in estimation problems) and as such one can limit attention to robustness of priors .

In the terminology of our particular problem of estimating N using D, the Bayes' estimators, $\tilde{N}$ may be sensitive not only to the assumptions about the sufficient statistic D, but also to changes in the structure of N which is the indexing parameter of the probability distribution $P(d|N)$ of D. Since all that is known about N apriori is believed to be appropriately modelled in the prior distribution $\pi(N)$, any inadequacy or change in $\pi(N)$ will surely affect the inference process about N. Here, we will not be concerned with robustness or sensitivity of the estimators with respect to the sample distribution since the choice of $P(d|N)$ is in a sense, less subjective than the choice of the loss function and prior distributions we used in the analysis.This is not to say that in general, robustness considerations concerning $P(d|N)$ cannot have a significant influence upon the decision rule. Rather what we are saying in this particular case is that in point of fact of remark 2.1.1, since $P(d|N)$ in (3.1) and (3.3) lead to the same likelihood function, one would necessarily arrive at the same posterior distribution and hence the same estimator.

Usually a loss function may not pose serious robustness difficulties except when the error is multiplied by a weighting factor, say $w(N)$ which indicates the importance of N. But in general, any such factor would have exactly the same effect on the estimator as the prior $\pi(N)$. Thus we will subsume robustness with respect to $w(N)$ in the discussion of sensitivity with respect to the prior.

4.2 Sensitivity with respect to prior distribution

In order to apply Bayesian analysis with confidence, a study of the prior robustness is crucial. The problem involved may be posed as follows: How much will the posterior inference (or desicion) change if we change the prior? This is the important concern we will look into in this section. Obviously the key to the Bayesian approach lies in the specification of the prior distribution, and it is here that difficulties can arise. The prior can seldom be specified accurately and errors in the specification can have adverse effect on the correctness of the final decision. Therefore to conduct a reliable Bayesian analysis, it is necessary to investigate the effects of inaccurate specification of the prior. It is suggested in the literature [see Berger(1980)] that the best way to do this would be often through examination of the risk function $R(N,\tilde{N})$ of the Bayes estimator, $\tilde{N}$ or alternatively through consideration of certain classical properties of $\tilde{N}$.

To investigate robustness with respect to the prior, we ideally specify a class Γ of plausible prior distributions, and see how the choice among the priors in Γ affects the analysis. This can be approached in two different ways: either through risk robustness considerations or through consideration of posterior robustness which measures the effect of the prior on the expected posterior loss.

4.2.1 Posterior robustness

Definition 4.2.1: *An action is said to be* ***posterior robust with respect to*** Γ, *if for all* $\pi \in \Gamma$, *the posterior expected loss of the action is close to the optimal expected loss.*

If for all $\pi \in \Gamma$, the posterior expected loss of an action is within ε of the optimal posterior expected loss, the action is said to be ε - **Posterior Robust.**

Proposition 4.2.1: *In estimation under squared-error loss, the posterior expected loss of an action is close to the optimal posterior expected loss whenever the action is close to the optimal Bayes action.*

Proof: *Under squared-error loss, the action minimizing the posterior expected loss is the mean, that is the Bayes action. Let this Bayes action be denoted by* a_π *and let* a_0 *be any other action. Then,*

$$E_{\pi(N|d)}[L(N,a_0)] - E_{\pi(N|d)}[L(N,a_\pi)] = E_{\pi(N|d)}[(N-a_0)^2 - (N-a_\pi)^2]$$

$$= E_{\pi(N|d)}[(a_\pi - a_0)(2N - a_\pi - a_0)]$$

$$= (a_\pi - a_0)(2a_\pi - a_\pi - a_0)$$

$$= (a_\pi - a_0)^2 .$$

The last expression implies that the posterior expected loss of a_0 is close to the optimal posterior expected loss if a_0 is close to the optimal Bayes action, a_π.

Here it is important to remark that analysis through posterior robustness will often depend crucially on which value of D is observed and as such it can best be investigated after the sample is at hand. The only situations in which posterior robustness is obvious are those in which the sample information are conclusive. Frequently posterior robustness is found to be lacking. However if the Bayes rule is posterior robust, there is no need for further investigation.

4.2.2 Risk robustness

There are many possible measures of risk robustness. The two most appealing measures are through examination of $R(N,\tilde{N})$ given in (2.14) or through examination of bayes risk. $r(\pi, \tilde{N})$. Examination of $R(N,\tilde{N})$ consists of simple calculation and comparison of the risk functions of the various $\tilde{N}$ under consideration to see which rules are sensitive to uncertainties in the prior specification and which are not. Although this is not a rigorous way, it is good practice, upon obtaining a Bayes rule, to look at $R(N,\tilde{N})$ for any unappealling features. The graphs of $R(N,\tilde{N})$ for the estimators, $\tilde{N}$ can be plotted for some numerical examples. Examination of bayes risk, $r(\pi, \tilde{N})$ is needed for a formal analysis of risk robustness which can be done by a so called Γ-minimax approach explicitly introduced by Robbins(1964).

Definition 4.2.2: *Γ-minimax risk of a rule $\tilde{N}$ - is defined as*

$$r_{\Gamma}(\tilde{N}) = \sup_{\pi\in\Gamma} r(\pi,\tilde{N})$$

Γ-minimax Principle :

$\tilde{N}_1$ should be prefered to $\tilde{N}_2$ if

$$r_{\Gamma}(\tilde{N}_1) < r_{\Gamma}(\tilde{N}_2)$$

Definition 4.2.3:

*Γ- **minimax Value of a decision problem** - is defined as*

$$r_{\Gamma} = \inf_{\pi\in\Gamma} r_{\Gamma}(\tilde{N}) = \inf_{\pi\in\Gamma}\sup_{\pi\in\Gamma} r(\pi,\tilde{N})$$

A rule $\tilde{N}^*$ is called the Γ - minimax rule if $r_{\Gamma}(\tilde{N}^*) = r_{\Gamma}$.

The quantity $r_{\Gamma}(\tilde{N}^*)$ represents the "worst Bayes risk" that can happen if $\pi \in \Gamma$, and our aim is to seek a rule for which this worst Bayes risk is as small as possible.

r_{Γ} is the lowest possible worst Bayes risk. Thus a Γ- minimax rule (if one exists) achieves the desired goal. Note that by setting Γ equal to only one prior, the Γ - minimax principle is made equivalent to the Bayes principle introduced in section 2.4.

Thus for a formal analysis of risk robustness, we specify Γ as the class

$$\Gamma = \{ \pi_{tb}(N; r1,p1,a,b) \ \pi_{tb}(N; r1,p1,a,b) \ \pi_{tp}(N;\lambda ,a,b,d) \}$$

of truncated power series prior distributions, and see how the choice among these priors in Γ affects the analysis. We examine the bayes risks $r(\pi, \tilde{N})$ associated with the priors, *tbprior, tnbprior,* and *tpprior* under the Bayes rules (estimators), $\tilde{N}_{tb}$, $\tilde{N}_{tnb}$ and $\tilde{N}_{tp}$. The matrices of Bayes risks were computed using MapleV3 [see appendix B]. Tables1-3 below show three examples of these matrices for the three situations corresponding to cases when the prior mean is greater than, equal to and less than the prior variance respectively. The computations in the tables were done using the following decomposition of the risk function;

$$r(N, \tilde{N}) = E_D[\tilde{N}(D) - \tilde{N}_B(D)]^2 + E_D[Var(N|D)]$$

where $\tilde{N}_B(D) = E[N|D]$ is the Bayes estimator.

Example (*i*). [prior mean, $M = 13$ and prior variance, $V = 1/9$, $a = 13$, $b = 24$]

Table 7. Matrix of Bayes Risks (r_{ij}) when prior mean is greater than prior variance.

	Bayes Estimator ($\tilde{N}$)		
Prior	$\tilde{N}_{tb}$ 1	$\tilde{N}_{tnb}$ 2	$\tilde{N}_{tp}$ 3
1) π_{tb}	0.0042850	0.0042850	0.0042850
2) π_{tnb}	0.0042886	0.0042886	0.0042886
3) π_{tp}	0.0042876	0.0042876	0.0042876

From table 7, we have the following results;

$$r_\Gamma(\widetilde{N}_{tb}) = \sup_{\pi \in \Gamma} r(\pi, \widetilde{N}_{tb}) = \sup\{r_{11}, r_{21}, r_{31}\} = r_{21} = 0.0042886$$

$$r_\Gamma(\widetilde{N}_{tnb}) = \sup_{\pi \in \Gamma} r(\pi, \widetilde{N}_{tnb}) = \sup\{r_{12}, r_{22}, r_{32}\} = r_{22} = 0.0042886$$

$$r_\Gamma(\widetilde{N}_{tp}) = \sup_{\pi \in \Gamma} r(\pi, \widetilde{N}_{tp}) = \sup\{r_{13}, r_{23}, r_{33}\} = r_{23} = 0.0042886$$

From the above results it turns out in this case that

$$r_\Gamma(\tilde{N}_{tb}) = r_\Gamma(\tilde{N}_{tnb}) = r_\Gamma(\tilde{N}_{tp}) .$$

Thus the Γ- minimax value of the problem would be

$$r_\Gamma = \inf\{r_{21}, r_{22}, r_{23}\} = r_{21} = r_{22} = r_{23} = 0.0042886$$

and as such there is no Γ - minimax rule in this case since all the rules are equally associated with the Γ - minimax value.

In summary, table 7 shows that all the three Bayes estimators considered in this case (when the prior mean is more than the prior variance) are insensitive to changes in prior distribution. This suggests that whenever prior information indicates that the possible values of N are concentrated around a target value M, it does not really matter much which prior is used among the three truncated power series priors considered. [However, it can be observed that the Bayes estimator resulting from the truncated binomial prior, $\tilde{N}_{tb}$ has the smallest variance followed by the estimator resulting from the truncated poisson prior, $\tilde{N}_{tp}$].

Next we consider the case when the given prior variance is equal to the given prior mean.

Example (*ii*). [prior mean, M = prior variance, V =10, a = 1, b = 30]

Table 8. Matrix of Bayes Risks (r_{ij}) when prior mean is equal to prior variance

	Bayes Estimator ($\tilde{N}$)		
Prior	$\tilde{N}_{tb}$ 1	$\tilde{N}_{tnb}$ 2	$\tilde{N}_{tp}$ 3
1) π_{tb}	8.6890532	NA	8.6891371
2) π_{tnb}	NA	NA	NA
3) π_{tp}	8.7669663	NA	8.7668823

From table 8, ignoring the "NA" entries which indicate the "Not-Available" computational results in those cases due to stack overflow problems, we have the following results;

$$r_\Gamma(\tilde{N}_{tb}) = \sup_{\pi \in \Gamma} r(\pi, \tilde{N}_{tb}) = \sup\{r_{11}, r_{21}, r_{31}\} = r_{31} = 8.76697$$

$$r_\Gamma(\tilde{N}_{tnb}) = \sup_{\pi \in \Gamma} r(\pi, \tilde{N}_{tnb}) = \sup\{r_{12}, r_{22}, r_{32}\} = NA$$

$$r_\Gamma(\tilde{N}_{tp}) = \sup_{\pi \in \Gamma} r(\pi, \tilde{N}_{tp}) = \sup\{r_{13}, r_{23}, r_{33}\} = r_{33} = 8.76688$$

Γ - minimax value = $\inf\{r_{31}, r_{33}\} = r_{33}$

Thus, in summary, table 8 shows that all three Bayes estimators considered in this case (when the prior mean is equal to the prior variance) are almost insensitive to changes in prior distribution. Also (ignoring the NA cases), the Bayes estimator resulting from the truncated poisson prior, ($\tilde{N}_{tp}$), has the Γ - minimax value.

Although we are dealing with truncated distributions, this result is in line with general distribution theory, where it is a well established fact that the poisson distribution is characterized by the equality of its mean and variance. This result alongside the previous result (where the $\tilde{N}_{tb}$ had the smallest variance when the prior mean is greater than the prior variance) suggest that truncation may not have much effect since if in practice, one had prior knowledge of the magnitude of the mean relative to the variance, making a choice of prior distribution from the class of power series distributions could be guided by such target values.

Example(*iii*) [prior mean, $M = 3$, prior variance, $V = 4$, $a = 1$, $b = 20$]

Table 9. Matrix of Bayes Risks (r_{ij}) when prior mean is less than the prior variance

Bayes Estimator (Nbayes)

Prior	$\tilde{N}_{tb}$ 1	$\tilde{N}_{tnb}$ 2	$\tilde{N}_{tp}$ 3
1) π_{tb}	2.290738	2.887738	2.307912
2) π_{tnb}	3.171090	2.340207	2.973408
3) π_{tp}	2.675422	3.135520	2.657536

From table 9, we have the following results;

$$r_\Gamma(\tilde{N}_{tb}) = \sup_{\pi \in \Gamma} r(\pi, \tilde{N}_{tb}) = \sup\{r_{11}, r_{21}, r_{31}\} = r_{21} = 3.171090$$

$$r_\Gamma(\tilde{N}_{tnb}) = \sup_{\pi \in \Gamma} r(\pi, \tilde{N}_{tnb}) = \sup\{r_{12}, r_{22}, r_{32}\} = r_{32} = 3.135520$$

$$r_\Gamma(\tilde{N}_{tp}) = \sup_{\pi \in \Gamma} r(\pi, \tilde{N}_{tp}) = \sup\{r_{13}, r_{23}, r_{33}\} = r_{23} = 2.973408$$

From the above results it tu ns out in this case that

$$r_\Gamma(\tilde{N}_{tb}) \geq r_\Gamma(\tilde{N}_{tnb}) \geq r_\Gamma(\tilde{N}_{tp}).$$

Thus the Γ - minimax value of the problem is

$$r_\Gamma = \inf\{r_{21}, r_{32}, r_{23}\} = r_{23} = 2.973408$$

and as such the Γ - minimax rule in this case is $\tilde{N}_{tp}$ since it is the rule that is associated with the Γ - minimax value.

In summary, table 9 shows that all the three Bayes estimators considered in this case (when the prior mean is less than the prior variance) are sensitive to changes in prior distribution. This suggests that when prior information indicates that the possible values of N could be scattered about a target value M, one has to be careful in the choice of prior distribution. In this particular example, the tpnbayes should be preferable since it has the Γ - minimax value. Note also that a closer look at the entries of the matrix shows that the $\tilde{N}_{tb}$ and $\tilde{N}_{tp}$ are only slightly sensitive with respect to each other.

Another approach to sensitivity analysis which we consider here is one which uses Bayes' relative efficiency [see Chaubey and Li (1993)].

Definition: *We define the* ***Bayes's Relative Efficiency, (BRE)*** *of an estimator* $\tilde{N}^*$ *relative to the Bayes' estimator* $\tilde{N}$ *by*

$$BRE(\tilde{N}^*) = \frac{r(\pi, \tilde{N})}{r(\pi, \tilde{N}^*)}$$

where $r(\pi, \tilde{N}^*)$ *is the Bayes' risk of* $\tilde{N}^*$.

Using table 9 above, the computations in the following table have been obtained in order to display the values of BRE for the different estimators.

Table 10. Matrix of BREs when $M = 3 < V = 4$, $a = 1$, $b = 20$

	Estimator ($\tilde{N}^*$)		
Prior	$\tilde{N}_{tb}$ 1	$\tilde{N}_{tnb}$ 2	$\tilde{N}_{tp}$ 3
1) π_{tb}	1.0000	0.7933	0.9926
2) π_{tnb}	0.7380	1.0000	0.7871
3) π_{tp}	0.9933	0.8475	1.0000

From table 10, it can be observed that under a truncated binomial prior, the BRE of the estimator coming from the truncated binomial prior is about 80% while the estimator coming from the truncated poisson prior has a BRE of about 99%. This suggests that under a truncated binomial prior, $\tilde{N}_{tp}$ is almost as efficient as $\tilde{N}_{tb}$ while $\tilde{N}_{tnb}$ is about 20% less efficient than $\tilde{N}_{tb}$. Under a truncated negative binomial prior, $\tilde{N}_{tb}$ and $\tilde{N}_{tp}$ have about 74% and 79% BRE respectively, while under a truncated poisson prior, $\tilde{N}_{tb}$ and $\tilde{N}_{tnb}$ have about 99% and 84% BRE respectively.

Chapter 5 Further Considerations

5.1 Case of "equiprobable" prior

Here we consider the extreme situation in which no prior knowledge is available except that $N \in \{a, a+1, ..., b\}$. The prior distribution in this case can be expressed as

$$p(N) = \begin{cases} C, & N = a, a+1, \ldots b \\ 0, & elsewhere \end{cases} \tag{5.1}$$

where C is a real positive constant.

In this case, the joint probability mass function of D and N is

$$p(d,N) = p(N)p(d|N)$$

$$= C\frac{N!}{N^{d+1}(N-d)!}, \tag{5.2}$$

for $d = 1,2,...,N$ and $N = max\{a,d\},...,b$ since $N > a$ and $N > d$ implies that $N > max\{a,d\}$. The marginal probability mass function of D is

$$p(d) = C\sum_{N=\max\{a,d\}}^{b} \frac{N!}{N^{d+1}(N-d)!} \tag{5.3}$$

Hence the posterior distribution of N given d is

$$p(N|d) = \frac{\dfrac{N!}{N^{d+1}(N-d)!}}{\displaystyle\sum_{i=\max\{a,d\}}^{b} \frac{i!}{i^{d+1}(i-d)!}}, \tag{5.4}$$

for $N = max\{a,d\},...,b$ and $p(N|d) = 0$ elsewhere. The denominator in (5.4) can be written as a constant function of d, a and b, $g(d,a,b)$ say, which is free of N. Then, it can be observed that the posterior distribution in this case is proportional to the sample distribuion. That is

$$p(N|d) \propto g(d,a,b)p(d|N).$$

Thus the Bayes estimator under squared-error loss function which we shall hereafter refer to as the *equiprobable prior Nbayes* [$\tilde{N}_{ep}(a,b,d)$], can be obtained as the mean of the posterior distribution in (5.4) given by

$$\tilde{N}_{ep}(a,b,d) = \frac{\sum_{N=a}^{b} Np(d|N)}{\sum_{N=a}^{b} p(d|N)} \qquad (5.5)$$

5.1.1 Equivalence of $\tilde{N}_{ep}(a,b,d)$ to MLE

When the posterior distribution in (5.4) is unimodal and symmetric(see appendix), its mode, median and mean coincide. Under this condition, the Bayes' estimator using any of the three loss functions defined in chapter 3 is the same. In what follows, we show that using the "0-1" loss function, the Bayes' estimator in this case [$\tilde{N}_{ep}(a,b,d)$] is equivalent to the MLE estimator. obtained by Ahmad et.al (1995).

Proposition 5.1.1 *Under the "0-1" loss function, the Bayes' estimator, [$\tilde{N}_{ep}(a,b,d)$] resulting from an equiprobable prior is in a sense equivalent to the MLE estimator.*

Proof: *Under the "0-1" loss, the Bayes estimator is the mode of $P(N|d)$ which is the value of N that makes $P(N|d)$ largest. That is, $\tilde{N}_{ep}(a,b,d)$ is the value of N corresponding to*

$$\max_{N\in[a,b]} P(N|d) = \max_{N\in[a,b]} C^{*}\, P(d|N) \qquad (5.6)$$

where $C^{*} = [m(d)]^{-1} = \left[\sum_{i=a}^{b} \frac{i!}{i^{d+1}(i-d)!}\right]^{-1}$ *is a constant.*

From (5.6), it can be observed that we only need to find the value of N that maximizes $P(d|N)$ since C^{*} is free of N. Infact we can write

$$\max_{N \in [a,b]} P(N|d) = \max_{N \in [a,b]} C^{*} P(d|N)$$

$$= \max_{N \in [a,b]} P(d|N), \quad \text{because } C^{*} \text{ does not depend on } N.$$

$$\equiv MLE \text{ of } N \text{ on } [a,b].$$

5.2 Case of "difference" prior

One interesting feature of the Bayesian method(though a major source of controversy) is the liberty to choose a prior distribution. When this choice is properly made, the chosen prior(subjective or non-subjective) leads to a good estimator. Often the choice of a prior requires among other things, a great deal of intuition on the part of the statistician. The priors we considered above were guided by intuition, the idea being that since N is necessarily a positive integer and finite, the class of power series distributions stands out clearly as a simple and natural candidate among several other possible choices that could be made. Guided by the same intuition, we propose yet another class of priors which can incorporate a wider range of the different forms the quantity N may assume in real life. This class of priors which we shall hereafter refer to as the **"difference" prior** arises as the difference between two independently distributed (but not necessarily identically distributed) random variables from the same family of distributions. Here we illustrate the idea by considering the poisson family which as we saw in chapter 4 outperformed the binomial and negative binomial.

5.2.1 Poisson difference prior

Irwin(1937), Skellam(1946), Fisz(1953), Katti(1960), Strackee and van der Gon(1962) etc., all have considered the differences of two Poisson variables. Katti(1960) obtained the moments of the absolute difference and absolute deviation of Poisson, Pascal and binomial distributions. Fisz(1953) considered the limiting distribution of the difference of two independent poisson random variables and showed that it tends to the normal distribution as their respective means tend to infinity. Consul(1986) considered the differences of two generalized Poisson variates.

The difference of two independent Poisson random variables finds application quite often in risk analysis. Also Strackee and van der Gon(1962) observed that the number of light quanta emitted or absorbed in a definite time is distributed according to a Poisson distribution and as such they proposed that the physical limit of perceptible contrast in vision can be studied in terms of the difference between two independent variates each having a Poisson distribution.

Consider a situation where the number of elements that join a population, N_1 and the number that leave the same population, N_2 can both be viewed as random variables following independently Poisson distributions $p(N_1;\ \theta_1)$ and $p(N_2;\ \theta_2)$ respectively and let $N = N_1 - N_2 + N_0$ denote actual size of the population in a definite time, where N_0, a real positive constant denotes the initial size of the population. In view thereof, N can be studied in terms of the difference between two independent variates each having a Poisson distribution or some non-central version of such a difference according as N_0 is equal to zero or not. That is N can be viewed as a residual or net effect of an input-output type mechanism. This research will be taken up elsewhere. For curiosity, we give here the basic formulation for the situation where we assume $N_0 = 0$.

When $E[N_1] = \theta_1 \neq E[N_2] = \theta_2$, the distribution of N, for $N > 0$, [see Johnson and Kotz (1969) or Consul(1986)] is given by

$$p(N_1 - N_2 = N) = e^{-\theta_1 - \theta_2}\left(\frac{\theta_1}{\theta_2}\right)^{\frac{N}{2}} I_{\frac{N}{2}}\left(2\sqrt{\theta_1\theta_2}\right). \tag{5.7}$$

where $I_m(n)$ is the first kind of modified Bessel function of order m and argument n. Noting that N_2 can be expressed as a fractional multiple of N_1 such that $\theta_2 = \alpha\theta_1$ where $0 < \alpha < 1$, the distribution in (5.7) can be written as

$$P(\boldsymbol{N} = N) = e^{-\theta(1+\alpha)}\alpha^{-(N/2)} I_{\frac{N}{2}}(2\theta\sqrt{\alpha}) \tag{5.8}$$

Thus a truncated form of the distribution in (5.8) which could in a sense be viewed as an "objective" prior distribution can be defined as follows;

Definition 5.2.1: Truncated Poisson Difference prior [$\pi_{tpd}(N, \theta, \alpha, a, b)$]

$$\pi_{tpd}(N, \theta, \alpha, a, b) = \frac{\alpha^{-(N/2)} I_{\frac{N}{2}}(2\theta\sqrt{\alpha})}{\sum_{t=a}^{b} \alpha^{-(t/2)} I_{\frac{t}{2}}(2\theta\sqrt{\alpha})} \tag{5.9}$$

Using the sampling distribution of D given in (3.3), the joint distribution of N and D is given by

$$P(N, \theta, \alpha, a, b, d) = \frac{dN!}{N^{d+1}(N-d)!}\pi_{tpd}(N, \theta, \alpha, a, b) \tag{5.10}$$

Thus the posterior distribution corresponding to the prior in (5.9) will also be defined as

Definition 5.2.2: Truncated Poisson difference posterior [$\pi_{tpd}(N|d,\lambda,\ \alpha,\ a,b)$]

$$\pi_{tpd}(N|d,\theta,\alpha,a,b)=\frac{\dfrac{N!}{N^{d+1}(N-d)!}\alpha^{-(N/2)}I_{\frac{N}{2}}(2\sqrt{\theta})}{\displaystyle\sum_{i=\max\{a,d\}}^{b}\frac{i!}{i^{d+1}(i-d)!}\alpha^{-(i/2)}I_{\frac{i}{2}}(2\sqrt{\theta})}\tag{5.11}$$

Definition 5.2.3:

Truncated Poisson difference Bayes estimator[$\tilde{N}_{tpd}(\theta,\ \alpha,a,b,d)$]

$$\tilde{N}_{tpd}(\theta,\alpha,a,b,d)=\sum_{N=\max\{a,d\}}^{b}\pi_{tpd}(N|d\theta,\alpha,a,b)\tag{5.12}$$

Hence under the extensive form of analysis using a squared-error loss function, the Bayes risk associated with the Bayes estimator given in (4 12) is given by

$$r_{\tilde{N}_{tpd}}(\theta,\alpha,a,b)=\sum_{d=1}^{b}m(d)\left[\left(\sum_{N=\max\{a,d\}}^{b}N^{2}\pi_{tpd}(N|d,\theta,\alpha,a,b)\right)-\left(\tilde{N}_{tpd}(\theta,\alpha,a,b,d)\right)^{2}\right]\tag{5.13}$$

where $m(d)$ is the marginal distribution of D given by

$$m(d)=\sum_{N=\max\{a,d\}}^{b}P(N\lambda,\alpha,a,b,d).$$

5.3 Limit property of the Bayes' estimators

In this section, we review further elements of the Bayesian methodology in order to study some features of the Bayes' estimators(rules) developed in the preceeding discussions. First we record the following useful definitions which can be found in Ferguson (1967), pg. 49.

Definition: Limiting Bayes' rules

A rule δ *is said to be a limit of Bayes rules* δ_n, *if for almost all* d, $\delta_n(d)\rightarrow\delta(d)$ *in the sense of distributions.*

Definition: Generalized Bayes rule

A rule δ_o is said to be a generalized Bayes rule if there exists a measure on $\aleph$ such that $\sum L(N,\delta)p(d|N)\pi(N)$ takes on a finite minimum value when $\delta = \delta_o$.

Definition 5.4.3: Extended Bayes rule

A rule δ_o is said to be extended Bayes if δ_o is ε-Bayes for every $\varepsilon > 0$.
In other words δ_o is extended bayes if for every $\varepsilon > 0$ there is a prior π such that δ_o is ε-Bayes with respect to π; that is, $r(\pi,\delta_o) \leq \inf_\delta r(\pi,\delta) + \varepsilon$.

Remark: This notion of Extended Bayes is an easy one which can represent an "almost" Bayes rule situation.

Now consider the following representation of the power series distributions introduced by Jain and Consul (1971);

$$\pi_{gnb}(N;r,p,\beta) = \frac{r\,\Gamma(r+\beta N)}{N!\,\Gamma(r+\beta N-N+1)}\,p^{N}(1-p)^{r+\beta N-N}$$

where $0 < p < 1$, $|p\beta| < 1$ and the parameter β may not be an integer.

This is called the **generalized negative binomial** distribution. It can be observed that for $\beta = 0$ or 1, the above reduces to the binomial or negative binomial respectively. Also when r and β are large, while p is very small, such that $rp = \lambda_1$, $pb = \lambda_2$, where λ_1 is finite and positive while $|\lambda_2| < 1$, the generalized negative binomial distribution can be approximated by using James Stirling's formula on the two gamma functions and can be simplified to the following form called **generalized poisson** distribution;

$$p(N;\lambda_1,\lambda_2) = \begin{cases} \dfrac{\lambda_1(\lambda_2 + N\lambda_2)^{N-1}e^{-(\lambda_1+N\lambda_2)}}{N!}, & N = 0,1,2,\ldots \\ & \lambda_1 > 0,\ |\lambda_2| < 1 \\ 0 & \text{for} \quad N \geq m \text{ if } \lambda_1 + m\lambda_2 \leq 0. \end{cases}$$

Note that for $\lambda_2 = 0$, the generalized poisson reduces to the poisson distribution.

Consul and Jain(1971) studied the behaviour of this generalized form of poisson distribution and remarked that it can be applied to a wide variety of observed data.

The mean and variance of this distribution are given by

$$E(N) = \frac{\lambda_1}{1-\lambda_2} \quad and \quad Var(N) = \frac{\lambda_1}{(1-\lambda_2)^3} = \frac{E(N)}{(1-\lambda_2)^2}.$$

It follows that the mean will be smaller than, equal to, or greater than the variance according as the value of λ_2 is positive, zero, or negative. In what follows, we introduce a modified form of this distribution which we shall call a **modified generalized poisson**.

Modified generalized poisson.

Let α be a real non-negative constant. Consider

$$p(N:\lambda_1,\lambda_2) = \frac{\lambda_1(\lambda_1 + N\lambda_2)^{N-1} e^{-(\lambda_1+N\lambda_2)}}{N!} \frac{N^\alpha}{C(\lambda_1,\lambda_2,\alpha)}$$

where

$$C(\lambda_1,\lambda_2,\alpha) = \sum_{i=1}^{\infty} \frac{\lambda_1(\lambda_1 + i\lambda_2)^{i-1} e^{-(\lambda_1+i\lambda_2)}}{i!}$$

A truncated form of this modified distribution can be obtained as

$$P(N|a \le N \le b) = \frac{\dfrac{\lambda_1(\lambda_1 + N\lambda_2)^{N-1} e^{-(\lambda_1+N\lambda_2)}}{N!} \dfrac{N^\alpha}{C(\lambda_1,\lambda_2,\alpha)}}{\sum_{i=a}^{b} \dfrac{\lambda_1(\lambda_1 + i\lambda_2)^{i-1} e^{-(\lambda_1+i\lambda_2)}}{i!} \dfrac{i^\alpha}{C(\lambda_1,\lambda_2,\alpha)}}.$$

Using the sampling distribution of D given in (3.3), the posterior distribution of N given $D = d$ is

$$p(N|d) = \frac{\dfrac{\lambda_1(\lambda_1 + N\lambda_2)^{N-1} e^{-N\lambda_2}}{(N-d)!} N^{\alpha-(d+1)}}{\sum_{i=\max\{a,d\}}^{b} \dfrac{\lambda_1(\lambda_1 + i\lambda_2)^{i-1} e^{-i\lambda_2}}{(i-d)!} i^{\alpha-(d+1)}}$$

The Bayes' estimator under squared-error loss which is the posterior mean is then given by

$$E(N|d) = \frac{\displaystyle\sum_{N=\max\{a,d\}}^{b} \frac{\lambda_1(\lambda_1 + N\lambda_2)^{N-1} e^{-N\lambda_2}}{(N-d)!} N^{\alpha-d}}{\displaystyle\sum_{i=\max\{a,d\}}^{b} \frac{\lambda_1(\lambda_1 + i\lambda_2)^{i-1} e^{-i\lambda_2}}{(i-d)!} i^{\alpha-(d+1)}}$$

Now when $\lambda_2 = 0$, this posterior mean reduces to that of a modified poisson which for $a = 1$, $\alpha = d+1$, simplifies to

$$E(N|d) = \frac{\displaystyle\sum_{N=d}^{b} N \frac{\lambda_1^{N}}{(N-d)!}}{\displaystyle\sum_{i=d}^{b} \frac{\lambda_1^{i}}{(i-d)!}}$$

$$= \frac{\displaystyle\sum_{y=0}^{b-d} (y+d) \frac{\lambda_1^{y}}{y!}}{\displaystyle\sum_{y=0}^{b-d} \frac{\lambda_1^{y}}{y!}}$$

$$= \frac{\displaystyle d\sum_{y=0}^{b-d} \frac{\lambda_1^{y}}{y!}}{\displaystyle\sum_{y=0}^{b-d} \frac{\lambda_1^{y}}{y!}} + \frac{\displaystyle\sum_{y=0}^{b-d} y \frac{\lambda_1^{y}}{y!}}{\displaystyle\sum_{y=0}^{b-d} \frac{\lambda_1^{y}}{y!}}$$

$$= d + \frac{\displaystyle\lambda_1 \sum_{y=1}^{b-d} \frac{\lambda_1^{y-1}}{(y-1)!}}{\displaystyle\sum_{y=0}^{b-d} \frac{\lambda_1^{y}}{y!}}$$

Now as $b \to \infty$, $E(N|d) \to g(d, \lambda_1)$ where

$$g(d,\lambda_1) = d + \frac{\lambda_1 \sum_{y=1}^{\infty} \frac{\lambda_1^{y-1}}{(y-1)!}}{\sum_{y=0}^{\infty} \frac{\lambda_1^{y}}{y!}}$$

$$= d + \frac{\lambda_1 \left[\sum_{y=0}^{\infty} \frac{\lambda_1^{y}}{y!} - 1 \right]}{\sum_{y=0}^{\infty} \frac{\lambda_1^{y}}{y!}}$$

$$= d + \lambda_1 (1 - e^{-\lambda_1})$$

This limiting result suggests that when the upper truncation point, b is very large, the Bayes' estimate of N can be taken as the sum of the observed number of distinct units, d and some fractional multiple of the initial target value, λ_1. Note however that this result is subject to the condition that $\alpha = d+1$.

5.5 Discussion

The Bayesian estimators considered in this thesis provide a wider range of choices for estimating the size of a finite population, N than the classical methods of estimation. Moreso, the results obtained in subsection 5.1.1 of chapter 5 show that the classical MLE can even be considered a special case of the $\tilde{N}$. The sensitivity analysis carried out in chapter 4 show that the Bayes' estimator $\tilde{N}_{tp}$ (formed by using a truncated poisson distribution) outperformed the $\tilde{N}_{tb}$ and $\tilde{N}_{tnb}$ (formed from the truncated binomial and negative binomial distributions respectively). We propose extending this research to situations where; (*i*) N can be viewed as a residual or net effect of a random input-output mechanism in which case the truncated poisson difference prior introduced and defined in subsection 5.2.1 of chapter 5 could serve as an "objective" prior distribution. (*ii*) the observations can be viewed as time dependent data. Work in this direction is in progress and we will give the results elsewhere in further research.

APPENDIX A: INTERACTIVE PROGRAM IN Maplev3 FOR DETERMINING PRIOR PARAMETERS

Example (i): Determination of truncated binomial prior when given prior mean, M =13, prior variance, V = 1/9

Note: The following interactive program gives an idea of the method we used in order to determine the prior parameters for example (i). The same method is applicable to other cases.

Notations:

tbprior = truncated binomial prior p.m.f, M1 = Mean of tbprior
M2 = second raw moment of tbprior, g1 = M1-M, g2 = M2-V.

```
> tbprior :=(N,r1,p1,a,b)->binomial(r1,N)*p1^N*(1-p1)^(-N)/(sum(binomial(r1,N)*p1^N
> *(1-p1)^(-N),N=a..b));
```

$$tbprior := (N, r1, p1, a, b) \to \frac{\mathrm{binomial}(r1, N)\, p1^N (1 - p1)^{(-N)}}{\sum_{N=a}^{b} \mathrm{binomial}(r1, N)\, p1^N (1 - p1)^{(-N)}}$$

```
> M1 := (r1,p1,a,b)->sum(N*tbprior(N,r1,p1,a,b),N=a...b);
```

$$M1 := (r1, p1, a, b) \to \sum_{N=a}^{b} N\ \mathrm{tbprior}(N, r1, p1, a, b)$$

```
> M2 := (r1,p1,a,b)->sum(N^2*tbprior(N,r1,p1,a,b),N=a...b);
```

$$M2 := (r1, p1, a, b) \to \sum_{N=a}^{b} N^2\ \mathrm{tbprior}(N, r1, p1, a, b)$$

```
> g1 :=(r1,p1,a,b)->M1(r1,p1,a,b)-13;
```

$$g1 := (r1, p1, a, b) \to \mathrm{M1}(r1, p1, a, b) - 13$$

```
> g2 :=(r1,p1,a,b)->M2(r1,p1,a,b)-(13)^2-(1/9);
```

$$g2 := (r1, p1, a, b) \to \mathrm{M2}(r1, p1, a, b) - \frac{1522}{9}$$

for a fixed r1, we want p1 such that g1-g2=0 and g1=g2 ~ 0.

```
>
> plot({g1(25,p1,13,24),g2(25,p1,13,24)},p1=0..1);
>
> fsolve(g1(25,p1,13,24)-g2(25,p1,13,24)=0);
```

.004944704457

```
> evalf(g1(25,fsolve(g1(25,p1,13,24)-g2(25,p1,13,24)=0),13,24));
```

.00427231

```
> evalf(g2(25,fsolve(g1(25,p1,13,24)-g2(25,p1,13,24)=0),13,24));
```

.0042724

```
> for b from 14 by 1 to 24 do print(b,fsolve(g1(25,p1,13,b)-g2(25,p1,13,b)=0), evalf(g
> 1(25,fsolve(g1(25,p1,13,b)-g2(25,p1,13,b)=0),13,b)), evalf(g2(25,fsolve(g1(25,p1,13,
> b)-g2(25,p1,13,b)=0),13,b))) od;
```

14, .004982206406, .00427351, .0042735

15, -1.941844746, -.02547826, -.0254782

16, .004944705105, .00427230, .0042723

17, -12.86130913, -.06457055, -.0645705

18, .004944704457, .00427231, .0042724

22, .004944704457, .00427231, .0042724

23, .004944704457, .00427231, .0042724

24, .004944704457, .00427231, .0042724

```
> for r1 from 24 by 1 to 30 do print(r1,fsolve(g1(r1,p1,13,24)-g2(r1,p1,13,24)=0), eval
> f(g1(r1,fsolve(g1(r1,p1,13,24)-g2(r1,p1,13,24)=0),13,24)), evalf(g2(r1,fsolve(g1(r1,p
> 1,13,24)-g2(r1,p1,13,24)=0),13,24))) od;
```

24, .005392135632, .00427231, .0042723

25, .004944704457, .00427231, .0042724

29, .003712480530, .00427227, .0042722

30, .003494756555, .00427227, .0042723

```
> fsolve(g1(25,p1,13,24)-g2(25,p1,13,24)=0);
```

.004944704457

```
>  for a from 10 by 1 to 13 do print(a,fsolve(g1(25,p1,a,24)-g2(25,p1,a,24)=0),g1(25,fs
> olve(g1(25,p1,a,24)-g2(25,p1,a,24)=0),a,24),g2( 25,fsolve(g1(25,p1,a,24)-g2(25,p1,a
> ,24)=0),a,24)) od;
```

10, .4874777394, -.15939777, -.1593978

12, .3809775210, -.04890208, -.0489021

13, .004944704457, .00427231, .0042724

Thus the prior parameters in this case are [r1 = 25, p1 = 0.00495, a = 13, b =25]

Appendix B: INTERACTIVE PROGRAM IN Maplev3 FOR COMPUTING BAYES' ESTIMATORS AND RISKS

Example (i) : Computation of Bayes estimator and matrix of risks when given prior mean, M = 13, prior variance, V = 1/9

Notations: cspmf = p.m.f of D
- tbprior = truncated binomial prior.
- tbjoint = truncated binomial joint distribution of N and d.
- tbmarg = truncated binomial marginal distribution of d.
- tbpost = truncated binomial posterior distribution of N given d.
- tbNbayes = truncated binomial Bayes estimator.
- tbpostvar = truncated binomial posterior variance of N given d.

Note: The above system of notation for the truncated binomial prior is applicable to other priors.

```
> cspmf := (d,N)->d*N!/(N^(d+1)*(N-d)!);
```

$$cspmf := (d, N) \to \frac{d\, N!}{N^{(d+1)}\,(N-d)!}$$

```
> tbprior:=(N,r1,p1,a,b)->binomial(r1,N)*p1^N*(1-p1)^(-N)/sum(binomial(r1,i)*p1^i*(1-
> p1)^(-i),i=a..b);
```

$$tbprior := (N, r1, p1, a, b) \to \frac{\mathrm{binomial}(r1, N)\, p1^N\, (1 - p1)^{(-N)}}{\displaystyle\sum_{i=a}^{b} \mathrm{binomial}(r1, i)\, p1^i\, (1 - p1)^{(-i)}}$$

```
> tbjoint := (N,r1,p1,a,b,d)->cspmf(d,N)*tbprior(N,r1,p1,a,b);
```

$$tbjoint := (N, r1, p1, a, b, d) \to \mathrm{cspmf}(d, N)\, \mathrm{tbprior}(N, r1, p1, a, b)$$

```
> tbmarg := (r1,p1,a,b,d)->sum(tbjoint(N,r1,p1,a,b,d),N=max(a,d)..b);
```

$$tbmarg := (r1, p1, a, b, d) \to \sum_{N = \max(a, d)}^{b} \mathrm{tbjoint}(N, r1, p1, a, b, d)$$

```
> tbpost := (N,r1,p1,a,b,d)->binomial(r1,N)*p1^N*(1-p1)^(-N)*N!*(N^(d+1)*(N-d)!)^(-1)/
> sum(binomial(r1,i)*p1^i*(1-p1)^(-i)*i!*(i^(d+1)*(i-d)!)^(-1),i=max(a,d)..b);
```

$tbpost :=$

$$(N, r1, p1, a, b, d) \rightarrow \frac{\dfrac{\mathrm{binomial}(r1, N)\, p1^{N} (1-p1)^{(-N)} N!}{N^{(d+1)} (N-d)!}}{\displaystyle\sum_{i=\max(a,d)}^{b} \frac{\mathrm{binomial}(r1, i)\, p1^{i} (1-p1)^{(-i)} i!}{i^{(d+1)} (i-d)!}}$$

```
> tbNbayes:= (r1,p1,a,b,d)->sum(N*tbpost(N,r1,p1,a,b,d), N=max(a,d)..b);
```

$$tbNbayes := (r1, p1, a, b, d) \rightarrow \sum_{N=\max(a,d)}^{b} N\, \mathrm{tbpost}(N, r1, p1, a, b, d)$$

```
> tbpostvar :=(r1,p1,a,b,d)->(sum(N^2*tbpost(N,r1,p1,a,b,d), N=max(a,d)..b))-(tbNba
> yes(r1,p1,a,b,d))^2;
```

$tbpostvar :=$

$$(r1, p1, a, b, d) \rightarrow \sum_{N=\max(a,d)}^{b} N^2\, \mathrm{tbpost}(N, r1, p1, a, b, d) - \mathrm{tbNbayes}(r1, p1, a, b, d)^2$$

```
> r11:= (r1,p1,a,b)->sum(tbpostvar(r1,p1,a,b,d)*tbmarg(r1,p1,a,b,d),d=1..b);
```

$$r11 := (r1, p1, a, b) \rightarrow \sum_{d=1}^{b} \mathrm{tbpostvar}(r1, p1, a, b, d)\, \mathrm{tbmarg}(r1, p1, a, b, d)$$

```
> r11(25,0.0049447,13,24);
```

$$.004285026544$$

```
> r12 := (r1,p1,a,b,r2,p2)->sum(tbpostvar(r1,p1,a,b,d)*tbmarg(r1,p1,a,b,d),d=1..b) + s
> um((tbNbayes(r1,p1,a,b,d)-tnbNbayes(r2,p2,a,b,d))^2*tbmarg(r1,p1,a,b,d),d=1..b);
>
```

$$r12 := (r1, p1, a, b, r2, p2) \rightarrow \sum_{d=1}^{b} \mathrm{tbpostvar}(r1, p1, a, b, d)\, \mathrm{tbmarg}(r1, p1, a, b, d)$$
$$+ \sum_{d=1}^{b} (\mathrm{tbNbayes}(r1, p1, a, b, d) - \mathrm{tnbNbayes}(r2, p2, a, b, d))^2\, \mathrm{tbmarg}(r1, p1, a, b, d)$$

```
> r12(25,0.0049447,13,24,0.0015678);
```

$$.004285026544$$

```
> r13 := (r1,p1,a,b,f)->sum(tbpostvar(r1,p1,a,b,d)*tbmarg(r1,p1,a,b,d),d=1..b) + sum((
> tbNbayes(r1,p1,a,b,d)-tpNbayes(f,a,b,d))^2*tbmarg(r1,p1,a,b,d),d=1..b);
>
```

$$r13 := (r1, p1, a, b, f) \rightarrow \sum_{d=1}^{b} \mathrm{tbpostvar}(r1, p1, a, b, d)\,\mathrm{tbmarg}(r1, p1, a, b, d)$$

$$+ \sum_{d=1}^{b} (\mathrm{tbNbayes}(r1, p1, a, b, d) - \mathrm{tpNbayes}(f, a, b, d))^2\,\mathrm{tbmarg}(r1, p1, a, b, d)$$

> r13(25,0.0049447,13,24,0.05959);

$$.004285026544$$

> r21 := (r? n2,a,b,r1,p1)->sum(tnbpostvar(r2,p2,a,b,d)*tnbmarg(r2,p2,a,b,d),d=1..b)
> + sum .bNbayes(r2,p2,a,b,d)-tbNbayes(r1,p1,a,b,d))^2*tnbmarg(r2,p2,a,b,d),d=1.
> .b);

$$r21 := (r, p2, a, b, p1) \rightarrow \sum_{d=1}^{b} \mathrm{tnbpostvar}(r, p2, a, b, d)\,\mathrm{tnbmarg}(r, p2, a, b, d)$$

$$+ \sum_{d=1}^{b} (\mathrm{tnbNbayes}(r, p2, a, b, d) - \mathrm{tbNbayes}(r, p1, a, b, d))^2\,\mathrm{tnbmarg}(r, p2, a, b, d)$$

> r21(25,0.0015678,13,24,0.0049447);

$$.004288581889$$

> r22 := (r2,p2,a,b)->sum(tnbpostvar(r2,p2,a,b,d)*tnbmarg(r2,p2,a,b,d),d=1..b);

$$r22 := (r, p2, a, b) \rightarrow \sum_{d=1}^{b} \mathrm{tnbpostvar}(r, p2, a, b, d)\,\mathrm{tnbmarg}(r, p2, a, b, d)$$

> r22(25, 0.0015678,13,24);

$$.004288581889$$

> r23 := (r2,p2,a,b,f)->sum(tnbpostvar(r2,p2,a,b,d)*tnbmarg(r2,p2,a,b,d),d=1..b) + su
> m((tnbNbayes(r2,p2,a,b,d)-tpNbayes(f,a,b,d))^2*tnbmarg(r2,p2,a,b,d),d=1..b);
>

$$r23 := (r, p2, a, b, f) \rightarrow \sum_{d=1}^{b} \mathrm{tnbpostvar}(r, p2, a, b, d)\,\mathrm{tnbmarg}(r, p2, a, b, d)$$

$$+ \sum_{d=1}^{b} (\mathrm{tnbNbayes}(r, p2, a, b, d) - \mathrm{tpNbayes}(f, a, b, d))^2\,\mathrm{tnbmarg}(r, p2, a, b, d)$$

> r23(25,0.0015678,13,24,0.05959);

.004288581889

```
> r31:= (f,a,b,r1,p1)->sum(tppostvar(f,a,b,d)*tpmarg(f,a,b,d),d=1..b) + sum((tpNbaye
> s(f,a,b,d)-tbNbayes(r1,p1,a,b,d))^2*tpmarg(f,a,b,d),d=1..b);
>
```

$$r31 := (f, a, b, r, p1) \rightarrow \sum_{d=1}^{b} \mathrm{tppostvar}(f, a, b, d)\ \mathrm{tpmarg}(f, a, b, d) + \sum_{d=1}^{b} (\mathrm{tpNbayes}(f, a, b, d) - \mathrm{tbNbayes}(r, p1, a, b, d))^2\ \mathrm{tpmarg}(f, a, b, d)$$

```
> r31(0.05959,13,24,25,0.0049447);
```

.004287760055

```
> r32:= (f,a,b,r2,p2)->sum(tppostvar(f,a,b,d)*tpmarg(f,a,b,d),d=1..b) + sum((tpNbaye
> s(f,a,b,d)-tnbNbayes(r2,p2,a,b,d))^2*tpmarg(f,a,b,d),d=1..b);
>
```

$$r32 := (f, a, b, r2, p2) \rightarrow \sum_{d=1}^{b} \mathrm{tppostvar}(f, a, b, d)\ \mathrm{tpmarg}(f, a, b, d) + \sum_{d=1}^{b} (\mathrm{tpNbayes}(f, a, b, d) - \mathrm{tnbNbayes}(r2, p2, a, b, d))^2\ \mathrm{tpmarg}(f, a, b, d)$$

```
> r32(0.05959,13,24,25, 0.0015678);
>
```

.004287760055

```
> r33 := (f,a,b)->sum(sum((N-tpNbayes(f,a,b,d))^2*cspmf(d,N)*tpprior(N,f,a,b),N=ma
> x(a,d)..b),d=1..b);
```

$$r33 := (f, a, b) \rightarrow \sum_{d=1}^{b} \sum_{N=\max(a,d)}^{b} (N - \mathrm{tpNbayes}(f, a, b, d))^2\ \mathrm{cspmf}(d, N)\ \mathrm{tpprior}(N, f, a, b)$$

```
> r33(0.05959,13,24);
```

.004287760055

REFERENCES

1. Ahmad, M., Alalouf, S. and Chaubey,Y. (1995). A comparison of Three Sampling Strategies for Estimating Population Size. *Technical Report* # 249, UQAM, Montreal.

2. Ahmad, M., Alalouf, S. and Chaubey, Y. (1995a). Estimation of Population Total when Population Size is Unknown. *Technical Report* No. 4/95, Concordia University, Montreal, Canada.

3. Basu, D. (1958). On sampling with and without replacement. *Sankhya,* A, **19**, 287-294.

4. Bayes, T. (1763) "Essay towards solving a problem in the doctrine of chances." Philosophical Transactions. Royal Society, London 53: 370-418. (Reprinted in *Biometrika*, 1958, 45:293-315)

5. Berger, J. O. (1980). *Statistical decision theory: Foundations, Concepts, and Methods.* Springer-Verlag, New York.

6. Box, G. E. P., and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis* Addison-Wesley, Reading Mass.

7. Chapman D. G. (1954). The estimation of biological populations. *Ann. Math. Statist.* **25**, 1-15.

8. Chaubey, Y. and Li, W.(1993). Estimation of fraction defective through batch sampling. Proceedings of the section on Quality and Productivity of the american Statistical Association. p.198-206.

9. Consul, P. C.(1986). On the differences of two generalized poisson variates. *Commun. Statist. -Simula.*, **15(3)**, 761-767.

10. Consul, P. C. , and Jain, G. C. (1973). A generalization of the poisson distribution. *Technometrics*, **15** (4), 791-799.

11. Cormack, R. M. (1968). The statistics for capture-recapture methods. *Oceanogr. Mar. Bio. Ann. Rev.* **6**, 455-506.

12. Craig, C. C. (1953). Use of marked specimens in estimating populations of flying

insects. *Biometrika*, **40**,170-6.

13. Darroch, J. :I. (1958). The multiple-recapture census. 1. Estimation of a closed population. *Biometrika*, **45**, 343-59.
14. De Groot, M. (1970). *Optimal Statistical Decisions.* Mc-Graw Hill , New York .
15. Ewing, G. M. (1969). *Calculus of Variations with Applications.* Norton, New york.
16. Feller, W. (1957). *An introduction to probability theory and its applications,* Vol. 1, 2nd ed. John Wiley, New York.
17. Fisher, R. A. (1922). On the Mathematical Foundations of Theoretical Statistics. *Phil. Trans. Roy. Soc., Series* A **222**, 309.
18. Ferguson, T. S. (1967). *Mathematical statistics: A decision theoretic approach* Academic Press, New York.
19. Fisz, M. (1953). The limiting distribution of the difference of two of two Poisson random variables, *Zastosowania Matematyki*, **1**, 41-45. (In polish)
20. Harris, B. (1968). Statistical inference inthe classical occupancy problem unbiased estimation of the number of classes. *J. A. S. A.* **63**, 837-847.
21. Irwin, J. O. (1937). The frequency distribution of the difference between two independent variates following the same Poisson distribution, *Journ. of the Royal Statistical Society, Series* A, 100, 415-416.
22. Iversen, G. R (1984). *Bayesian Statistical Inference* . Sage University paper No.43, London
23. Jaynes, E. T. (1968). Prior Probabilities. *IEEE Transactions on Systems Science and Cybernetics*, **SSC-4**, 277-241.
24. Johnson, N., and Kotz, S. (1969). *Distributions in statistics: Discrete distributions* Houghton Mifflin, Boston.
25. Katti, S. K. (1960). Moments of the absolute difference and and absolute deviation of discrete distributions. *Ann. Math. Statist.* **31**, 78-85.
26. Lehmann, E.L. (1983). *Theory of Point Estimation.* Wiley, New York.

27. Lincoln, F. C. (1930). U.S. Dept. Agric. Circ. No. 118, 1-4.

28. MAPLE V3, Waterloo, Ontario, Canada.

29. Nayak. T. K. (1988). Estimating population size by recapture sampling. *Biometrika.*

30. Otis, D. L., Burnham, K. P., White, G. C. and Anderson, D. R. (1978). *Statistical Inference from capture data on closed animal populations.* Wildlife Monographs 62. Washington, D.C.: The Wildlife Society.

31. Robbins, H. E., (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35**, 1-20.

32. Schnabel, Z. E. (1938). The Estimation of the total fish population of a lake. *Amer. Math. Monthly*, **45**, 348-352.

33. Seber, G. A. F. (1973). *The estimation of Animal Abundance and Related Parameters.* London: Griffin.

34. Skellam, J. G. (1946). The frequency distribution of the difference between two Poisson variates belonging to different populations. *Journ. of the Royal Statistical Society, Series* A, **109**, 296.

35. Strackee, J. and van der Gon, J.J.D. (1962). The frequency distribution of the difference between two Poisson variates, *Statistica Neerlandica,* **16,** 17-23.