7-14-2017

On New Procedures of Estimation for Binary Data

Latia Carraway

ON NEW PROCEDURES OF ESTIMATION FOR BINARY DATA

by

Latia Carraway

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

Major: Mathematical Sciences

The University of Memphis

August 2017

Acknowledgments

First, I would like to thank God for giving me the strength to persevere. All the times I wanted to give up or could go no farther, God helped me to continue to the next step and not worry about the big picture. He closed doors that would have allowed me to quit, and opened new ones that allowed me to finish.

I have to thank my family. My mom and my husband for supporting me and encouraging me, always listening to me, and feeling my pain. They always made up for my shortcomings because I was working. I can never thank them enough for their help.

I want to thank Dr. Bowman. She went above and beyond to help me finish my degree. She worked weekends to accommodate my schedule because I have a job and small kids. She was kind, patient, and encouraging. I know there are not many people like her, and I am so thankful that God let me work with her.

Abstract

Carraway, Latia, Rachelle. PhD. The University of Memphis. August, 2017. On New Procedures of Estimation for Binary Data. Dale Bowman, PhD.

In developmental toxicity studies, current methods divide animals equally among all treatment groups. New procedures are introduced for estimating correlated binary data. Instead of allocating an equal number to each treatment, observe clusters one at a time until a desired number of clusters have a chosen number of responses or more. Dose levels, or treatments, known to have many responses would not need as many animals. This procedure could save animals but not sacrifice any information. Focusing on exchangeable binary data, a new procedure for estimating the probability of a response is investigated. This alternate design is analyzed through a simulation study and applied to a clinical data set. Comparisons are made between past estimators and the new estimator given.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Investigators of animal studies generally attempt to determine whether exposure to a chemical causes adverse effects in the experimental subjects. To do so, the investigator estimates the probabilities, $\theta_0, \ldots, \theta_g$, of positive response for each dose group, and a dose-response curve may be estimated to show the relationship between the dose level and the probability of an adverse effect. This information is used to determine safe exposure levels of the chemical for humans.

Human exposure typically occurs at low dose levels. The dose levels used in studies are usually higher than the expected levels for human exposure. Investigators must extrapolate from the observed risks at high doses in order to estimate risks at low doses which apply to humans (Chen and Kodell, 1989). In addition, two different distributions may fit the observed data adequately at the high doses but may give very different estimates at the lower end of the curve. Many animals are unnecessarily sacrificed at high doses, where response rates are high. This raises the question of animal allocation.

Further research needs to be done on the current method used to allocate the total number of animals among the different treatment groups. Finney (1964) reported the total number of animals used should be divided equally between all the doses used unless there are specific reasons not to do so. Animals would be saved if fewer were assigned to the high dose levels where more responses occur. Instead of assigning the same number of litters to each dose, a desired number of successes could be chosen and only the animals needed to reach the desired number of responses would need to be sacrificed.

There are many different types of animal studies. Research areas such as biomedical, pharmaceutical, and environmental research use animals in an effort to know more about safe exposure. In regular animal studies, each animal receives the toxin, and the response is measured on the animal. Each animal is independent so the binomial distribution can be used to count the number of positive responses. In developmental toxicity studies female animals are used to determine safe exposure to the offspring. Since the offspring are the individuals used to determine the response, a litter of responses are not independent but are correlated and respond similarly. According to the U.S. FDA (1993), a minimum number of 20 rats, mice, rabbits, or hamsters should be used in developmental toxicity testing. There should be at least three test groups and one control group. At the high dose, no more than 10% parental mortality, and at the lose dose, there should be no observable effects to the parent attributable to the test substance. The low dose should be set at a level which is expected to provide a margin of safety. The intermediate doses should be spaced to allow an arithmetic or geometric progression between the low and high doses. The addition of one or more groups is preferable to the use of large intervals between doses.

Currently, work is being done to develop alternatives to the use of animals. Two areas discussed by the Society of Toxicology are using cell or organ cultures rather than whole organisms and specific human gene mutations induced and quantified in cell culture. The Society of Toxicology claim, even with these other testing options, whole animal testing will still be needed in the future to validate the results of non-animal methods and as a last protective step before exposure of humans. Designs for developmental toxicity studies will require different analysis to reduce the number of animals needed.

Dette, Pepelyshev, and Wong (2009) investigated optimal designs for dose-finding experiments in toxicity studies. At the time of publication, there were few theoretical articles that used statistical principles to design toxicology studies. Since toxicology studies are expensive and labor intensive, Dette et al. (2009) claim an efficient design could significantly reduce the number of animals

needed. In their results, they found locally optimal designs and compare the results to uniform designs that are more commonly used. The conditional covariance in their work is incorrectly written, and it is incorrectly assumed that the litter size only depends on the dose level. George et al. (2016) give the correct covariance matrix for exchangeable multinomial data. With this contribution, the information matrix of the design could be corrected to include the correct conditional covariance matrix. Optimality criteria for designs involve the information matrix. This correction could improve procedures and give better results. Another approach to designs of this type could be using a negative binomial model instead of the traditional binomial model as discussed later.

Often data are observed in clusters where independence of response is not a valid assumption. Cluster sampling is useful when the population of interest has natural grouping or occur in natural clusters. Often government agencies or private research organizations will use cluster sampling by demographic regions. Cluster sampling involves randomly selecting clusters and using all individuals within those clusters for the sample. Similarly, nested designs are designs that have multiple observations within each object (or cluster). Nested designs occur in toxicology studies, neuroscience, biomedical research, when more than one observation is taken from an individual (person or animal). Nested designs can also occur in the social sciences (children nested in classes), behavioral genetics (relatives nested in families), and the field of medicine (patients nested in clinics) (Emmeke Aarts et al.,2014). A problem that arises in nested designs is that observations from the same object tend to be more similar than when taken from different objects. This results in a "within cluster" correlation that can make results misleading if not considered.

Another issue is sample size. In all statistical studies, the balance of having a large enough sample for reliable, meaningful results and minimizing the cost to obtain enough individuals is key to determining the sample size. To maximize the effect size, it is recommended to set the treatment as low as possible in the control group and high as ethically possible in the experimental group. This

advice is not easily followed when you think of testing on animals or in the social sciences when it can be difficult to find an intervention with any noticeable affect (Gelman, 2007). In the negative binomial distribution the sample size is the random variable. If clusters could be drawn until the desired number had enough responses for meaningful results, sample sizes could be minimized. In addition, cost could be minimized; yet, the same or more information could be attained. The recent advances and future advances of reducing sample sizes (animals used) in toxicology studies was discussed earlier, but the approach of sampling one cluster at a time has not been investigated. The nested design, or cluster data, discussed in this paper is the developmental toxicity study. Currently, pregnant females are equally divided among the doses and are randomly assigned to receive a toxic substance at varying dose levels. The females are sacrificed before their term, and the offspring are examined for binary responses, such as tumor or no tumor. The responses are recorded as Bernoulli random variables.

In the case of non-clustered data, $X_1, ..., X_n$, are independent Bernoulli trials, where the probability of success for each trial is μ. Define $R = \sum_{i=1}^{n} X_i$ as the number of successes in the n trials. Then the probability distribution function for R is $P(R = r) = \binom{n}{r}\mu^r(1-\mu)^{n-r}$ for $r = 0, 1, ..., n$. Under those conditions, R is a binomial random variable with $E(R) = n\mu$ and $Var(R) = n\mu(1-\mu)$. Instead of fixing the number of experimental units, n, we could fix the number of positive responses, r of interest. Then we would observe experimental units until there are r responses, making the number of trials Y the random variable of interest. Then, Y follows a negative binomial distribution with parameters r and μ, and

$$P(y|r,\mu) = P(Y = y) = \binom{y-1}{r-1}\mu^r(1-\mu)^{y-r},$$

where $y = r, r+1, \ldots$ gives the probability that exactly y trials are required to observe r success when the trials occur sequentially. Under those conditions, the expected number of trials is $E(Y) = \frac{r}{\mu}$, with variance, $Var(Y) = \frac{r}{\mu^2}$.

The binomial and negative binomial models just discussed are not

appropriate for developmental toxicity studies. The assumption of independence is needed, and it cannot be assumed that the offspring in a litter are independent. Offspring from the same litter tend to respond more similarly to a stimulus than fetuses from different litters. This within litter correlation causes over-dispersion, which means the variance of the responses is greater than the nominal variance. Recent statistical procedures account for this litter effect. Bowman and George (1995) introduce a non-parametric model called the exchangeable binary model where exchangeablility is assumed instead of independence.

Dose-response estimation is used to study the relationship between the dose of a toxic substance given and the probability of a response. It is often of interest to estimate the dose-response curve and the effective dose (ED_α). The effective dose is the dose level such that the proportion with an effect is α. Dose-response curves can be estimated using parametric models such as probit and logit functions (Prentice, 1976). Morgan (1992) provided a comprehensive review of parametric estimation methods. Non-parametric models were introduced to enhance the robustness of estimation. Mukhopadhyay (2000) developed a Bayesian nonparametric approach based on the Dirichlet process prior. Dette, Neumeyer, and Pilz (2005) constructed a nonparametric estimate of the quantile response curve and classical density curve. Dette and Scheder (2010) gave a finite comparison of nonparametric estimates of the effective dose in quantal bioassay. Yuan and Yin (2011) construct semi-parametric estimates of the dose-response curve to retain the advantages of parametric and non-parametric approaches.

Another important estimate to investigators is the no-observed-adverse-effect-level (NOAEL). This is the environmental dose level just below the lowest dose level with responses that are significantly different from the control. The EPA give guidelines, but this has been under scrutiny. Chen and Kodell (1989) proposed the benchmark dose (BD) estimated from a dose-response curve.

Estimating the probabilities, $\theta_0, \theta_1, \ldots, \theta_g$, of positive response is another important aspect of developmental toxicity studies. Investigators would want to

know for instance the probability of a positive response, θ_1, at each dose level. Bowman and George (1995) and Stefanescu and Turnbull (2003) discuss estimating the probability of a positive response and more for equal and random cluster sizes.

Chapter 2

Exchangeable Binomial Distribution

2.1 Introduction

For a nested design suppose there are g treatment groups. Clusters are assigned to each treatment such that there are m_i clusters in treatment i, $i = 1, \ldots, g$. The j^{th} cluster in the i^{th} dose group has n_{ij} individuals. Each individual is examined for a response. A success is recorded if the individual has the desired response (or response of interest), otherwise it is denoted a failure.

In a typical developmental toxicity experiment, pregnant females are randomly assigned to different treatment groups. The responses of the offspring from the female that has been exposed to a toxin are observed. A response is considered a success if there is an adverse affect observed such as death or malformation. The choice of dose groups start from a control group dose of 0 to the highest dose. Consider a developmental study with g dose groups, $d_1, d_2, ..., d_g$. Clusters are assigned to each dose group. The guidelines for the number of clusters assigned to a dose was discussed earlier. Each individual is examined for a response, and success is denoted $X_{ijk} = 1$. A failure is denoted $X_{ijk} = 0$.

$$X_{ijk} = \begin{cases} 1 & response(death, malformation, ...) \\ 0 & otherwise \end{cases}$$

where X_{ijk} is the response of the k^{th} fetus in the j^{th} litter of the i^{th} dose group

for $i = 1, ..., g$, $j = 1, ..., m_i$, $k = 1, ..., n_{ij}$, and n_{ij} is the size of the j^{th} litter of the i^{th} dose group. Further, denote the probability of success (i.e. probability of observing the response of interest) $P(X_{ijk} = 1) = \mu_i$ in the i^{th} dose group. Then, X_{ijk} is a Bernoulli random variable with probability of success μ_i.

Consider a single cluster of size n in a fixed dose group, $X_1, \ldots, X_n$ are the observations of the units within the cluster. Let $R = \sum_{k=1}^{n} X_k$ be the total number of successes in the n trials. Since $X_1, ..., X_n$ are not independent, let $\rho = corr(X_a, X_b)$, where $a \neq b$, be the within litter correlation. Then by definition, $\rho = corr(X_a, X_b) = \frac{cov(X_a, X_b)}{\sigma_{X_a}\sigma_{X_b}}$. It follows that
$Var(R) = Var(\sum_{k=1}^{n} X_k) = \sum_{k=1}^{n} Var(X_k) + 2\sum\sum_{a<b} cov(X_a, X_b) =$
$n\mu(1-\mu) + n(n-1)\rho\mu(1-\mu) = n\mu(1-\mu)(1+\rho(n-1))$. Observations in a litter tend to respond similarly, so ρ is expected to be positive. In this case, the correlation factor $[1 + \rho(n-1)]$ is greater than 1 and represents over-dispersion relative to the binomial model.

Bowman and George (1995) introduced a non-parametric model called the exchangeable binary model. This model is based on the assumption of exchangeability between litter mates. Although the litter-mates' responses are not independent, it may be reasonable to assume they are exchangeable. Exchangeability means for any different permutation of litter-mates the probability of responses stays the same. Suppose $X_1, X_2, ...$ is a finite or countable sequence of random variables. These variables are exchangeable if for any vector $(X_1, X_2, ...)$ and for any n, the
$P(X_{\pi(1)} = x_1, ..., X_{\pi(n)} = x_n) = P(X_1 = x_1, ...X_n = x_n)$ for any permutation $\pi(1), ..\pi(n)$ of $1, ...n$.

Although the definitions are the same for finite or infinite sequences, there are probabilistic differences in the properties of finite and infinite exchangeable sequences. These differences are important in applications to modeling data. One such difference involves the fundamental theorem of de Finetti, which states that given an infinite sequence of exchangeable binary random variables

$X_1, X_2, ...$ there exists a distribution function F on $[0,1]$ such that

$$P\Big(\sum_{k=1}^{n} X_k = r\Big) = \binom{n}{r} \int_0^1 u^m (1-u)^{n-m} dF(u).$$

This theorem is not necessarily true for a finite sequence of exchangeable random variables (Freedman and Diaconis, 1982). Therefore, George and Bowman (1995) gave the joint distribution of any finite set of exchangeable binary random variables in terms of the probability of similar response among members of a cluster.

Correlated binary data are commonly analyzed by modeling the marginal response by the beta-binomial (Williams, 1975,1987; Prentice, 1986) and quasi-likelihood techniques and generalized estimating equations for estimating the mean response (Zeger and Liang, 1986; Ryan, 1992). Generalized estimating equations model the mean and variance parameters and use working matrices to specify the third and fourth moments while ignoring higher moments (Bowman and George, 1995). Under the assumption of exchangeability, moments of all orders can be efficiently estimated. For correlated binary data such as observations from some familial studies, developmental toxicity experiments, and ophthalmologic clinical trials, the assumption that data from the same dam, individual, or cluster are exchangeable may be reasonable.

An important subclass of multivariate binary distributions is the family of exchangeable binary distributions. Let $X_1, X_2, ..., X_n$ be a set of exchangeable binary random variables, and let

$$\lambda_k = P(X_1 = 1, ..., X_k = 1) \tag{2.1}$$

where $\lambda_0 = 1$. Using inclusion and exclusion principles George and Bowman (1995) obtained $P(X_1 = x_1, ..., X_n = x_n) = \sum_{j=0}^{n-r} (-1)^j \binom{n-r}{j} \lambda_{r+j}$, where

$r = \sum_{k=1}^{n} x_k$. They also derive the following exchangeable binomial distribution,

$$P(R = r) = \binom{n}{r} \sum_{j=0}^{n-r} (-1)^j \binom{n-r}{j} \lambda_{r+j}, \tag{2.2}$$

where $R = \sum_{k=1}^{n} X_k$. The expected number of responses, $E(R) = n\lambda_1$, and the variance of R can be written as $Var(R) = n\lambda_1 + n(n-1)\lambda_2 - n^2\lambda_1^2$.

To define ρ_k, the correlation of order k, using the definition

$$(var(X_1))^{k/2}\rho_k = E(X_{i_n} - \mu), ..., E(X_{i_k} - \mu) = E(X_1 - \mu), ..., E(X_k - \mu), \tag{2.3}$$

where $E(X) = \mu = \lambda_1$. Bowman and George (1995) obtained

$$\rho_k = \frac{\sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} \lambda_1^{k-j} \lambda_j}{[\lambda_1(1-\lambda_1)]^{k/2}}. \tag{2.4}$$

For a set of exchangeable binary data, $X_1, ..., X_n$, higher moments may now be expressed in terms of λ_k's (Bowman and George, 1995).

2.2 Estimation with Exchangeable Binary Data with Equal Cluster Sizes

Bowman and George (1995) show how the MLE's of λ_k can be obtained and used to compute the MLE's of the joint probabilities, marginal means, moments, and correlations of all orders for data when the cluster sizes are the same. Let $(X_{j1}, ..., X_{jn})$ be independent vectors of binary random variables such that each of the vectors has a common exchangeable distribution where $j = 1, ..., m$ denotes the cluster. Let $R_j = \sum_{k=1}^{n} X_{jk}$. Then $(R_1, ..., R_m)$ is a random sample from a population with discrete probability function given by $P(R = r)$ from equation 2.2. Let A_k be the number of samples for which R_j equals k,

$k = 0, ..., n$ and $j = 1, ..., m$, and let

$$p_r = \sum_{k=0}^{n-r} (-1)^k \binom{n-r}{k} \lambda_{r+k} \tag{2.5}$$

$r = 0, 1, ..., n$. Then $P(R = r) = \binom{n}{r} p_r$ and $(A_0, A_1, ..., A_n)$ have a multinomial distribution with parameters $(n, p_0, \binom{n}{1} p_1, \binom{n}{2} p_2, ..., p_n)$. As a result, the MLE of p_r is given by

$$\hat{p}_r = \frac{A_r}{\left[\binom{n}{r} m\right]}. \tag{2.6}$$

An inversion of equation 2.5 gives

$$\lambda_\ell = \sum_{j=0}^{n-\ell} \binom{n-\ell}{j} p_{n-j} \tag{2.7}$$

and hence the MLE of λ_ℓ is given by

$$\hat{\lambda}_\ell = \frac{1}{m} \sum_{j=0}^{n-\ell} \binom{n-\ell}{j} \frac{A_{n-j}}{\binom{n}{j}}, \tag{2.8}$$

$\ell = 1, ..., n$. It is shown in Bowman and George (1995) that the $\hat{\lambda}_\ell$ are unbiased. Using the properties of transformations of estimates from a multinomial population described in Bickel and Doksum (1977) and Bowman and George (1995) show that the estimated variance of $\hat{\lambda}_\ell$ is given by

$$\hat{var}(\hat{\lambda}_\ell) = \frac{1}{m} \left\{ \sum_{i=\ell}^{n} \frac{A_i}{m} \frac{\binom{n-\ell}{n-i}}{\binom{n}{i}^2} - \left(\sum_{i=\ell}^{n} \frac{A_i}{m} \frac{\binom{n-\ell}{n-i}}{\binom{n}{i}^2} \right)^2 \right\}. \tag{2.9}$$

When all clusters are of common size, n, the likelihood for m clusters is proportional to the multinomial likelihood

$$\prod_{r=0}^{n} P(R = r)^{A_r} \tag{2.10}$$

where A_r is the number of litters containing exactly r successes for $(0 \leq r \leq n)$.

As an example of exchangeable data with equal cluster sizes, Bowman and

George (1995) discuss an application to a clinical trial that compares two antibiotics for ear infections in children. The data set is from a double-blind randomized clinical trial comparing cefaclor (CEF) and amoxicillian (AMO), used for the treatment of acute otitis media (OME). Seventy-five children have OME in both ears at the beginning of the study and are randomly assigned to a 14-day treatment of CEF or AMO. X_1 is defined to be 1 if the right ear is clear at the $14th$ day, 0 otherwise, and X_2 is similarly defined in terms of the left ear. It is discussed how exchangeability is appropriate for this data set, and MLE's of λ_1, λ_2, and ρ_2 are given.

2.3 Parameter Estimation for Data with Random Cluster Sizes

Bowman and George (1995) provide estimates for λ_ℓ in the case where cluster sizes are not equal, however, Xu and Prorok (2003) showed that these estimates are not maximum likelihood estimates. In most applications involving cluster sampling, the litter size is random. Consider a sample in which m litters are independently chosen, but of varying sizes, with the j^{th} litter having n_j observations, denoted $(X_{j1}, ..., X_{jn_j})$ for $j = 1, \ldots, m$. Let n denote the cluster size, and because we are only concerned with finite clusters, assume the maximum value for n is K. Clusters of different sizes can be viewed as coming from clusters of equal-size, K, but with $(K - n)$ observations missing at random; even if this is not the true nature of the missing values. To make inferences based on combined information from litters of varying sizes, the parameters will need to have the same meaning irrespective of the litter size. This is termed the interpretability assumption by Stefanescu and Turnbull (2003). Under interpretability, $\lambda_{r,n}$, the probability of observing r responses in a cluster of size n, is the same for all clusters $K \geq n$, in other words, $\lambda_{r,n} = \lambda_{r,K} = \lambda_r$. This assumption needs to be justified in any given application. Without it, it is difficult to combine the information from clusters of different sizes without

imposing some model assumption for the effect of cluster size on the joint distribution. Stefanescu and Turnbull (2003) give testing procedures for the interpretability assumption. In addition, Szabo and George (2010) and Pang and Kuk (2007) give the setting and testing procedure for the interpretability assumption, also called marginal compatibility.

Stefanescu and Turbull (2003) propose using the EM algorithm which takes natural advantage of the statistical structure of the problem. The Expectation-Maximization (EM) Algorithm is an approach to approximate the maximum likelihood estimates when some of the data are missing. Under the interpretability assumption, the likelihood of any particular litter of size n is the same as the likelihood of those outcomes arising from a larger cluster of size K, but with $K-n$ observations missing completely at random. Thus, it is natural to consider using the EM algorithm to obtain maximum likelihood estimates.

To use the EM algorithm, first show the likelihood of the complete data with clusters of equal size K. Define $A_{r,n}$ to be the number of clusters of size n with exactly r successes. The full data likelihood is

$$L = \prod_{n=1}^{K} \prod_{r=0}^{n} \left\{ \binom{n}{r} \sum_{j=0}^{n-r} (-1)^j \binom{n-r}{j} \lambda_{r+j} \right\}^{A_{n,r}}. \tag{2.11}$$

Let $q_{r,n} = P(R=r) = \binom{n}{r} \sum_{j=0}^{n-r} (-1)^j \binom{n-r}{j} \lambda_{r+j}$, then the complete data likelihood could be reduced to $L = \prod_{n=1}^{K} \prod_{r=0}^{n} {q_{r,n}}^{A_{r,n}}$. For a generic cluster of size n, with observations $(x_1, \ldots, x_n)$ and responses $\sum_{i=1}^{n} x_i = r$, the probability conditional on cluster size n is

$$P(x_1, ..., x_n | N = n) = \frac{q_{r,n}}{\binom{n}{r}} = \sum_{j=0}^{n-r} (-1)^j \binom{n-r}{j} \lambda_{r+j}. \tag{2.12}$$

For a cluster of size n where $n < K$, there are $K-n$ observations missing completely at random. If all K were obtained the maximum number of responses, T, would be greater than or equal to r. If we observe r responses in n observations, T could be any value from r to $K-(n-r)$. By induction on

$K-n$ missing observations, $P(x_1,...,x_n|N=n)$, equation 2.12, becomes

$$= \sum_{t=r}^{K-n+r} \binom{K-n}{t-r} \sum_{j=0}^{K-t} (-1)^j \binom{K-t}{j} \lambda_{t+j}.$$

Simplify back into terms of $q_{r,n}$,

$$= \sum_{t=r}^{K-n+r} \binom{K-n}{t-r} \frac{q_{t,K}}{\binom{K}{t}}, \tag{2.13}$$

which by exchangeability is the probability of observing $(x_1,...,x_n)$ in a cluster of size K, when responses $(x_{n+1},...,x_K)$ are missing completely at random. And, under the interpretability assumption, the likelihood of any particular cluster of size n is the same as the likelihood of those outcomes from a larger cluster of size K, but with $K-n$ missing. The incomplete-data likelihood is then,

$$L = \prod_{n=1}^{K} \prod_{r=0}^{n} \left\{ \sum_{t=r}^{K-n+r} q_{t,K} \frac{\binom{K-n}{t-r}}{\binom{K}{t}} \right\}^{A_{n,r}}. \tag{2.14}$$

Now start the algorithm with initial estimates for $P(T=t) = q_{t,K} = \frac{1}{K+1}$ for $t = 0,...,K$. Let $A_{t,K}$ be the number of clusters with exactly t successes if the missing data were present so that all clusters have common size K. The complete data log-likelihood is then given by
$l = log\left(\prod_{n=1}^{K} \prod_{r=0}^{n} {q_{r,n}}^{A_{r,n}}\right) = \sum_{t=0}^{K} A_{t,K} log(q_{t,K})$.

For any cluster j with $n < K$ observations, $R_j = X_{j1} + ... + X_{jn}$ is the incomplete data and $T_j = R_j + X_{j,r+1} + ... + X_{j,K}$ is the complete data. Then denote

$$\begin{aligned}
p_{t,r,n} &= P(T=t|R=r) \\
&= \frac{P(R=r|T=t)P(T=t)}{\sum_{t'} P(R=r|T=t')P(T=t')} \\
&= \frac{\left\{\binom{t}{r}\left(\binom{K-t}{n-r}\right)/\binom{K}{n}\right\} q_t}{\sum_{t'=r}^{K-n+r} \left\{\binom{t'}{r}\left(\binom{K-t'}{n-r}\right)/\binom{K}{n}\right\} q_{t'}} \\
&= \frac{\left\{\binom{t}{r}\left(\binom{K-t}{n-r}\right)\right\} q_t}{\sum_{t'=r}^{K-n+r} \left\{\binom{t'}{r}\left(\binom{K-t'}{n-r}\right)\right\} q_{t'}}
\end{aligned}$$

for $t = r, r+1, ..., K-n+r$, where we have used the exchangeability of the X's. Also, let $p_{K,r,t} = 1$ if $r = t$ and 0 otherwise. To estimate $A_{t,K}^{(i)}$, for iteration i, conditional on the observed data $\{R_j\}$

$$A_{t,K}^{(i)} = E(At, K|\{R_j\}) = \sum_{n=1}^{K} \sum_{r=max(0,t+n-k)}^{min(t,n)} A_{r,n} p_{t,r,n} \tag{2.15}$$

where $A_{r,n}$ is the number of clusters size n with r responses. Update the estimates of $q_{0,K}^{(i)}, q_{1,K}^{(i)}, ..., q_{K,K}^{(i)}$, with $q_t^{(i)} = \frac{A_{t,K}^{(i)}}{K}$ at each iteration, (i), until convergence. The MLE's of $\hat{\lambda}_1$, $\hat{\rho}_2$, and $\hat{E}(X_1, X_2)$ can then be estimated.

Chapter 3

Exchangeable Negative Binomial Distribution

Tan et al. (2010) introduce the exchangeable negative binomial distribution as another way to model count data. Poison regression is the standard method used to model count data. However, the Poisson distribution requires the equality of its mean and variance, an assumption which is rarely met in real data. What often happens is that the variance of data is larger than the mean which was discussed earlier as over-dispersion. The standard parametric model to account for Poisson over-dispersion is the negative binomial distribution, and negative binomial regression is finding increased use (Hilbe, 2007). The negative binomial random variable can be viewed as the count to get the desired number of successes in a series of independent and identically distributed Bernoulli trials. When independence can not be assumed, an assumption of exchangeability is often used as an alternative to independence. The exchangeable negative binomial distribution has many advantages over some existing models. In an exchangeable model, the joint distribution is expressed in terms of marginal probabilities. The correlations of all orders are given by these probabilities so that an exchangeable model incorporates higher order moments and makes the full use of the information in them (Tan et al., 2010).

The data, $X_1, X_2, ...$, is assumed to be a sequence of exchangeable Bernoulli random variables. Let r be the desired number of successes. Then, the probability that Y trials are needed to obtain r successes is given by Tan,

Rayner, Wang, and Peng 2010 to be

$$P(Y=y)=\binom{y-1}{r-1}\sum_{k=0}^{y-r}(-1)^k\binom{y-r}{k}\lambda_{r+k}, \tag{3.1}$$

where $y=r,r+1,....$ The exchangeable Bernoulli sequence $X_1,X_2,...$ may be unobservable, but we can observe Y, the number of trials to get the first r successes which is the same as observing the number of failures to get the first r successes. Let S be the number of failures, then $Y=S+r$. Equation 3.1 can be rewritten

$$P(S=s)=\binom{s+r-1}{r-1}\sum_{k=0}^{s}(-1)^k\binom{s}{k}\lambda_{r+k} \tag{3.2}$$

where $s=0,1,\ldots$ (Tan et al., 2010).

Tan et al. (2010) go on to justify the exchangeable negative binomial distribution as a probability distribution, derive the moment generating function, and derive the mean and variance. Let $Y\sim\mathbf{ENB}(\boldsymbol{\lambda},r)$ with $\boldsymbol{\lambda}=\{\lambda_r,\lambda_{r+1},...\}$. The moment generating function of Y by definition is given by

$$M_Y(t)=E(e^{tY})=\sum_{y=r}^{\infty}e^{tY}P(Y=y),\quad t\in\mathbf{N}.$$

Using the de Finetti theorem, Tan et al. (2010) obtain

$$P(Y=y)=\binom{y-1}{r-1}\int_0^1 u^r(1-u)^{y-r}dQ(u) \tag{3.3}$$

where $y=r,r+1,\ldots.$ Assuming convergence they substitute 3.3 into the moment generating function definition and swap the summation and integration. By the Taylor expansion of the infinite negative binomial series, $\sum_{k=0}^{\infty}\binom{r+k-1}{r-1}(1-u)^k=u^{-r}$ they obtain

$$M_Y(t)=e^{tr}\int_0^1 u^r[1-(1-u)e^t]^{-r}dQ(u). \tag{3.4}$$

Differentiating the moment generating function and evaluating it where t=0

yields

$$E(Y) = \frac{d}{dt} M_Y(t)\Big|_{t=0} = M_Y'(0) = r \int_0^1 u^{-1} dQ(u). \tag{3.5}$$

When $r = 1$, the Taylor series of negative binomial series,
$\sum_{k=0}^{\infty} \binom{r+k-1}{r-1}(1-u)^k = u^{-r}$, becomes $\sum_{k=0}^{\infty}(1-u)^k = u^{-1}$. We can then write equation 3.5 as

$$E(Y) = r \int_0^1 \sum_{k=0}^{\infty} (1-u)^k dQ(u). \tag{3.6}$$

If the sum and integral converge, we can interchange the integral and summation; changing the expected value to

$$E(Y) = r \sum_{k=0}^{\infty} \int_0^1 (1-u)^k dQ(u). \tag{3.7}$$

Now, using the Hausdorff theorem (Feller, 1971) which states

> *to every infinite sequence of exchangeable binary random variables* $X_1, X_2, ...$ *there corresponds a probability distribution* Q *concentrated on* $[0,1]$ *such that for* $y = \ell + 1, \ell + 2, ...$
>
> $$P(X_1 = 1, ..., X_\ell = 1, X_{\ell+1} = 0, ..., X_y = 0) = \int_0^1 u^\ell (1-u)^{y-\ell} dQ(u)$$

If $\ell = 0$ in the Hausdorff Theorem, then
$P(X_1 = 0, ..., X_k = 0) = \int_0^1 (1-u)^k dQ(u)$. This is the probability that the number of successes is zero, $P(R = 0)$, in the exchangeable binary model. According to equation 2.2 previously defined,
$P(R = 0) = P(X_1 = 0, ..., X_n = 0) = \sum_{j=0}^{n} (-1)^j \binom{n}{j} \lambda_j$. Hence, using Hausdorff Theorem with $\ell = 0$, the expected value from equation 3.7 becomes

$$E(Y) = r \sum_{n=1}^{\infty} \sum_{j=0}^{n} (-1)^j \binom{n}{j} \lambda_j. \tag{3.8}$$

In the same manner the variance of Y can be defined. By definition the

$Var(Y) = E(Y^2) - (E(Y))^2$. The $E(Y^2)$ can be found by

$$E(Y^2) = M_y''(0) = r(r+1)\int_0^1 u^{-2}dQ(u) - r\int_0^1 u^{-1}dQ(u).$$

.

It has already been shown that $E(Y) = r\int_0^1 u^{-1}dQ(u)$, so $E(Y^2) = r(r+1)\int_0^1 u^{-2}dQ(u) - E(Y)$. The only part of the equation not previously discussed is $r(r+1)\int_0^1 u^{-2}dQ(u)$. Using the Taylor series of negative binomial series, where $r = 2$,$u^{-2} = \sum_{k=0}^{\infty}(k+1)(1-u)^k$. This gives $r(r+1)\int_0^1 u^{-2}dQ(u) = r(r+1)\int_0^1 \sum_{k=0}^{\infty}(k+1)(1-u)^k dQ(u)$. Assuming the sum and integral converge and using Hausdorff theorem when $\ell = 0$, we get

$$\begin{aligned} E(Y^2) &= r(r+1)\sum_{k=0}^{\infty}(k+1)\sum_{j=0}^{k}(-1)^j\binom{k}{j}\lambda_j - E(Y) \\ &= r(r+1)\sum_{k=0}^{\infty}(k+1)\sum_{j=0}^{k}(-1)^j\binom{k}{j}\lambda_j - r\sum_{k=1}^{\infty}\sum_{j=0}^{k}(-1)^j\binom{k}{j}\lambda_j \\ &= r(r+1)\sum_{k=0}^{\infty}k\sum_{j=0}^{k}(-1)^j\binom{k}{j}\lambda_j + r^2\sum_{k=1}^{\infty}\sum_{j=0}^{k}(-1)^j\binom{k}{j}\lambda_j \end{aligned}$$

.

The variance of Y is then given by,

$$Var(Y) = r(r+1)\sum_{k=0}^{\infty}k\sum_{j=0}^{k}(-1)^j\binom{k}{j}\lambda_j + r^2\sum_{k=1}^{\infty}\sum_{j=0}^{k}(-1)^j\binom{k}{j}\lambda_j - \left(r\sum_{k=1}^{\infty}\sum_{j=0}^{k}(-1)^j\binom{k}{j}\lambda_j\right)^2. \tag{3.9}$$

.

Tan et al. (2010) apply their procedure to a clinical burn wound data set to show the benefits of the exchangeable negative binomial distribution.

Chapter 4

An Alternate Design for Estimation using Exchangeable Binary Data

4.1 Motivation

The use of animals in toxicity studies began in the 1920's when J.W. Trevan introduced the use of the 50% lethal dose ($LD50$). In the 1960's, regulatory agencies made mandatory the submission of toxicity profiles for investigating new drugs. In the 1980's OECD, Organization for Economic Co-operation and Development, and ICH, International Conference on Harmonization, brought out guidelines for toxicity testing of pharmaceutical substances (Parasuraman, 2011). As discussed in the introduction, animals will always be used in toxicology research because the experimental environment of a whole animal can not be reproduced in a laboratory. However, as more awareness surrounds animal testing, organizations such as the National Anti-Vivisection Society (NAVS) and People for the Ethical Treatment of Animals (PETA) continue to advocate for fewer or no animals to be used in testing. The Society for Toxicology discuss that other options are becoming available, but animals will always be needed as the last method of ensuring safety of certain toxins for humans. So a different approach may be needed from what Finney (1964) suggested, assigning the same number of animals to each dose group or treatment group.

One type of animal study is the developmental toxicity study. In these studies pregnant females are randomly assigned to different doses, low to high,

and fetuses are observed for an effect. Another approach could be to do testing one litter at a time, by assigning a litter to a treatment and observing the responses. If this cluster had the desired number of responses or more, you could consider that a success; if not it could be considered a failure. Assign another cluster to the treatment and observe the responses, assigning success or failure based on the number of responses. This process would continue until you reach a desired number of clusters with the number of responses or more of interest. So for doses with a large number of expected responses, such as $LD50$ previously discussed, fewer animals would be assigned because the desired number of success could be observed with fewer animals needed. Each cluster would be correlated, since each female would have offspring that respond similarly. The success or failure of each cluster (i.e. whether or not they had the desired number of responses) would be a Bernoulli random variable, and the total number of clusters needed would follow the negative binomial distribution. This process could be more time consuming, but if it reduced the number of animals needed and did not sacrifice information, it could be very valuable. The following discussion investigates this alternate design for estimation using exchangeable binary data.

4.2 Conditional Probability

It is of interest to see how many clusters of size n are needed to observe t clusters having r or more responses. Each cluster is examined, let

$$E_j = \begin{cases} 1 & if \ \# \ of \ responses \geq r. \\ 0 & otherwise \end{cases}$$

Stop examining clusters when there are t litters with r or more successes. Therefore, $\sum_{j=1}^{M_t} E_j = t$. Where M_t is the total number of clusters needed to get t successes. Since each cluster is independent, M_t would follow a negative binomial distribution.

Table 4.1 illustrates an example where $n = 5$ and $r = 2$. For cluster 1 there are $n = 5$ observations with $\sum_{k=1}^{n} x_k = r = 2$. Therefore that cluster is considered a success, $e_1 = 1$, since there were $r = 2$ or more responses. The same can be seen of cluster 2, with a total of $r = 3$ responses, $e_2 = 1$. Cluster 3 has $r = 1$ responses which is not the desired 2 or more, so cluster 3 is considered a failure, e_3=0. This continues for each cluster j until there are M_t clusters where the total successes, $\sum_{j=1}^{M_t} e_j = t$, is t.

Table 4.1: Motivating Example

Litter	$(X_1, X_2, X_3, X_4, X_5)$	E
1	$(0, 1, 0, 0, 1)$	1
2	$(1, 1, 0, 0, 1)$	1
3	$(0, 0, 1, 0, 0)$	0
...	...	...
M_t	$(1, 1, 1, 1, 0)$	1
		$\sum_{i=1}^{m_t} E_i = t$

Where $(X_1, ..., X_n)$ are exchangeable binary random variables within each cluster and $R = \sum_{i=1}^{n} X_i$. Then, R is an exchangeable binomial random variable with r success in n trials. The probability of having r responses in a litter of size n is

$$P(R = r) = \binom{n}{r} \sum_{j=0}^{n-r} (-1)^j \binom{n-r}{j} \lambda_{r+j} \tag{4.1}$$

as defined in Bowman and George (1995). E_i is a Bernoulli random variable with probability of success $p_{r,n}$

$$\begin{aligned} p_{r,n} &= P(R_i = r) + P(R_i = r+1) + ... + P(R_i = n) \\ &= \sum_{k=r}^{n} \binom{n}{k} \sum_{j=0}^{n-k} (-1)^j \binom{n-k}{j} \lambda_{k+j} \end{aligned} \tag{4.2}$$

.

As discussed earlier, M_t is the number of clusters needed to reach t successes; a success is defined as a cluster having r or more responses. The probability of having r or more responses was just given as $p_{r,n}$. The the

probability of $M_t = m_t$ is given by

$$P(M_t = m_t|p_{r,n}) = \binom{m_t - 1}{t - 1} p_{r,n}^t (1 - p_{r,n})^{m_t - t}. \tag{4.3}$$

Let A_0 be the number of litters with 0 responses, A_1 be the number of litters with 1 response, and so on up to A_n, the number of litters with n responses. The total number of litters $A_0 + A_1 + \cdots + A_n = M_t$. M_t is a negative binomial random variable. To use the multinomial distribution for the A_i's, we can condition on knowing the value of M_t, the total number of litters. Let P_0 be the probability of having 0 responses in a litter of size n, which is $P(R = 0)$ as defined in Equation 4.1. In the same manner,
$P_1 = P(R = 1) = \binom{n}{1} \sum_{j=0}^{n-1} (-1)^j \binom{n-1}{j} \lambda_{1+j}$. Similarly define P_2,...,P_n. Then, conditional on $M_t = m_t$, $(A_0, ..., A_n)$ follow a multinomial distribution with parameters $(m_t, P_0, P_1, ..., P_n)$. Therefore,

$$P(A_0 = a_0, A_1 = a_1, ..., A_n = a_n|M_t = m_t) = \frac{m_t!}{a_0!a_1! \dots a_n!} P_0^{a_0} \dots P_n^{a_n} \tag{4.4}$$

for $\sum_{k=1}^n P_k = 1$ and $\sum_{k=1}^n a_i = m_t$. The joint probability distribution of $\mathbf{A}$ and M_t is

$$P(\mathbf{A}|M_t = m_t)P(M_t = m_t) = \left(\frac{m_t!}{a_0!a_1! \dots a_n!} P_0^{a_0} \dots P_n^{a_n} \right) \left(\binom{m_t - 1}{t - 1} p_{r,n}^t (1 - p_{r,n})^{m_t - t} \right) \tag{4.5}$$

and the marginal of $\mathbf{A}$ is

$$\sum_{m_t=1}^{\infty} \left(\frac{m_t!}{a_0!a_1! \dots a_n!} P_0^{a_0} \dots P_n^{a_n} \right) \left(\binom{m_t - 1}{t - 1} p_{r,n}^t (1 - p_{r,n})^{m_t - t}. \right) \tag{4.6}$$

The conditional likelihood in (4.4) must be maximized subject to the constraint $\sum_{k=0}^n P_k = 1$. The conditional maximum likelihood estimates of the P_i's are found by taking the partial derivatives, set equal to 0, and solving subject to the

constraints listed above using LaGrange multipliers.

$$\frac{d\ell}{dP_i} = \frac{d}{dP_i}\left[log\frac{m_t!}{a_0!a_1!\cdots a_n!} + \sum_{i=0}^{n} a_i log P_i - \lambda(\sum_{i=0}^{n} P_i - 1)\right].$$

Then, the conditional MLE for P_i is,

$$\hat{P}_i = \frac{A_i}{m_t} \tag{4.7}$$

where $i = 0, ...n$.

To find the maximum likelihood estimate of λ_ℓ, equation (4.2) is inverted. Start with $p_{n,n} = \lambda_n$, then find $p_{n-1,n} = \binom{n}{n-1}[\lambda_{n-1} - \lambda_n] + \lambda_n$. Substitute $p_{n,n}$ into $p_{n-1,n}$ to get $p_{n-1,n} = \binom{n}{n-1}[\lambda_{n-1} - \lambda_n] + p_{n,n}$ and solve for λ_{n-1}.

$$\lambda_{n-1} = \frac{1}{\binom{n}{n-1}}[p_{n-1,n} - p_{n,n}] + p_{n,n}$$

Continue finding λ_{n-2} by expanding $p_{n-2,n}$,
$p_{n-2,n} = \binom{n}{n-2}[\lambda_{n-2} - 2\lambda_{n-1} + \lambda_n] + p_{n-1,n}$. Solve for λ_{n-2}, and substitute into the equation λ_{n-1} and λ_n. Simplify like terms to get

$$\lambda_{n-2} = \frac{1}{\binom{n}{n-2}}[p_{n-2,n} - p_{n-1,n}] + 2\left(\frac{1}{\binom{n}{n-1}}[p_{n-1,n} - p_{n,n}]\right) + p_{n,n}.$$

Find λ_{n-3} in the same manner,
$p_{n-3,n} = \binom{n}{n-3}[\lambda_{n-3} - 3\lambda_{n-2} + 3\lambda_{n-2} - \lambda_n] + p_{n-2,n}$. Solve for λ_{n-3}, substitute into the equation λ_{n-2}, λ_{n-1}, and λ_n; then simplify.

$$\lambda_{n-3} = \frac{1}{\binom{n}{n-3}}[p_{n-3,n} - p_{n-2,n}] + 3\left(\frac{1}{\binom{n}{n-2}}[p_{n-2,n} - p_{n-1,n}]\right) + 3\left(\frac{1}{\binom{n}{n-1}}[p_{n-1,n} - p_{n,n}]\right) + p_{n,n}.$$

Examining λ_{n-3}, λ_{n-2}, and λ_{n-1} the following equation can be created

$$\lambda_\ell = \sum_{j=0}^{n-\ell}\binom{n-\ell}{j}\frac{1}{\binom{n}{\ell+j}}[p_{\ell+j} - p_{\ell+j+1}], \tag{4.8}$$

where $p_{n+1} = 0$ since the sample size is n. By definition of equation (4.2),

$p_{0,n} = \sum_{k=0}^{n} P(R = k)$, $p_{1,n} = \sum_{k=1}^{n} P(R = k)$,
$p_{2,n} = \sum_{k=2}^{n} P(R = k)$,...,$p_{n,n} = \sum_{k=n}^{n} P(R = k)$. Then, $P_0 = p_{0,n} - p_{1,n}$,
$P_1 = p_{1,n} - p_{2,n}$, and so on up to P_n.

Then, λ_ℓ can be written

$$\lambda_\ell = \sum_{j=0}^{n} \binom{n-\ell}{j} \frac{1}{\binom{n}{\ell+j}} P_{\ell+j} \tag{4.9}$$

which corresponds to the unconditional estimates of Bowman and George, 1995. Using equations (4.7) and (4.9) the conditional MLE of λ_ℓ is given by

$$\hat{\lambda}_\ell = \sum_{j=0}^{n-\ell} \binom{n-\ell}{j} \frac{1}{\binom{n}{\ell+j}} \frac{A_{\ell+j}}{m_t} \tag{4.10}$$

With our conditional MLE of λ_ℓ we can now find an estimate for the correlation between litter-mates and estimate over-dispersion. Bowman and George (1995) defined ρ_k, the correlation up to order k. Using this the correlation between two litter-mates would be

$$\rho_2 = \frac{\lambda_2 - \lambda_1^2}{\lambda_1(1 - \lambda_1)}. \tag{4.11}$$

Using our conditional likelihood estimates $\hat{\lambda}_1$ and $\hat{\lambda}_2$, the estimate for ρ_2 is

$$\hat{\rho}_2 = \frac{\hat{\lambda}_2 - \hat{\lambda}_1^2}{\hat{\lambda}_1(1 - \hat{\lambda}_1)}. \tag{4.12}$$

The conditional variance of $\hat{\lambda}_1$, is given by

$$\begin{aligned} Var(\hat{\lambda}_\ell | M_t = m_t) &= \sum_{j=0}^{n-\ell} \frac{\binom{n-\ell}{j}^2}{\binom{n}{\ell+j}^2} \frac{1}{m_t^2} Var(A_{\ell+j}) + \sum_{j=0}^{n-\ell-1} \sum_{\substack{i=0 \\ i \neq j}}^{n-\ell} \frac{\binom{n-\ell}{j}}{\binom{n}{\ell+j}} \frac{\binom{n-\ell}{i}}{\binom{n}{\ell+i}} \frac{1}{m_t^2} Cov(A_{\ell+j}, A_{\ell+i}) \\ &= \sum_{j=0}^{n-\ell} \frac{\binom{n-\ell}{j}^2}{\binom{n}{\ell+j}^2} \frac{1}{m_t} P_{\ell+j}(1 - P_{\ell+j}) + \sum_{j=0}^{n-\ell-1} \sum_{\substack{i=0 \\ i \neq j}}^{n-\ell} \frac{\binom{n-\ell}{j}}{\binom{n}{\ell+j}} \frac{\binom{n-\ell}{i}}{\binom{n}{\ell+i}} \frac{1}{m_t^2} (-m_t P_{\ell+j} P_{\ell+i}). \end{aligned} \tag{4.13}$$

The estimated conditional variance of $\hat{\lambda}_1$ would then be

$$\hat{Var}(\hat{\lambda}_\ell|M_t = m_t) = \sum_{j=0}^{n-\ell} \frac{\binom{n-\ell}{j}^2}{\binom{n}{\ell+j}^2} \frac{1}{m_t^2} A_{\ell+j} \left(1 - \frac{A_{\ell+j}}{m_t}\right) + \sum_{j=0}^{n-\ell-1} \sum_{\substack{i=0 \\ i\neq j}}^{n-\ell} \frac{\binom{n-\ell}{j}}{\binom{n}{\ell+j}} \frac{\binom{n-\ell}{i}}{\binom{n}{\ell+i}} \frac{1}{m_t^3} (-A_{\ell+j} A_{\ell+i}). \tag{4.14}$$

4.3 Unconditional Expectation and Variance

The unconditional expectation of λ_ℓ is investigated for bias, and the unconditional variance of λ_ℓ is derived for comparison with other unconditional estimation procedures. The unconditional expectation of λ_ℓ is given by,

$$\begin{aligned}
E(\hat{\lambda}_\ell) &= E(E(\hat{\lambda}_\ell|M_t = m_t)) \\
&= E\left(E\left(\sum_{j=0}^{n-\ell} \binom{n-\ell}{j} \frac{1}{\binom{n}{\ell+j}} \frac{A_{\ell+j}}{M_t}\right)\right) \\
&= E\left(\sum_{j=0}^{n-\ell} \binom{n-\ell}{j} \frac{1}{\binom{n}{\ell+j}} \frac{E(A_{\ell+j})}{M_t}\right) \\
&= E\left(\sum_{j=0}^{n-\ell} \binom{n-\ell}{j} \frac{1}{\binom{n}{\ell+j}} P_{\ell+j}\right) \\
&= E(\lambda_\ell) \\
&= \lambda_\ell
\end{aligned} \tag{4.15}$$

.

This shows our estimator is an unbiased estimator.

The unconditional variance of $\hat{\lambda}_\ell$ is found from

$$Var(\hat{\lambda}_\ell) = E(Var(\hat{\lambda}_\ell|M_t = m_t)) + Var(E(\hat{\lambda}_\ell|M_t = m_t)).$$

As shown, $E(\hat{\lambda}_\ell | M_t = m_t)) = \lambda_\ell$. Therefore,

$$\begin{aligned}
Var(\hat{\lambda}_\ell) &= E(Var(\hat{\lambda}_\ell | M_t = m_t)) + Var(E(\hat{\lambda}_\ell | M_t = m_t)) \\
&= E\left(\sum_{j=0}^{n-\ell} \frac{\binom{n-\ell}{j}^2}{\binom{n}{\ell+j}^2} \frac{1}{M_t^2} Var(A_{\ell+j}) + \sum_{j=0}^{n-\ell-1} \sum_{\substack{i=0 \\ i\neq j}}^{n-\ell} \frac{\binom{n-\ell}{j}}{\binom{n}{\ell+j}} \frac{\binom{n-\ell}{i}}{\binom{n}{\ell+i}} \frac{1}{M_t^2} Cov(A_{\ell+j}, A_{\ell+i}) \right) + Var(\lambda_\ell) \\
&= E\left(\sum_{j=0}^{n-\ell} \frac{\binom{n-\ell}{j}^2}{\binom{n}{\ell+j}^2} \left[\frac{1}{M_t} P_{\ell+j}(1 - P_{\ell+j}) \right] + \sum_{j=0}^{n-\ell-1} \sum_{\substack{i=0 \\ i\neq j}}^{n-\ell} \frac{\binom{n-\ell}{j}}{\binom{n}{\ell+j}} \frac{\binom{n-\ell}{i}}{\binom{n}{\ell+i}} \frac{1}{M_t} (-P_{\ell+j} P_{\ell+i}) \right) + 0 \\
&= \sum_{j=0}^{n-\ell} \frac{\binom{n-\ell}{j}^2}{\binom{n}{\ell+j}^2} \left[E\left(\frac{1}{M_t} \right) P_{\ell+j}(1 - P_{\ell+j}) \right] + \sum_{j=0}^{n-\ell-1} \sum_{\substack{i=0 \\ i\neq j}}^{n-\ell} \frac{\binom{n-\ell}{j}}{\binom{n}{\ell+j}} \frac{\binom{n-\ell}{i}}{\binom{n}{\ell+i}} E\left(\frac{1}{M_t} \right) (-P_{\ell+j} P_{\ell+i}).
\end{aligned} \tag{4.16}$$

Using a Taylor series expansion, we can find an approximation for $E(\frac{1}{M_t})$. Let $f(x) = \frac{1}{x}$; using Taylor series expansion to the second moment about $x = a$ we have

$$\frac{1}{x} \approx \frac{1}{a} + \frac{1}{a^2}(x - a) + \frac{1}{a^3}(x - a)^2.$$

Evaluate at $a = E(x)$, then $\frac{1}{x}$ becomes

$$\frac{1}{x} \approx \frac{1}{E(x)} + \frac{1}{E(x)^2}(x - E(x)) + \frac{1}{E(x)^3}(x - E(x))^2$$

Taking the Expectation of both sides we get

$$\begin{aligned}
E\left(\frac{1}{x} \right) &\approx E\left(\frac{1}{E(x)} + \frac{1}{E(x)^2}(x - E(x)) + \frac{1}{E(x)^3}(x - E(x))^2 \right) \\
&\approx \frac{1}{E(x)} + \frac{1}{E(x)^3} Var(X).
\end{aligned} \tag{4.17}$$

For $X \sim NB(r, p)$, where $p = P(success)$, $E(X) = \frac{r}{p}$ and $Var(X) = \frac{r}{p^2}$. Thus we get

$$E\left(\frac{1}{X} \right) \approx \frac{p}{r} + \frac{p}{r^2}.$$

For $M_t \sim NB(t, p_{r,n})$, the $E\left(\frac{1}{M_t}\right) \approx \frac{p_{r,n}}{t} + \frac{p_{r,n}}{t^2}$. Therefore the $Var(\hat{\lambda}_\ell)$ becomes,

$$\begin{aligned} Var(\hat{\lambda}_\ell) &\approx \sum_{j=0}^{n-\ell} \frac{\binom{n-\ell}{j}^2}{\binom{n}{\ell+j}^2} \left(\frac{p_{r,n}}{t} + \frac{p_{r,n}}{t^2}\right) P_{\ell+j}(1 - P_{\ell+j}) \\ &\quad + \sum_{j=0}^{n-\ell-1} \sum_{\substack{i=0 \\ i\neq j}}^{n-\ell} \frac{\binom{n-\ell}{j}}{\binom{n}{\ell+j}} \frac{\binom{n-\ell}{i}}{\binom{n}{\ell+i}} \left(\frac{p_{r,n}}{t} + \frac{p_{r,n}}{t^2}\right) (-P_{\ell+j}P_{\ell+i}) \\ &\quad \approx \left(\frac{p_{r,n}}{t} + \frac{p_{r,n}}{t^2}\right) \left(\sum_{j=0}^{n-\ell} \frac{\binom{n-\ell}{j}^2}{\binom{n}{\ell+j}^2} [P_{\ell+j}(1 - P_{\ell+j})] + \sum_{j=0}^{n-\ell-1} \sum_{\substack{i=0 \\ i\neq j}}^{n-\ell} \frac{\binom{n-\ell}{j}}{\binom{n}{\ell+j}} \frac{\binom{n-\ell}{i}}{\binom{n}{\ell+i}} (-P_{\ell+j}P_{\ell+i}) \right), \end{aligned} \tag{4.18}$$

and we can estimate this by

$$\begin{aligned} &\hat{Var}(\hat{\lambda}_\ell) \\ &\approx \frac{\sum_{i=\ell}^{n} \frac{A_i}{m_t}}{t} + \frac{\sum_{i=\ell}^{n} \frac{A_i}{m_t}}{t^2} \left(\sum_{j=0}^{n-\ell} \frac{\binom{n-\ell}{j}^2}{\binom{n}{\ell+j}^2} \frac{A_{\ell+j}}{m_t} (1 - \frac{A_{\ell+j}}{m_t}) + \sum_{j=0}^{n-\ell-1} \sum_{\substack{i=0 \\ i\neq j}}^{n-\ell} \frac{\binom{n-\ell}{j}}{\binom{n}{\ell+j}} \frac{\binom{n-\ell}{i}}{\binom{n}{\ell+i}} (\frac{-A_{\ell+j}}{m_t} \frac{A_{\ell+i}}{m_t}) \right). \\ &\approx \frac{\sum_{i=\ell}^{n} \frac{A_i}{m_t}}{t} + \frac{\sum_{i=\ell}^{n} \frac{A_i}{m_t}}{t^2} \left(\sum_{j=0}^{n-\ell} \frac{\binom{n-\ell}{j}^2}{\binom{n}{\ell+j}^2} \frac{A_{\ell+j}}{m_t} (1 - \frac{A_{\ell+j}}{m_t}) + \frac{1}{m_t^2} \sum_{j=0}^{n-\ell-1} \sum_{\substack{i=0 \\ i\neq j}}^{n-\ell} \frac{\binom{n-\ell}{j}}{\binom{n}{\ell+j}} \frac{\binom{n-\ell}{i}}{\binom{n}{\ell+i}} (-A_{\ell+j}A_{\ell+i}) \right). \end{aligned} \tag{4.19}$$

Below in Tables 4.2 to 4.7 are comparisons for the conditional standard error of $\hat{\lambda}_1$ from sampling in the simulation study to the unconditional variance (unconditional standard error) defined here. Values for t were chosen based on the results of the simulation study; the simulation performed better when t and r were higher than the expected number of responses. Also, since the folded logistic model performed better overall, the folded logistic model is used here.

Table 4.2: $\lambda_1 = 0.05$, $n = 5$, $E(R) = 0.25$, $r = 1$

t	$mean(\hat{\lambda}_1)$	$Var(\hat{\lambda}_1 \mid M_t = m_t)$	$se(\hat{\lambda}_1 \mid M_t = m_t)$	$Var(\hat{\lambda}_1)$	$se(\hat{\lambda}_1)$
4	0.0608	0.0008	0.0289	0.0010	0.0323
5	0.0654	0.0007	0.0268	0.0009	0.0294
6	0.0676	0.0008	0.0289	0.0010	0.0312
7	0.0535	0.0004	0.0192	0.0004	0.0205
8	0.0617	0.0005	0.0227	0.0006	0.0241
9	0.0539	0.0003	0.0173	0.0003	0.0182
10	0.0767	0.0006	0.0239	0.0006	0.0251

Table 4.3: $\lambda_1 = 0.05$, $n = 10$, $E(R) = 0.5$, $r = 1$

t	$mean(\hat{\lambda}_1)$	$Var(\hat{\lambda}_1 \mid M_t = m_t)$	$se(\hat{\lambda}_1 \mid M_t = m_t)$	$Var(\hat{\lambda}_1)$	$se(\hat{\lambda}_1)$
4	0.0476	0.0004	0.0197	0.0005	0.0220
5	0.0562	0.0007	0.0265	0.0008	0.0291
6	0.0409	0.0002	0.0125	0.0002	0.0134
7	0.0623	0.0006	0.0250	0.0007	0.0267
8	0.0460	0.0002	0.0156	0.0003	0.0165
9	0.0463	0.0003	0.0162	0.0003	0.0171
10	0.0637	0.0005	0.0229	0.0006	0.0241

Table 4.4: $\lambda_1 = 0.30$, ,$n = 5$, $E(R) = 1.5$, $r = 2$

t	$mean(\hat{\lambda}_1)$	$Var(\hat{\lambda}_1 \mid M_t = m_t)$	$se(\hat{\lambda}_1 \mid M_t = m_t)$	$Var(\hat{\lambda}_1)$	$se(\hat{\lambda}_1)$
4	0.3622	0.0076	0.0873	0.0150	0.1224
5	0.2752	0.0054	0.0736	0.0147	0.1212
6	0.3471	0.0035	0.0592	0.0075	0.0867
7	0.3227	0.0035	0.0595	0.0065	0.0803
8	0.3378	0.0030	0.0545	0.0048	0.0693
9	0.3337	0.0041	0.0643	0.0076	0.0871
10	0.3136	0.0030	0.0547	0.0050	0.0709

Table 4.5: $\lambda_1 = 0.30$, $n = 10$, $E(R) = 3$, $r = 3$

t	$mean(\hat{\lambda}_1)$	$Var(\hat{\lambda}_1 \mid M_t = m_t)$	$se(\hat{\lambda}_1 \mid M_t = m_t)$	$Var(\hat{\lambda}_1)$	$se(\hat{\lambda}_1)$
4	0.3774	0.0094	0.0968	0.0160	0.1265
5	0.2630	0.0026	0.0505	0.0069	0.0831
6	0.3323	0.0073	0.0855	0.0132	0.1151
7	0.2642	0.0022	0.0465	0.0052	0.0722
8	0.2938	0.0027	0.0520	0.0056	0.0751
9	0.3664	0.0042	0.0645	0.0078	0.0880
10	0.2932	0.0023	0.0475	0.0043	0.0653

Table 4.6: $\lambda_1 = 0.85$, $n = 5$, $E(R) = 4.25$, $r = 5$

t	$mean(\hat{\lambda}_1)$	$Var(\hat{\lambda}_1 \mid M_t = m_t)$	$se(\hat{\lambda}_1 \mid M_t = m_t)$	$Var(\hat{\lambda}_1)$	$se(\hat{\lambda}_1)$
5	0.8650	0.0047	0.0689	0.0095	0.0973
6	0.8774	0.0017	0.0411	0.0038	0.0619
7	0.8172	0.0067	0.0821	0.0123	0.1110
8	0.8318	0.0019	0.0435	0.0046	0.0677
9	0.8427	0.0030	0.0551	0.0066	0.0815
10	0.8795	0.0011	0.0331	0.0023	0.0475

Table 4.7: $\lambda_1 = 0.85$, $n = 10$, $E(R) = 8.5$, $r = 9$

t	$mean(\hat{\lambda}_1)$	$Var(\hat{\lambda}_1 \mid M_t = m_t)$	$se(\hat{\lambda}_1 \mid M_t = m_t)$	$Var(\hat{\lambda}_1)$	$se(\hat{\lambda}_1)$
9	0.8409	0.0056	0.0749	0.0075	0.0867
10	0.8921	0.0011	0.0326	0.0017	0.0407

As expected the unconditional standard error and variance are higher than the conditional standard error and variance. The conditional standard error decreases as t increases. The unconditional standard error does not consistently decrease as t increases like the conditional standard error.

Chapter 5

Simulation Study and Application

5.1 Simulation Study

To investigate the properties of the conditional estimators $\hat{\lambda}_1$, $\hat{\lambda}_2$, and $\hat{\rho}_2$. A simulation study was conducted. The folded logistic model and power family model were used to generate the λ's to give a known value for comparison. For the folded logistic model, described in George and Bowman (1995), the follow function for λ_k is used.

$$\lambda_k(\beta) = \frac{2}{1 + (k+1)^{\beta}} \tag{5.1}$$

where $k \geq 0$ and $\beta > 0$. The power family model, described in Kuk (2004), allows us to use the following as a different function for λ_k.

$$\lambda_k = p^{k^{\gamma}} \tag{5.2}$$

with $0 \leq p, \gamma \leq 1$. The data, as previously described, involves clusters, $(X_1, ..., X_n)$, of binary random variables, and $R = \sum_{i=1}^{n} X_i$ is the sum of exchangeable binary random variables. The probability of having r responses in n exchangeable trials is given by $P(R = r) = \binom{n}{r} \sum_{j=0}^{n-r} (-1)^j \binom{n-r}{j} \lambda_{r+j}$. Where λ_{r+j} is generated from either the folded logistic model or the power family model. The cumulative distribution function, F, is

$$F(x) = \sum_{r=0}^{x} P(R = r)$$

.

The exchangeable responses within each cluster are generated by the following scheme. A uniform random variable is generated. According to what interval of the cumulative distribution function values, $F(0), F(1), \ldots, F(n)$, the uniform random number, u, falls in, the value of the number of responses in a litter of size n, is recorded. If $0 < u \leq F(0)$, then $r = 0$; if $F(0) < u \leq F(1)$, then $r = 1$ and so on. If r is greater than or equal to the fixed number of responses of interest, then $E = 1$, as described in chapter 4, otherwise $E = 0$. The process stops when $\sum_{i=1}^{M_t} E_i = t$ where t is the chosen number of clusters with r or more responses.

In developmental toxicity studies it is typical to have low probability of response at a low dose and high probability of response at a high dose. To pattern results like a typical study, three cases for λ_1 are investigated, Low : $\lambda_1 = 0.05$, Medium : $\lambda_1 = 0.30$, and High : $\lambda_1 = 0.85$. For each case, we want to investigate different values of r and t. Based on the expected number of responses, $E(R) = n\lambda_1$, values for r are chosen. Values from $2, ..., 10$ are investigated for possible t values. Table 5.1 shows all combinations of parameters used in the simulation study for both power family and folded logistic models.

Table 5.1: Table of parameter values investigated

λ_1	n	Expected # responses	r	t	Total # cases
0.05	5	0.25	1	$2, \ldots, 10$	9
0.05	10	0.5	1	$2, \ldots, 10$	9
0.30	5	1.5	1,2	$2, \ldots, 10$	18
0.30	10	3	2,3,4	$2, \ldots, 10$	27
0.85	5	4.25	4,5	$2, \ldots, 10$	18
0.85	10	8.5	8,9	$2, \ldots, 10$	18

For each of the 99 different scenarios, the following is done for 1000 iterations:

- Generate a data set of responses as described above until you get t litters with at least r successes.

- Once the data set is generated, re-sample from this data 1000 times. For each iteration, $\hat{\lambda}_1$, $\hat{\lambda}_2$, and $\hat{\rho}_2$ are calculated. Let $\hat{\lambda}_1^{(i)}$ and $\hat{\rho}_2^{(i)}$ be the estimates in the i^{th} iteration, $i = 1, \ldots, 1000$, where $\hat{\lambda}_\ell = \sum_{j=0}^{n-\ell} \binom{n-\ell}{j} \frac{1}{\binom{n}{\ell+j}} \frac{A_{\ell+j}}{m_t}$.
- $mean(\hat{\lambda}_1) = mean(\hat{\lambda}_1^{(1)}, \hat{\lambda}_1^{(2)}, \ldots, \hat{\lambda}_1^{(1000)})$ and
- $se(\hat{\lambda}_1) = \sqrt{Var(\hat{\lambda}_1^{(1)}, \hat{\lambda}_1^{(2)} \ldots, \hat{\lambda}_1^{(1000)})}$
- Compare $mean(\hat{\lambda}_1)$ to the true value of λ_1 used to simulate the data.
- Let $mean(\hat{\rho}_2) = mean(\hat{\rho}_2^{(1)}, \ldots, \hat{\rho}_2^{(1000)})$ and
- $se(\hat{\rho}_2) = \sqrt{Var(\hat{\rho}_2^{(1)}, \ldots, \hat{\rho}_2^{(1000)})}$
- Compare $mean(\hat{\rho}_2)$ to the actual value of ρ_2 for the simulated data.

5.1.1 Simulation Result

All the simulation results are listed at the end of section 5.1. Our estimator $\hat{\lambda}_1$, with a 95% confidence interval, almost always contained the value of λ_1. The simulation code had more consistent results close to parameter values at $n = 10$ than $n = 5$ and when the number of responses were low, $r = 1, 2, 3, 4$. When λ_1 was high, the value for t also needed to be high for better results. Overall, when data was generated from the folded logistic model my estimate for λ_1 performed better than the estimates from the power family model. There does not seem to be any consistent bias in estimating λ_1 as t changes. When t is above the expected value of R, our estimators, $\hat{\lambda}_1$ and $\hat{\rho}_2$, were closer to their parameter values. In general the average number of clusters needed to reach t litters with at least r successes increases as t increases, as expected. The distribution of M_t is almost always right skewed, but is less so when t is higher values. The histograms for M_t are roughly bell-shaped at $\lambda_1 = 0.30$ using the power family model.

For the value of the correlation within clusters, ρ_2, it is known to be notoriously more difficult to get good consistent estimates, than for λ_1. During the simulation study, some iterations would result in all responses being the

highest value, n, making the estimate for $\lambda_1 = 1$. This, in turn, caused our estimate for $\hat{\rho}_2$ to be NA or NAN because of division by 0. Since these occurrences were few in a sequence of 1000 iterations, those were omitted so values could be given for our estimate of $\hat{\rho}_2$.

5.1.1.1 Folded Logistic

Table 5.2 and Table 5.3 at the end of section 5.1 represent low probability of response, when $\lambda_1 = 0.05$. When $n = 10$, our estimator, $\hat{\lambda}_1$, is closer to the parameter value of λ_1, than when $n = 5$. The standard error for our estimator is also lower at $n = 10$ than $n = 5$. And, the estimate for ρ_2 is more consistent with smaller standard error, than at $n = 5$. The average of M_t's does not seem to increase with t as we would expect when response probability is small. In Figures 5.1 and 5.2, for $\lambda_1 = 0.05$, histograms for M_t are skewed right, but are more bell-shaped as t increases.

Tables 5.4 through 5.8 represent a middle probability of response, when $\lambda_1 = 0.30$. Again, when $n = 10$ the estimator seems to be more consistent at estimating the value of the parameter λ_1 with smaller standard error than when $n = 5$. When r was higher than the expected value, the estimate for ρ_2 seemed to be better, and when n was higher the standard error of ρ_2 was smaller. Average values for M_t increase as t increases, as expected. Standard error for M_t is smaller when values for the number of responses is smaller. Histograms when $\lambda_1 = 0.30$, Figures 5.3 through 5.7, are all right skewed, but less so when r is higher than the $E(R)$ for high values of t.

Tables 5.9 through 5.12 represent a high probability of response, when $\lambda_1 = 0.85$. When $n = 10$ our estimator $\hat{\lambda}_1$ is very close to the parameter value with smaller standard error than $n = 5$. In most cases ρ_2 is underestimated whether $n = 5$ or $n = 10$ when λ_1 is high. Average values for M_t again increase as t increases, and standard error for M_t remains small and about the same for all values of n and r. Histograms for M_t, Figures 5.8 through 5.11, are again right skewed, but less so at $n = 5$ and r greater than $E(R)$.

5.1.1.2 Power Family

When $\lambda_1 = 0.05$, Tables 5.13 and 5.14, estimates of ρ_2 severely under estimate the true parameter value for both cases $n = 5, 10$. The underestimation is decreased when $\lambda_1 = 0.30$ and $r = 2$ and $r = 3$. For both cases $n = 5, 10$, estimates for λ_1 are very close to the parameter value with small standard error. The average of M_t's does not seem to increase with t as we would expect when response probability is small. Figures 5.12 and 5.13 show histograms for M_t when $\lambda_1 = 0.05$, they are again right skewed but less so when t is high.

When $\lambda_1 = 0.30$ and $n = 5$, Tables 5.15 and 5.16, The estimates for λ_1 underestimate the true parameter values at $r = 1$. However, at $n = 5$, $r = 2$, estimates for λ_1 are close to the parameter value. The standard errors for $\hat{\lambda}_1$ are slightly higher when $n = 5$, $r = 2$. Estimates for ρ_2 are underestimated at $n = 5$, and standard errors for ρ_2 are about the same for both cases of $n = 5$. M_t increases as t increases for both cases when $n = 5$. The standard errors for M_t are lower at $r = 1$ but increase with t when $r = 2$. When n increases to 10, Tables 5.17 through 5.19 ($\lambda_1 = 0.30$), estimates for λ_1 are close to the parameter value for all cases $r = 2, 3, 4$ and standard errors are all small and seem to decrease with t. Estimates for ρ_2 underestimate the true value and have about the same standard error. Values for M_t increase as t increases but so does the standard error. For figures 5.14 through 5.18, the histograms for M_t are almost bell-shaped at high values for t.

When $\lambda_1 = 0.85$, Tables 5.20 through 5.23, estimates for λ_1 are better when r is higher than $E(R)$ for both $n = 5, 10$, and the standard error is small and about the same for all cases. Estimates for ρ_2 show no consistent bias. Average values for M_t increase as t increases in all cases and the standard error stays consistently small. Again in figures 5.19 through 5.22, the histograms for M_t are right skewed but less so at high values of t.

5.1.1.3 Folded Logistic compared to Power Family

At low values of λ_1, $\lambda_1 = 0.05$, the power family estimates of λ_1 are closer to the parameter value. The folded logistic model does a better job at estimating ρ_2 when $n = 10$; otherwise, they both underestimate ρ_2. When $\lambda_1 = 0.30$, the folded logistic model has estimates, $\hat{\lambda}_1$, very close to the parameter value and closer estimates of ρ_2 to the parameter value when n and r are high. The power family does not perform as well an underestimates λ_1 and ρ_2 more often than the folded logistic. At $\lambda_1 = 0.85$, the folded logistic model again does better at high values of n and r by have closer estimates of λ_1 and ρ_2 to the parameter values with smaller standard error. The power family has high standard error estimates for $\hat{\rho}_2$ and underestimates λ_1 more often than the folded logistic model. As stated earlier, overall the folded logistic model had more consistent estimates that were closer to the stated parameter values.

5.1.1.4 Tables of Results and Histograms of $M_t = m_t$

Simulation Results Folded Logistic

Table 5.2: $\lambda_1 = 0.05$,$\rho_2 = 0.0736$,$E(R) = 0.25$,$n = 5$,$r = 1$

t	$mean(\hat{\lambda}_1)$	$se(\hat{\lambda}_1)$	$mean\hat{\rho}_2$	$se(\hat{\rho}_2)$	$mean(m_t)$	$se(m_t)$
2	0.0798	0.0474	-0.0899	0.0603	6.9010	4.0089
3	0.0922	0.0529	0.0176	0.0879	11.0840	5.5273
4	0.0643	0.0328	0.1008	0.0617	22.9460	10.6915
5	0.1455	0.0470	-0.0822	0.0693	8.9980	2.6835
6	0.0594	0.0273	0.1515	0.1078	34.7910	12.8452
7	0.0498	0.0178	-0.0528	0.0203	31.4410	10.4045
8	0.0757	0.0280	0.1296	0.0833	34.9800	11.0419
9	0.0540	0.0165	-0.0574	0.0188	36.2310	10.3772
10	0.0476	0.0159	0.0971	0.0403	64.3290	18.4358

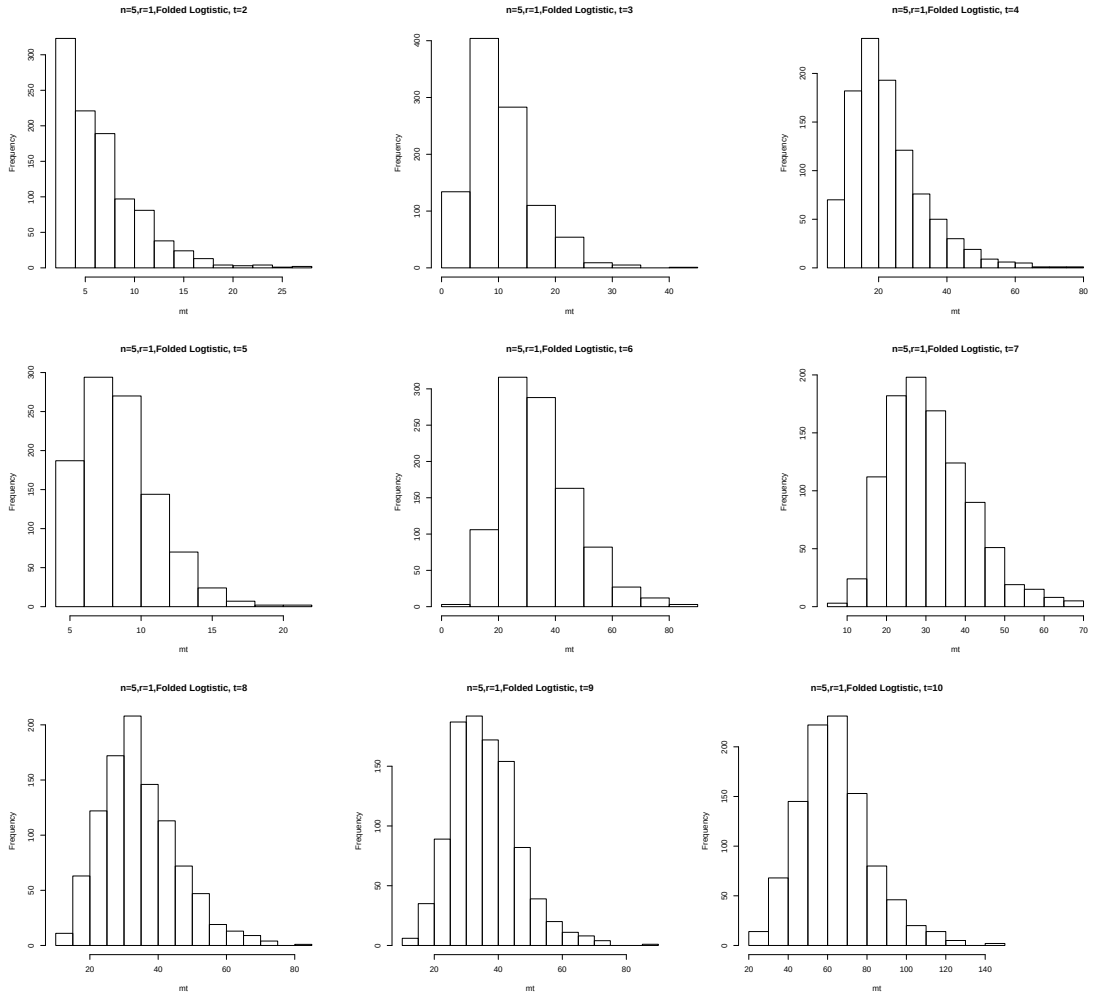

Figure 5.1: $\lambda_1 = 0.05$,$E(R) = 0.25$, $n = 5$, $r = 1$, $t = 2, \ldots, 10$

Table 5.3: $\lambda_1 = 0.05$,$\rho_2 = 0.0736$,$E(R) = 0.5$,$n = 10$,$r = 1$

t	$mean(\hat{\lambda}_1)$	$se(\hat{\lambda}_1)$	$mean\hat{\rho}_2$	$se(\hat{\rho}_2)$	$mean(m_t)$	$se(m_t)$
2	0.0346	0.0290	0.0277	0.0461	13.1150	8.2835
3	0.0278	0.0158	-0.0289	0.0173	13.9680	7.0258
4	0.0387	0.0232	0.0588	0.0608	19.1550	8.4589
5	0.0574	0.0247	0.0749	0.0359	20.1430	7.7641
6	0.0593	0.0244	0.0325	0.0452	17.0370	5.7289
7	0.0486	0.0193	0.0345	0.0421	22.8770	7.0985
8	0.0475	0.0163	0.0270	0.0389	25.0330	7.0603
9	0.0487	0.0175	0.0691	0.0580	31.0620	8.8748
10	0.0379	0.0122	0.0608	0.0370	42.4820	11.9202

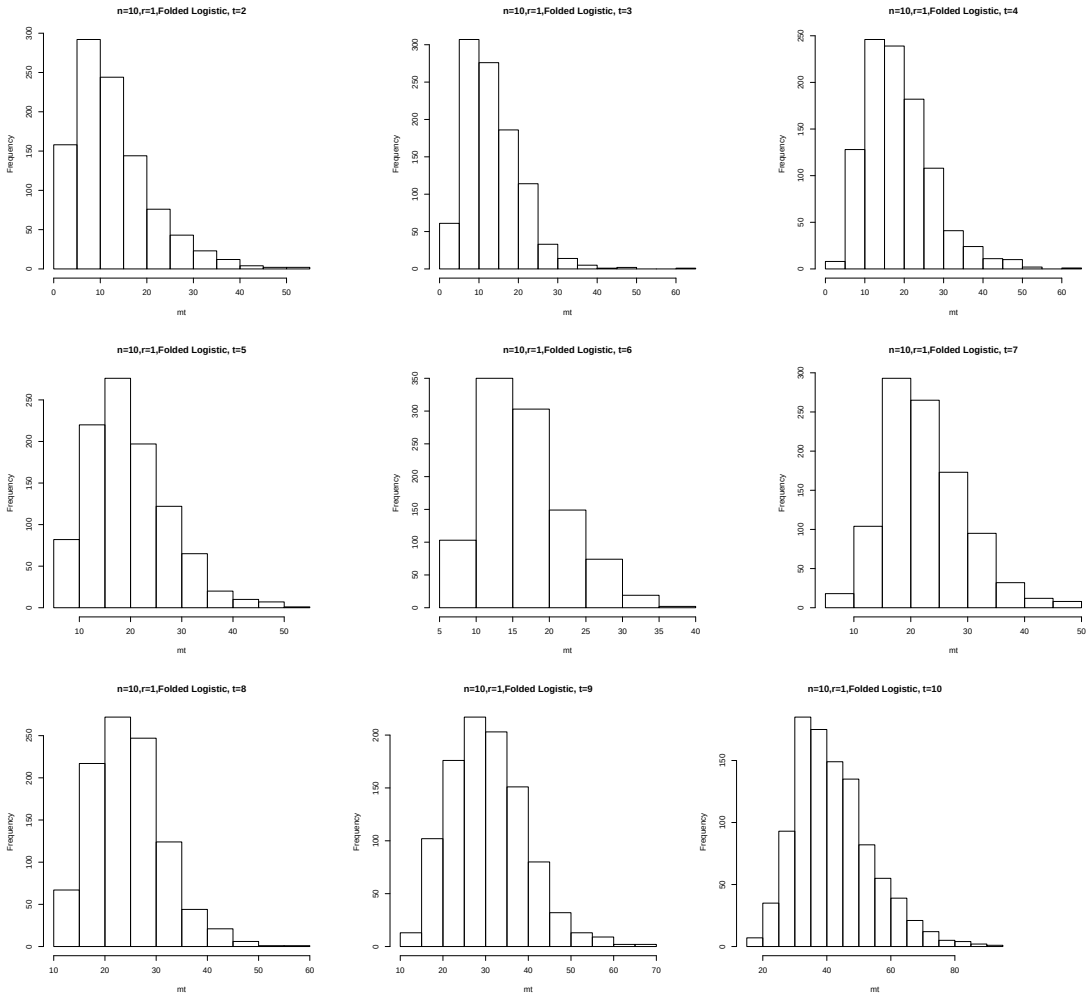

Figure 5.2: $\lambda_1 = 0.05$,$E(R) = 0.5$, $n = 10$, $r = 1$, $t = 2, \ldots, 10$

Table 5.4: $\lambda_1 = 0.30$,$\rho_2 = 0.1441$,$E(R) = 1.5$,$n = 5$,$r = 1$

t	$mean(\hat{\lambda}_1)$	$se(\hat{\lambda}_1)$	$mean\hat{\rho}_2$	$se(\hat{\rho}_2)$	$mean(m_t)$	$se(m_t)$
2	0.1821	0.0905	-0.0529	0.1199	4.0710	2.1261
3	0.3188	0.1084	-0.0314	0.1423	4.0140	1.1922
4	0.4490	0.0440	-0.2225	0.0202	4.0000	0.0000
5	0.3610	0.1229	0.2337	0.1620	7.0660	1.6954
6	0.3048	0.1045	0.2899	0.2276	9.0050	2.0802
7	0.1978	0.0590	0.0232	0.0903	11.8980	3.0031
8	0.2240	0.0460	-0.0815	0.0577	11.0690	2.0441
9	0.3172	0.0644	-0.0016	0.1144	11.0360	1.6191
10	0.3834	0.0752	0.0507	0.0992	11.0580	1.1008

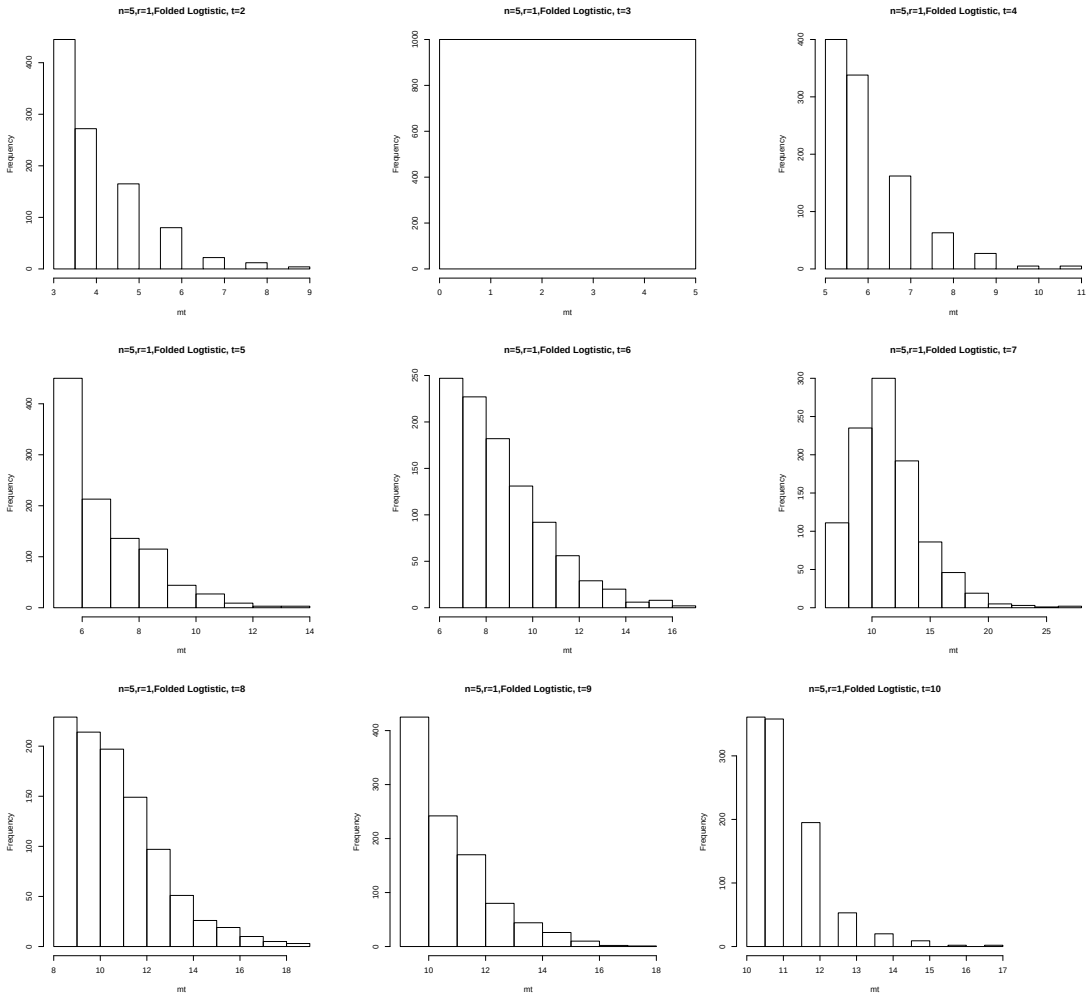

Figure 5.3: $\lambda_1 = 0.30$,$E(R) = 1.5$, $n = 5$, $r = 1$, $t = 2, \ldots, 10$

Table 5.5: $\lambda_1 = 0.30$,$\rho_2 = 0.1441$,$E(R) = 1.5$,$n = 5$,$r = 2$

t	$mean(\hat{\lambda}_1)$	$se(\hat{\lambda}_1)$	$mean\hat{\rho}_2$	$se(\hat{\rho}_2)$	$mean(m_t)$	$se(m_t)$
2	0.3405	0.1306	0.0398	0.1591	6.9340	4.2225
3	0.2255	0.0657	-0.0678	0.0789	9.0400	4.4947
4	0.2412	0.0767	0.0847	0.0952	12.0390	5.0322
5	0.3245	0.1001	0.3257	0.1656	12.0760	4.0421
6	0.3201	0.0736	0.1675	0.1418	15.9010	5.2649
7	0.2401	0.0711	0.2036	0.1035	17.9900	5.4525
8	0.3323	0.0623	0.1369	0.0845	20.0590	5.3054
9	0.3519	0.0709	0.2232	0.1188	18.7900	4.6006
10	0.4121	0.0674	0.1254	0.0994	19.0490	4.3083

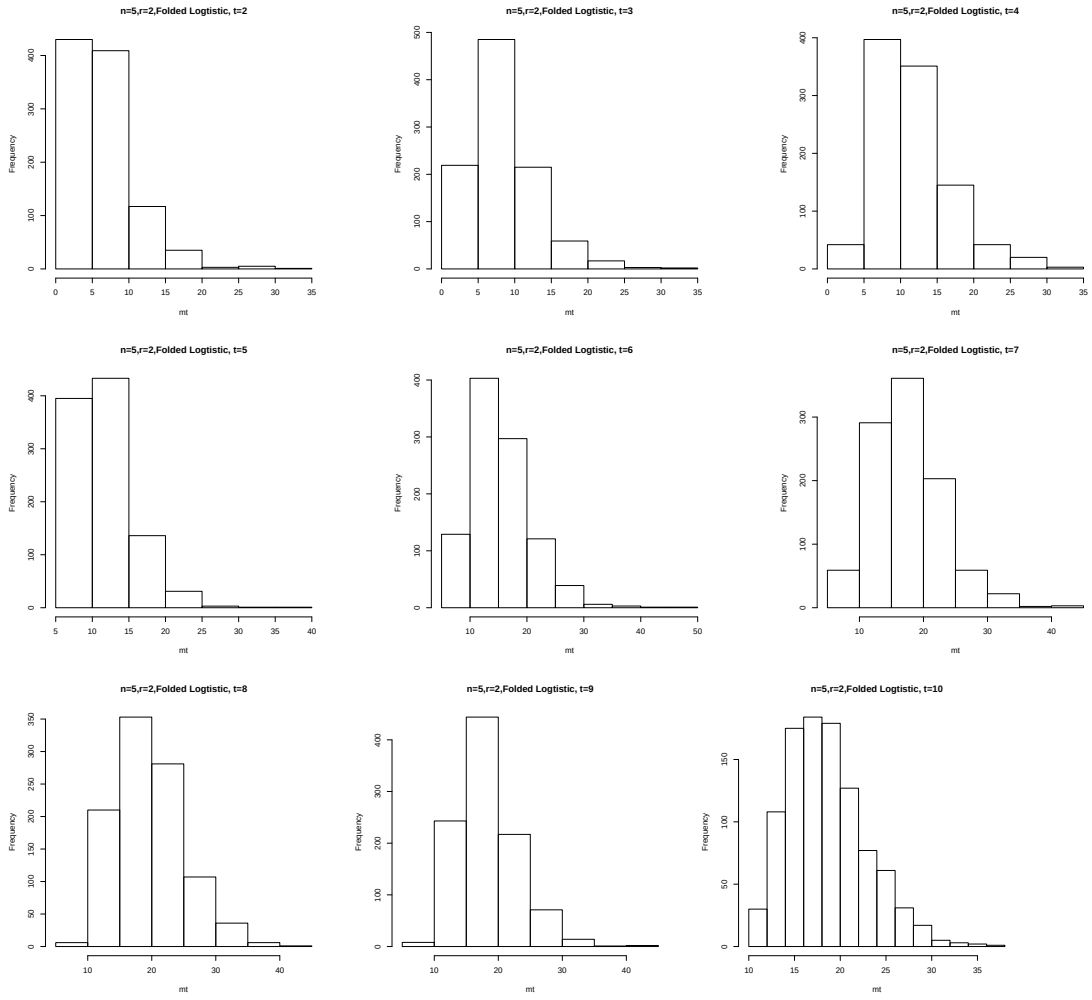

Figure 5.4: $\lambda_1 = 0.30$,$E(R) = 1.5$, $n = 5$, $r = 2$, $t = 2, \ldots, 10$

Table 5.6: $\lambda_1 = 0.30$,$\rho_2 = 0.1441$,$E(R) = 3$,$n = 10$,$r = 2$

t	$mean(\hat{\lambda}_1)$	$se(\hat{\lambda}_1)$	$mean\hat{\rho}_2$	$se(\hat{\rho}_2)$	$mean(m_t)$	$se(m_t)$
2	0.1717	0.0570	-0.0294	0.0587	4.0670	2.0408
3	0.2365	0.0797	0.0403	0.0833	5.1640	1.8763
4	0.2744	0.0690	0.0066	0.0798	4.9620	1.0981
5	0.2369	0.0544	0.0045	0.0552	7.9810	2.2382
6	0.2391	0.0570	0.0895	0.0545	11.0260	3.0260
7	0.2555	0.0433	0.0208	0.0298	13.0420	3.3023
8	0.2942	0.0614	0.0633	0.1054	10.0650	1.6839
9	0.2774	0.0469	0.0266	0.0526	11.8680	1.9719
10	0.2774	0.0469	0.0266	0.0526	11.8680	1.9719

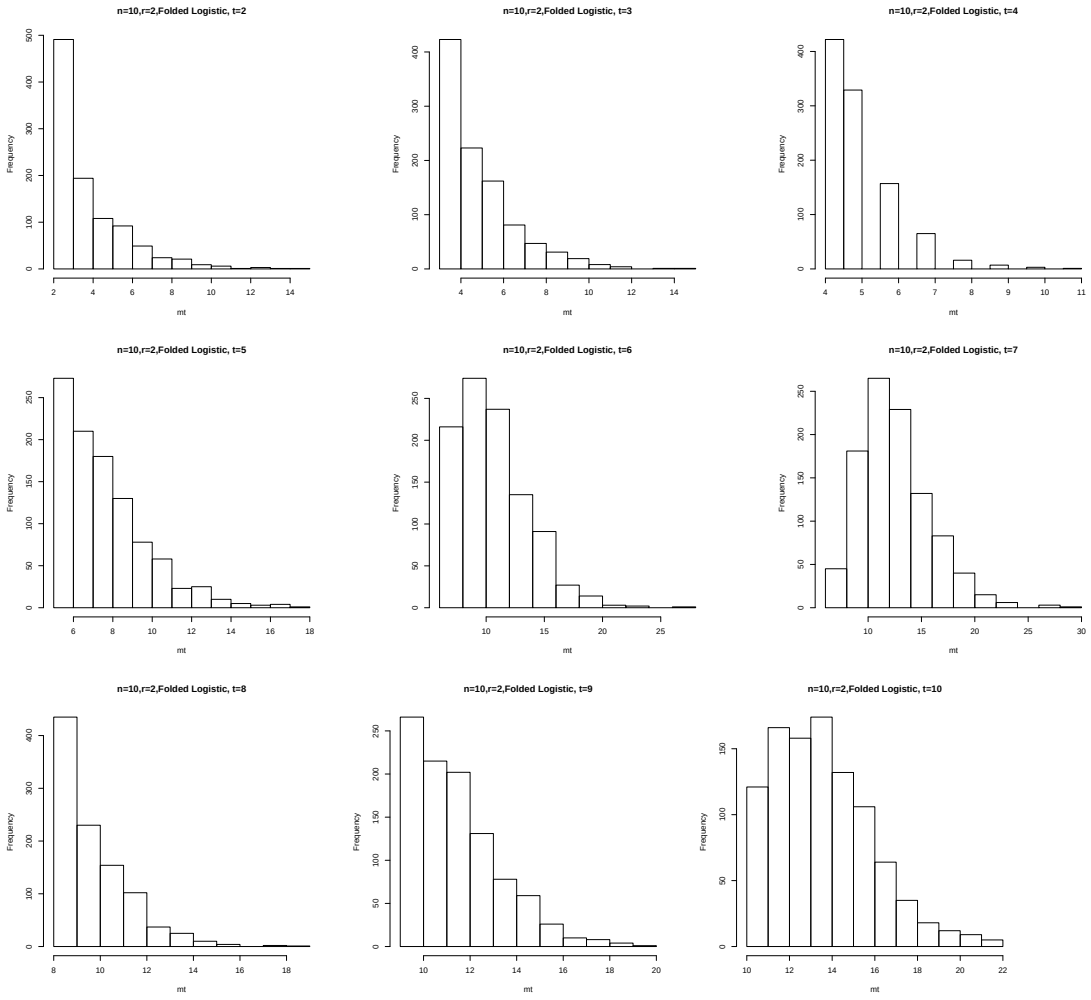

Figure 5.5: $\lambda_1 = 0.30$,$E(R) = 3$, $n = 10$, $r = 2$, $t = 2, \ldots, 10$

Table 5.7: $\lambda_1 = 0.30$,$\rho_2 = 0.1441$,$E(R) = 3$,$n = 10$,$r = 3$

t	$mean(\hat{\lambda}_1)$	$se(\hat{\lambda}_1)$	$mean\hat{\rho}_2$	$se(\hat{\rho}_2)$	$mean(m_t)$	$se(m_t)$
2	0.3347	0.0821	-0.0194	0.0591	3.9780	1.9192
3	0.3179	0.1053	0.1596	0.1434	6.9750	3.1161
4	0.3588	0.0935	0.0814	0.1049	5.9100	1.6860
5	0.4424	0.0682	0.0089	0.0685	6.0380	1.0990
6	0.2894	0.0680	0.1249	0.0755	11.9290	3.4664
7	0.3435	0.0617	0.1058	0.0696	12.8920	3.4044
8	0.3211	0.0649	0.2321	0.1268	17.1070	4.1726
9	0.3607	0.0541	0.0892	0.0644	15.9140	3.4944
10	0.3184	0.0530	0.1382	0.0603	18.9700	4.1423

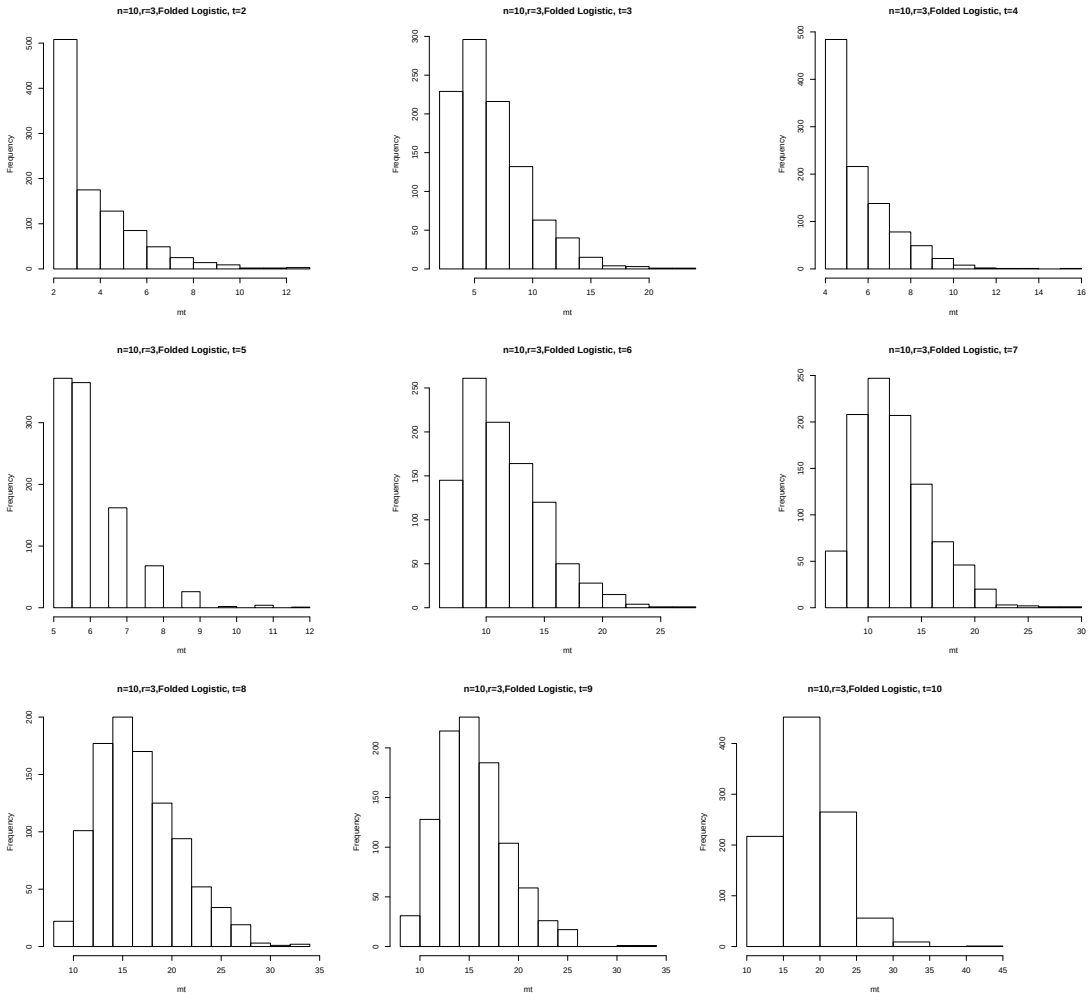

Figure 5.6: $\lambda_1 = 0.30$,$E(R) = 3$, $n = 10$, $r = 3$, $t = 2, \ldots, 10$

Table 5.8: $\lambda_1 = 0.30$,$\rho_2 = 0.1441$,$E(R) = 3$,$n = 10$,$r = 4$

t	$mean(\hat{\lambda}_1)$	$se(\hat{\lambda}_1)$	$mean\hat{\rho}_2$	$se(\hat{\rho}_2)$	$mean(m_t)$	$se(m_t)$
2	0.4126	0.0451	-0.0927	0.0145	3.0040	1.2319
3	0.3392	0.0655	-0.0273	0.0511	5.0670	1.8730
4	0.4080	0.1095	0.2796	0.0991	9.9870	3.9177
5	0.3863	0.0817	0.0992	0.0940	7.8650	2.0888
6	0.3048	0.0665	0.1593	0.0643	16.1730	5.2219
7	0.3204	0.0502	0.0698	0.0423	17.8270	5.4583
8	0.3342	0.0560	0.1997	0.0897	23.8360	6.7172
9	0.3411	0.0453	0.0778	0.0519	22.7550	5.8463
10	0.2987	0.0376	0.1303	0.0763	35.6810	8.9923

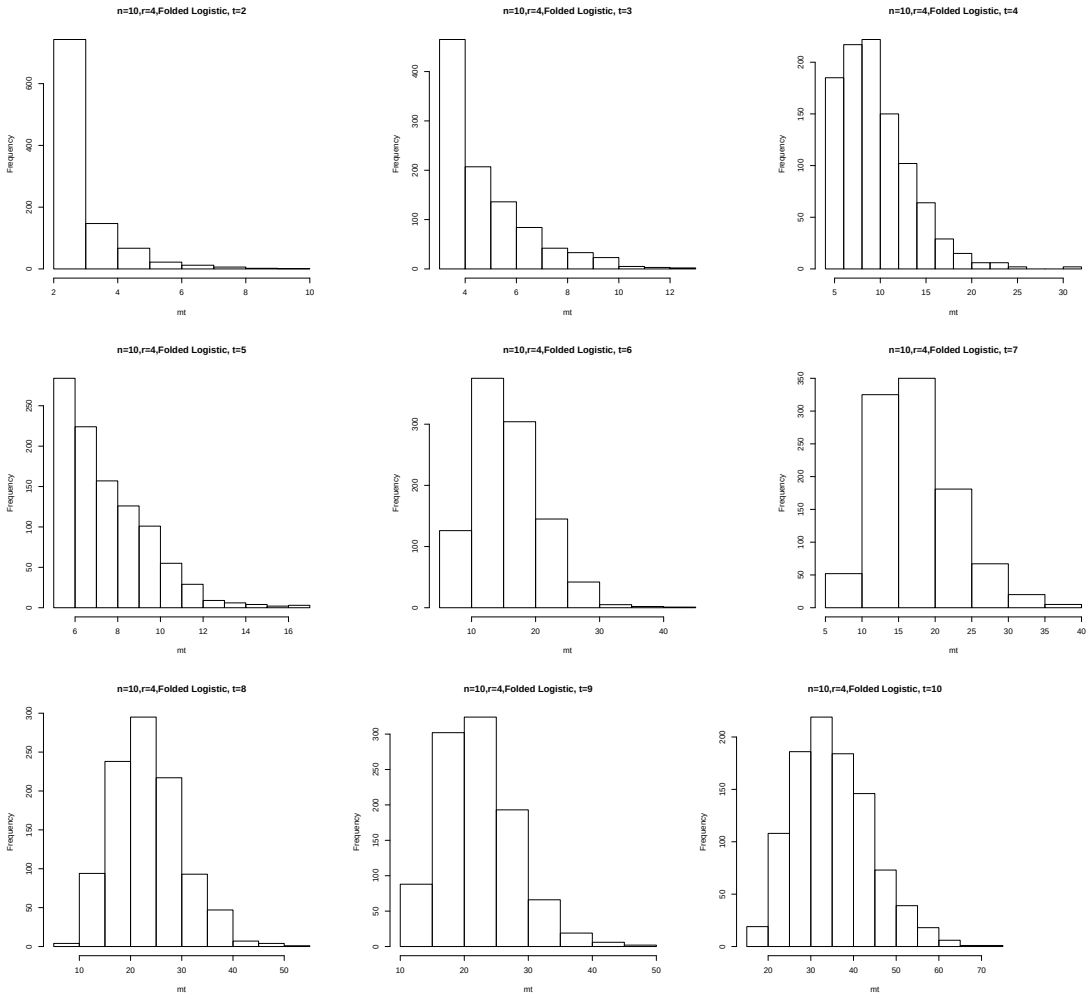

Figure 5.7: $\lambda_1 = 0.30$,$E(R) = 3$, $n = 10$, $r = 4$, $t = 2, \ldots, 10$

Table 5.9: $\lambda_1 = 0.85$,$\rho_2 = 0.3328$,$E(R) = 4.25$,$n = 5$,$r = 4$

t	$mean(\hat{\lambda}_1)$	$se(\hat{\lambda}_1)$	$mean\hat{\rho}_2$	$se(\hat{\rho}_2)$	$mean(m_t)$	$se(m_t)$
2	0.9029	0.0942	0.0880	0.0569	2.9840	1.1735
3	0.8395	0.1111	0.2259	0.1132	4.9460	1.8370
4	0.8265	0.1260	0.3302	0.2968	5.0040	1.1469
5	0.8703	0.0650	0.0408	0.0830	6.9810	1.6815
6	0.9248	0.0718	0.4271	0.0345	7.0050	1.0845
7	0.9221	0.0524	0.1691	0.0387	8.9420	1.5688
8	0.7687	0.0850	0.1891	0.2761	10.0220	1.6050
9	0.8273	0.0796	0.4452	0.1812	12.0290	1.9125
10	0.8357	0.0657	0.3151	0.1434	14.1060	2.3619

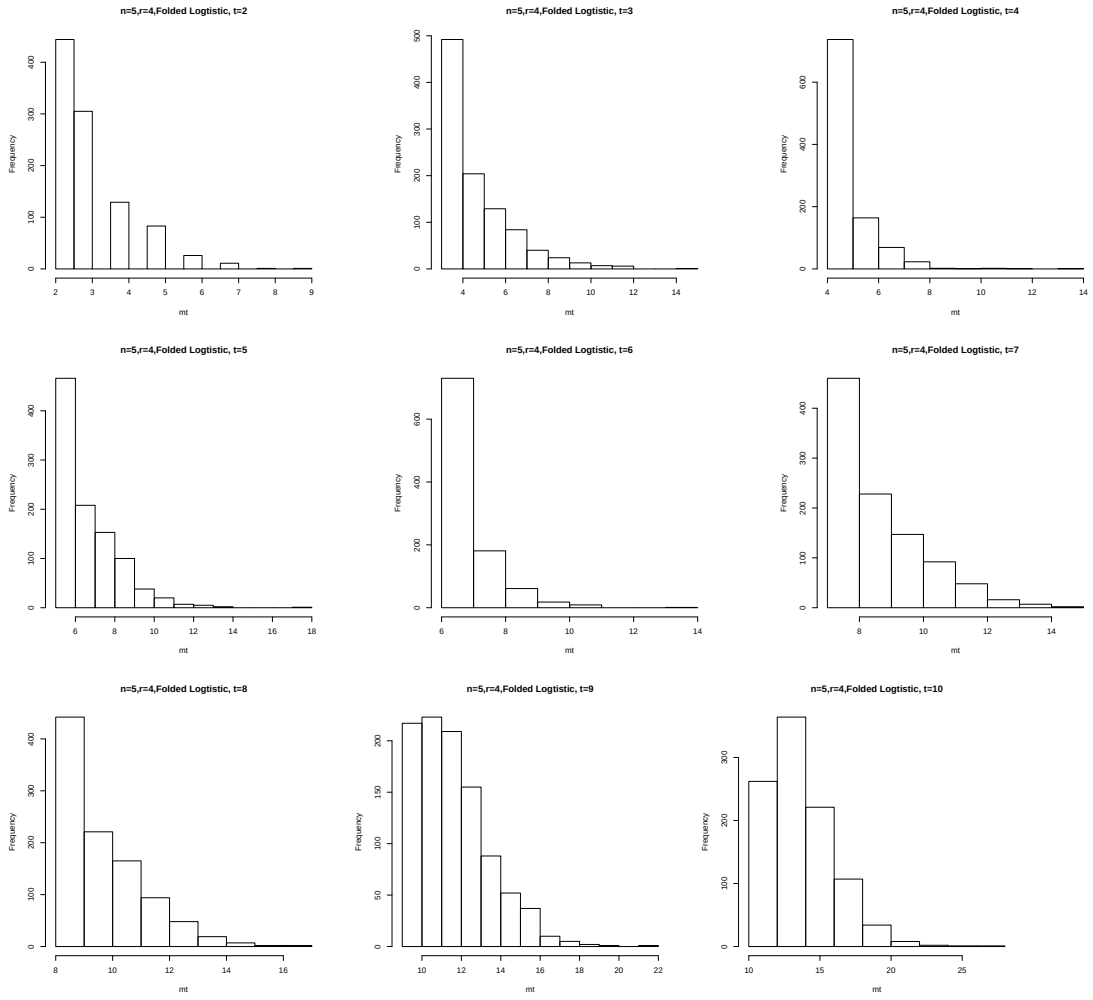

Figure 5.8: $\lambda_1 = 0.85$,$E(R) = 4.25$, $n = 5$, $r = 4$, $t = 2, \ldots, 10$

Table 5.10: $\lambda_1 = 0.85$,$\rho_2 = 0.3328$,$E(R) = 4.25$,$n = 5$,$r = 5$

t	$mean(\hat{\lambda}_1)$	$se(\hat{\lambda}_1)$	$mean\hat{\rho}_2$	$se(\hat{\rho}_2)$	$mean(m_t)$	$se(m_t)$
2	0.8447	0.1050	0.0472	0.0755	4.0500	2.0755
3	0.8909	0.0637	-0.0282	0.0776	5.9610	2.5353
4	0.8070	0.1149	0.3765	0.1902	6.8660	2.2599
5	0.8992	0.0653	0.1411	0.0491	7.0420	1.70566
6	0.8902	0.0719	0.2937	0.1012	7.9540	1.5486
7	0.8286	0.0603	0.1453	0.0965	13.8200	3.5995
8	0.8286	0.0682	0.2639	0.1421	13.9850	3.2972
9	0.7872	0.0782	0.4801	0.2257	17.0280	3.7764
10	0.8203	0.0576	0.2303	0.1165	18.9730	4.1878

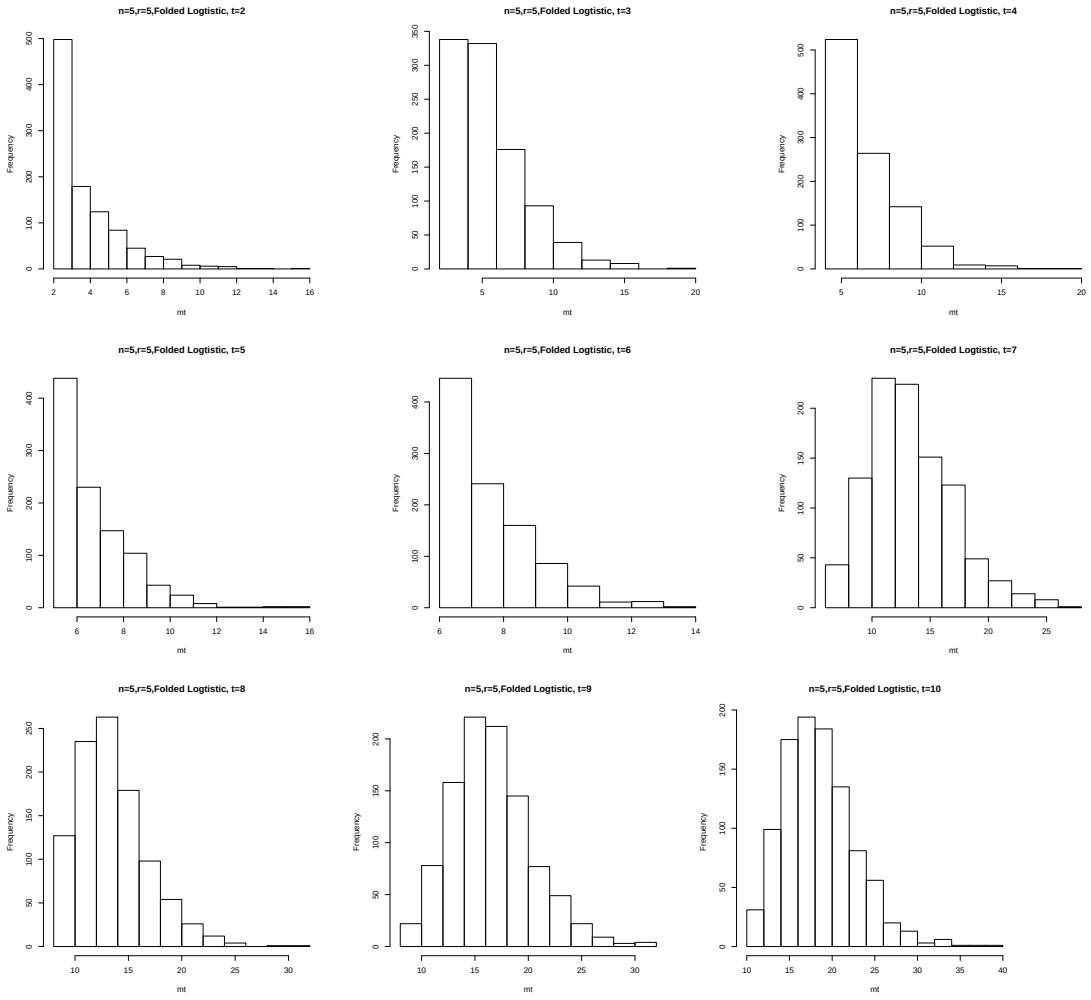

Figure 5.9: $\lambda_1 = 0.85$,$E(R) = 4.25$, $n = 5$, $r = 5$, $t = 2, \ldots, 10$

Table 5.11: $\lambda_1 = 0.85$,$\rho_2 = 0.3328$,$E(R) = 8.5$,$n = 10$,$r = 8$

t	$mean(\hat{\lambda}_1)$	$se(\hat{\lambda}_1)$	$mean\hat{\rho}_2$	$se(\hat{\rho}_2)$	$mean(m_t)$	$se(m_t)$
2	0.8023	0.1140	0.0678	0.1167	3.0420	1.2969
3	0.8981	0.0727	0.0822	0.1123	3.9470	1.1382
4	0.8486	0.0603	0.0419	0.0682	5.9310	1.6921
5	0.8658	0.0559	0.0687	0.0664	7.0340	1.7081
6	0.8909	0.0525	0.1015	0.0737	7.0600	1.0870
7	0.8574	0.0956	0.5378	0.3574	8.9820	1.5768
8	0.8605	0.0480	0.1453	0.0680	11.8500	2.3093
9	0.8510	0.0680	0.3338	0.1697	12.0290	1.9749
10	0.8447	0.0597	0.2368	0.1676	13.0360	1.9821

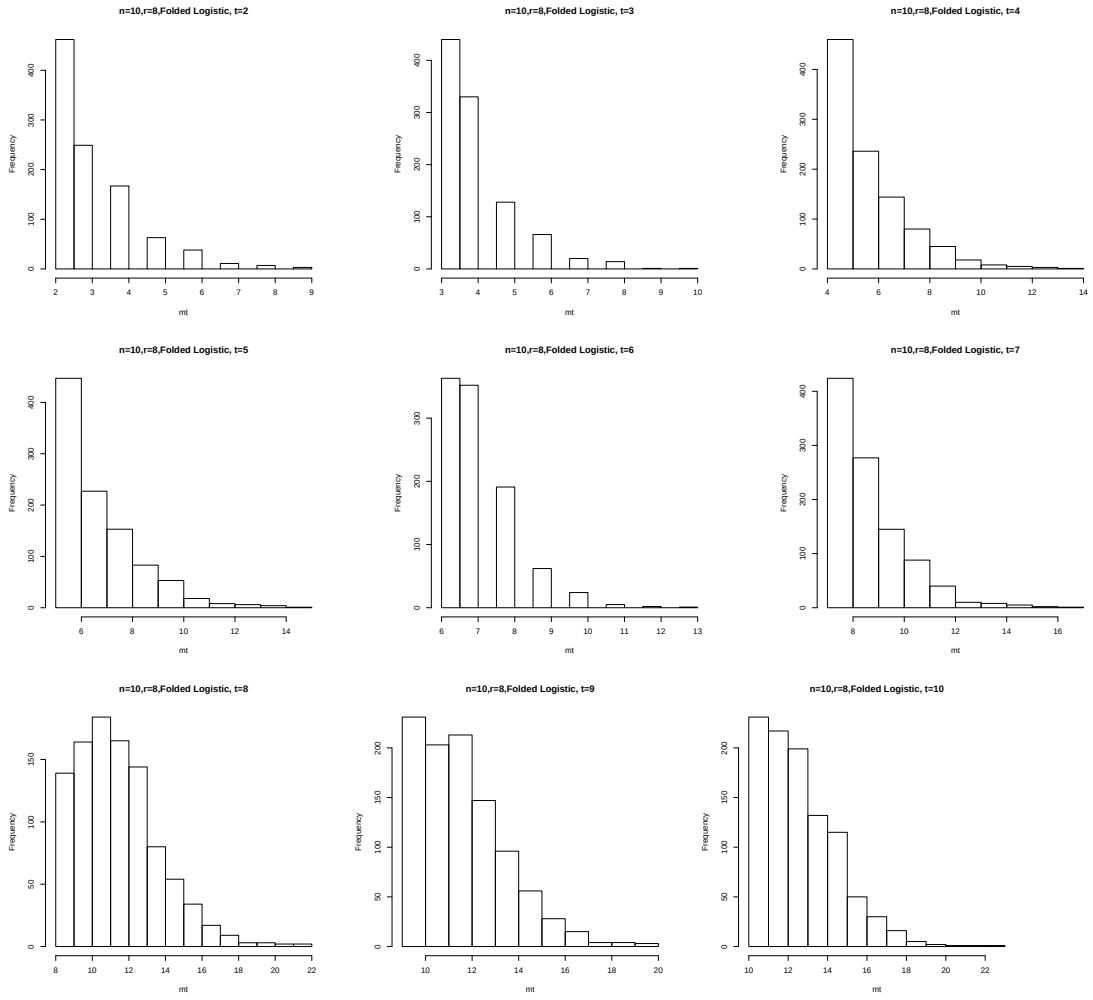

Figure 5.10: $\lambda_1 = 0.85$,$E(R) = 8.5$, $n = 10$, $r = 8$, $t = 2, \ldots, 10$

Table 5.12: $\lambda_1 = 0.85$,$\rho_2 = 0.3328$,$E(R) = 8.5$,$n = 10$,$r = 9$

t	$mean(\hat{\lambda}_1)$	$se(\hat{\lambda}_1)$	$mean\hat{\rho}_2$	$se(\hat{\rho}_2)$	$mean(m_t)$	$se(m_t)$
2	0.7659	0.1748	0.3197	0.3022	2.9730	1.1513
3	0.9114	0.0564	0.0275	0.0738	4.0610	1.2020
4	0.9336	0.0639	0.2468	0.0366	5.0070	1.1376
5	0.9069	0.0524	0.0718	0.1046	6.0220	1.0801
6	0.8848	0.0467	0.0550	0.0739	7.9690	1.5958
7	0.8561	0.0753	0.2775	0.1653	9.0180	1.6341
8	0.8551	0.0596	0.2430	0.0747	11.9490	2.4618
9	0.8400	0.0598	0.2249	0.1537	13.0320	2.4583
10	0.8267	0.0798	0.5122	0.1691	13.9550	2.3350

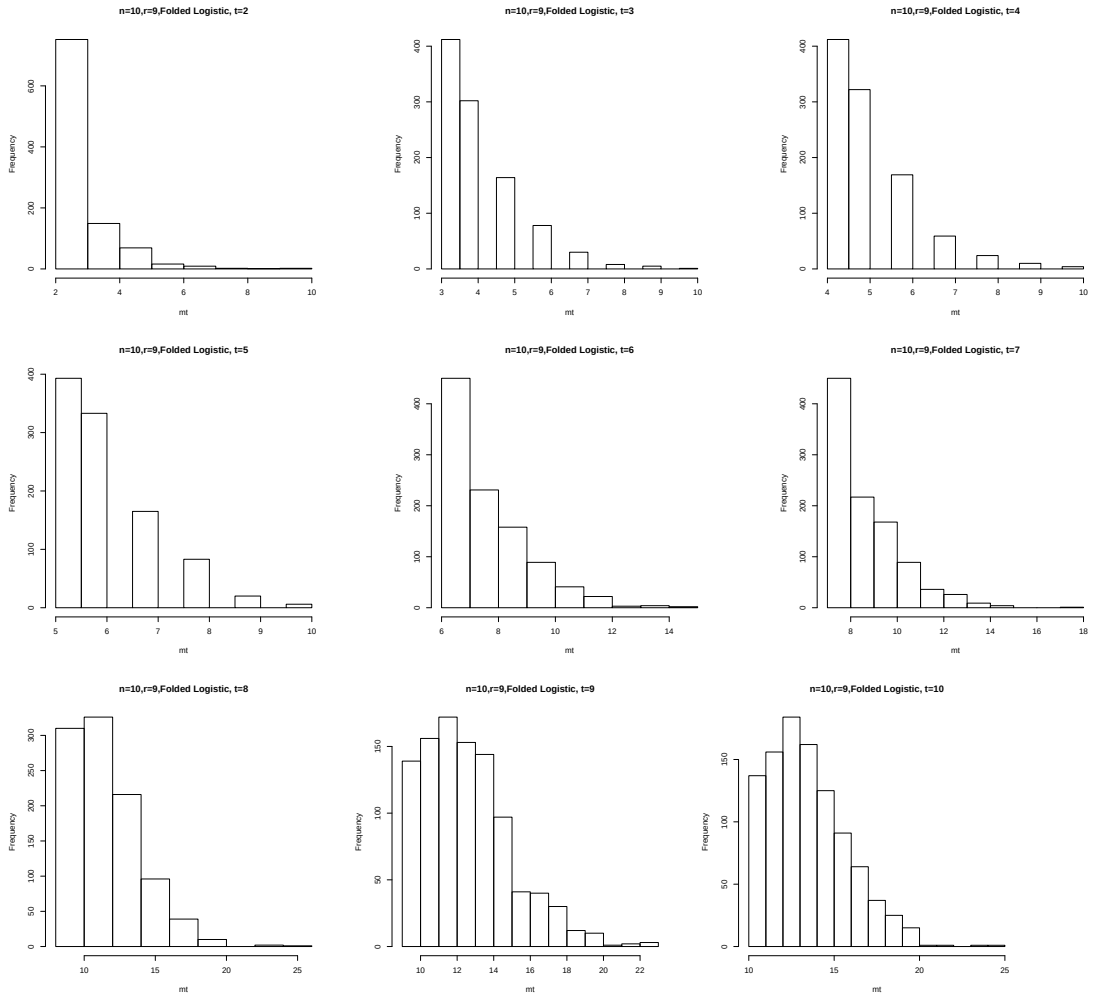

Figure 5.11: $\lambda_1 = 0.85$,$E(R) = 8.5$, $n = 10$, $r = 9$, $t = 2, \ldots, 10$

Simulation Results Power Family $\gamma = 0.50$

Table 5.13: $\lambda_1 = 0.05$, $\rho_2 = 0.2517$, $E(R) = 0.25$, $r = 1$, $n = 5$

t	$mean(\hat{\lambda}_1)$	$se(\hat{\lambda}_1)$	$mean(\hat{\rho}_2)$	$se(\hat{\rho}_2)$	$mean(m_t)$	$se(m_t)$
2	0.0569	0.0425	-0.0627	0.0522	10.8200	6.7847
3	0.0443	0.0264	-0.0472	0.0304	18.2090	9.8407
4	0.0568	0.0321	0.0354	0.0690	22.0780	10.0619
5	0.0332	0.0150	-0.0346	0.0165	35.7280	14.6006
6	0.0312	0.0137	-0.0324	0.0150	44.8530	17.0190
7	0.0505	0.0182	-0.0535	0.0209	31.0860	10.5149
8	0.0625	0.0196	-0.0672	0.0229	28.1110	8.8688
9	0.0453	0.0153	0.0009	0.0404	48.4490	14.6647
10	0.0594	0.0215	0.1409	0.1640	46.8250	12.5548

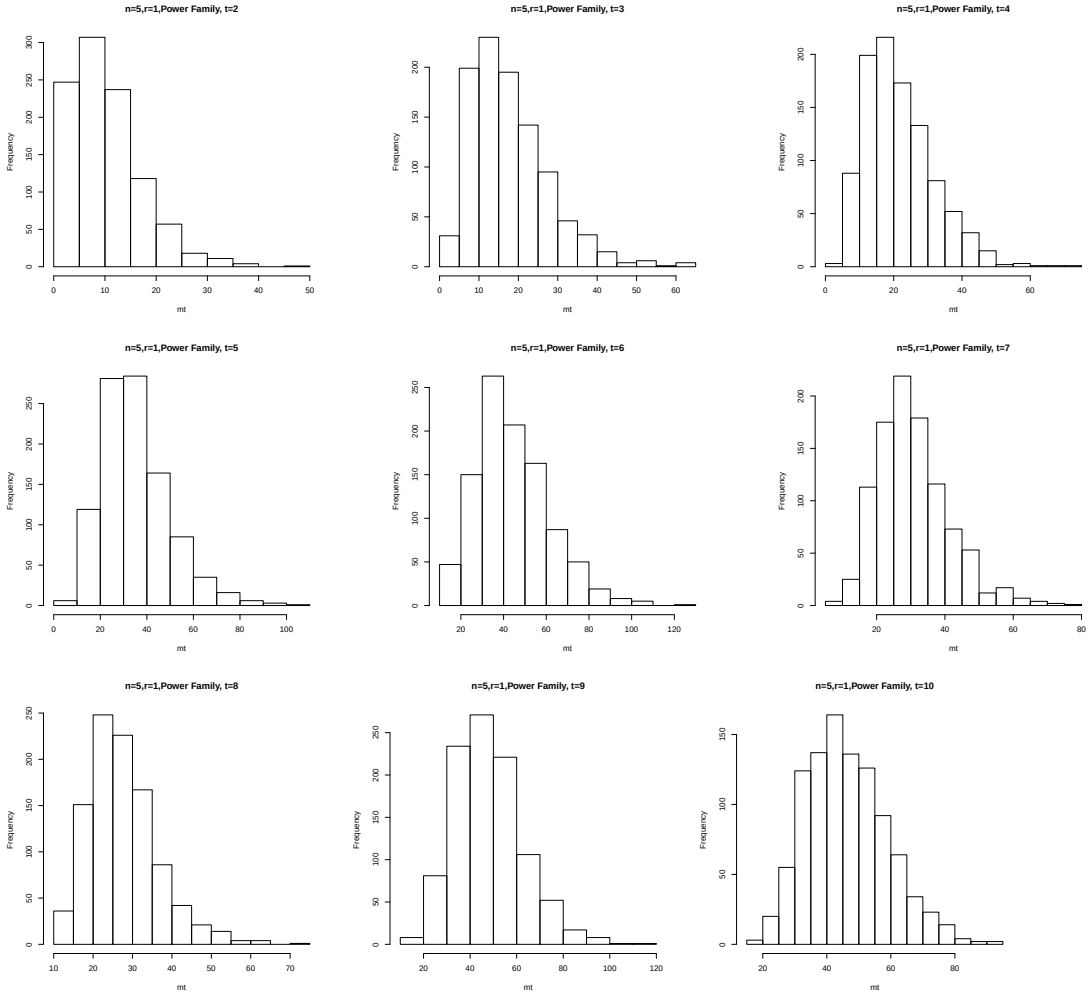

Figure 5.12: $\lambda_1 = 0.05$,$E(R) = 0.25$, $n = 5$, $r = 1$, $t = 2, \ldots, 10$

Table 5.14: $\lambda_1 = 0.05$, $\rho_2 = 0.2517$, $E(R) = 0.5$, $r = 1$, $n = 10$

t	$mean(\hat{\lambda}_1)$	$se(\hat{\lambda}_1)$	$mean(\hat{\rho}_2)$	$se(\hat{\rho}_2)$	$mean(m_t)$	$se(m_t)$
2	0.0318	0.0211	-0.0333	0.0233	9.0570	5.3927
3	0.0822	0.0562	0.1036	0.1080	9.1600	4.3669
4	0.0425	0.0176	-0.0448	0.0196	11.0950	4.6493
5	0.0492	0.0231	0.0530	0.0480	18.8450	7.3723
6	0.0593	0.0282	0.0655	0.0850	16.8890	5.5141
7	0.0450	0.0128	-0.0473	0.0142	16.8280	4.7536
8	0.0486	0.0145	-0.0277	0.0211	19.8710	5.3728
9	0.0440	0.0119	-0.0462	0.0132	21.9230	5.7842
10	0.0395	0.0154	0.1018	0.0933	40.5650	11.1275

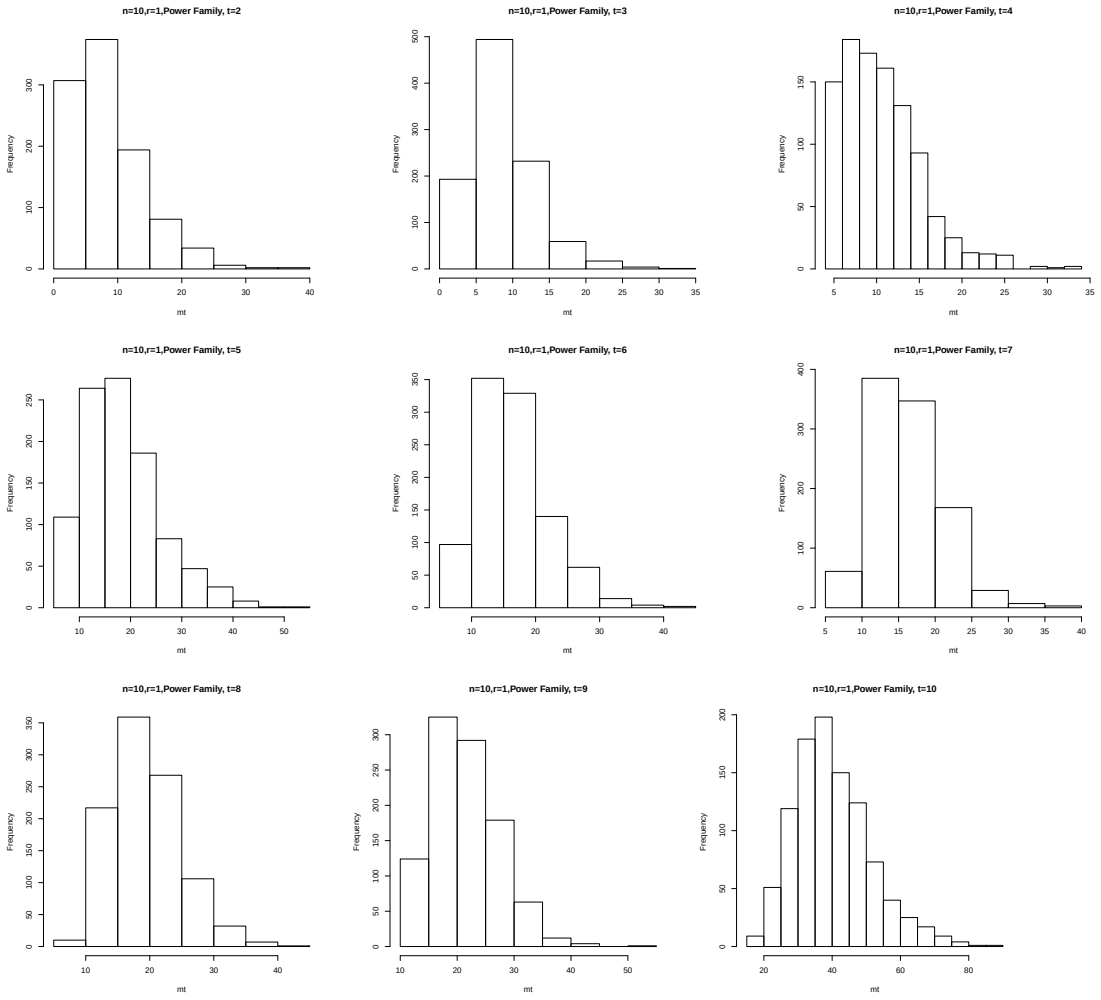

Figure 5.13: $\lambda_1 = 0.05$,$E(R) = 0.5$, $n = 10$, $r = 1$, $t = 2, \ldots, 10$

Table 5.15: $\lambda_1 = 0.30$, $\rho_2 = 0.439$, $E(R) = 1.5$, $r = 1$, $n = 5$

t	$mean(\hat{\lambda}_1)$	$se(\hat{\lambda}_1)$	$mean(\hat{\rho}_2)$	$se(\hat{\rho}_2)$	$mean(m_t)$	$se(m_t)$
2	0.1508	0.0465	-0.1811	0.0640	2.9930	1.1977
3	0.1837	0.0694	-0.0936	0.0908	4.8850	1.7807
4	0.2118	0.0857	-0.0156	0.1409	6.0140	1.7677
5	0.1724	0.0261	-0.2095	0.0376	5.9500	1.0112
6	0.2864	0.0751	0.0205	0.0942	7.9820	1.6292
7	0.1777	0.0218	-0.2169	0.0318	8.0130	1.1100
8	0.4093	0.1163	0.3052	0.1609	9.0230	1.0685
9	0.3631	0.0958	0.2072	0.1705	9.9880	0.9924
10	0.2800	0.0775	0.1814	0.1781	13.0080	1.9758

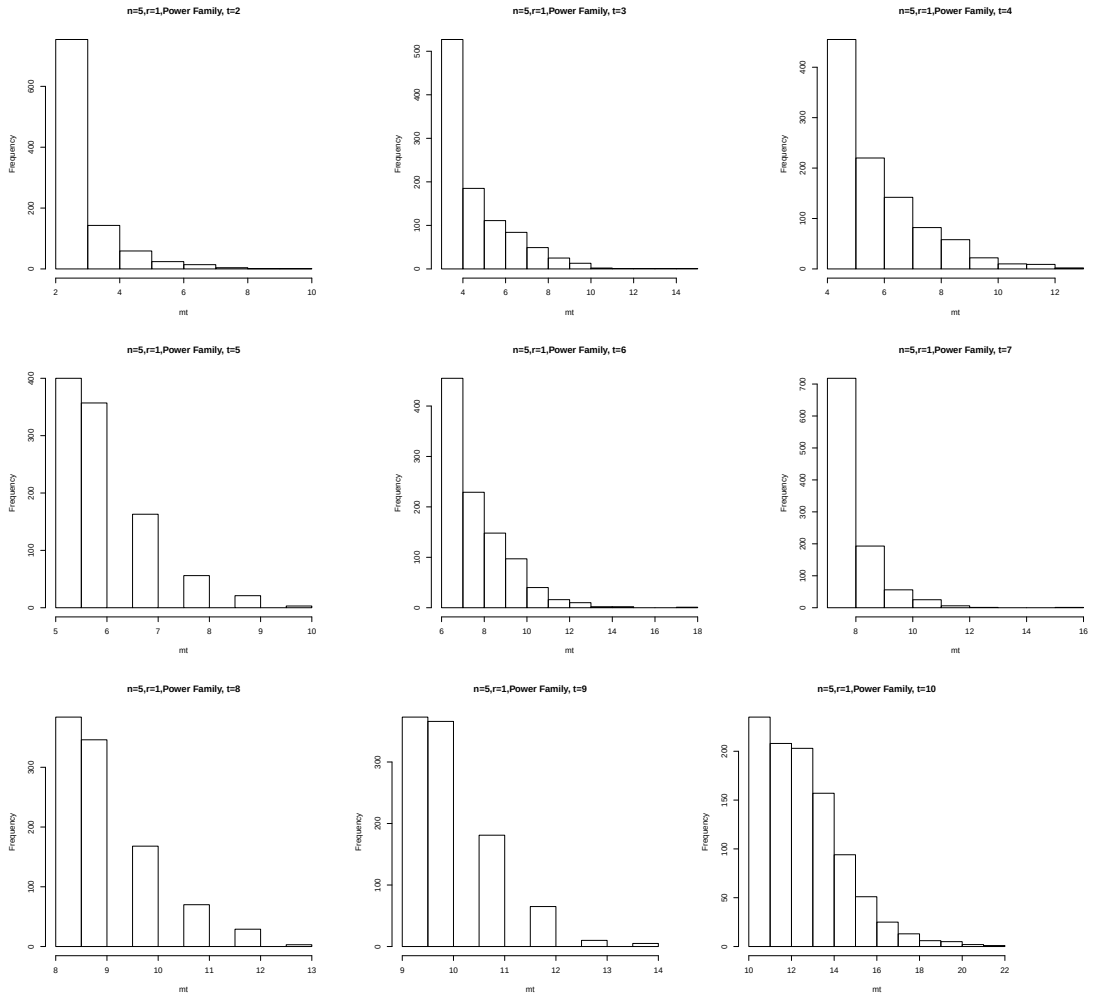

Figure 5.14: $\lambda_1 = 0.30$,$E(R) = 1.5$, $n = 5$, $r = 1$, $t = 2, \ldots, 10$

Table 5.16: $\lambda_1 = 0.30$, $\rho_2 = 0.439$, $E(R) = 1.5$, $r = 2$, $n = 5$

t	$mean(\hat{\lambda}_1)$	$se(\hat{\lambda}_1)$	$mean(\hat{\rho}_2)$	$se(\hat{\rho}_2)$	$mean(m_t)$	$se(m_t)$
2	0.3713	0.1360	0.0900	0.1441	7.1930	4.2905
3	0.4663	0.1134	0.0792	0.0975	7.1490	3.1271
4	0.3330	0.0748	0.0549	0.0938	14.9070	6.4396
5	0.3956	0.1183	0.4922	0.1514	14.9580	5.7165
6	0.3467	0.0751	0.2016	0.1172	19.0440	6.3081
7	0.3932	0.0830	0.3403	0.1185	20.9210	6.4947
8	0.4087	0.0673	0.1884	0.0961	21.8380	6.2041
9	0.3164	0.0554	0.2644	0.0994	37.4790	10.9310
10	0.3093	0.0445	0.1455	0.0962	41.9250	11.2484

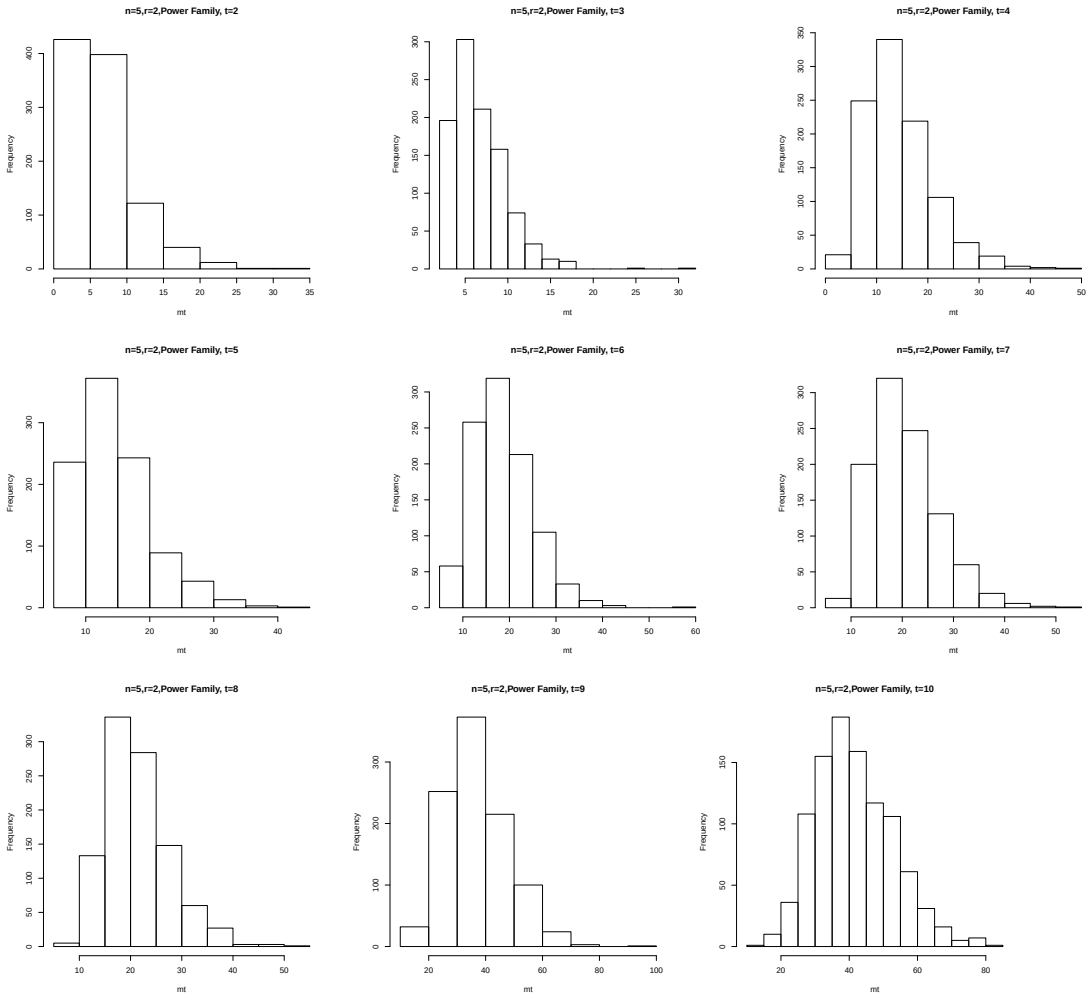

Figure 5.15: $\lambda_1 = 0.30$,$E(R) = 1.5$, $n = 5$, $r = 2$, $t = 2, \ldots, 10$

Table 5.17: $\lambda_1 = 0.30$, $\rho_2 = 0.439$, $E(R) = 3$, $r = 2$, $n = 10$

t	$mean(\hat{\lambda}_1)$	$se(\hat{\lambda}_1)$	$mean(\hat{\rho}_2)$	$se(\hat{\rho}_2)$	$mean(m_t)$	$se(m_t)$
2	0.4470	0.1360	0.0644	0.1390	2.9720	1.2102
3	0.3850	0.1378	0.4279	0.1288	9.0040	4.0778
4	0.2601	0.0827	0.2465	0.1018	14.8670	6.2703
5	0.2967	0.0872	0.3725	0.0759	19.7930	7.9906
6	0.3055	0.0687	0.2040	0.0671	14.9660	4.5629
7	0.3405	0.0810	0.3778	0.0885	19.7200	6.0747
8	0.2931	0.0635	0.3528	0.0796	27.1090	7.9121
9	0.2794	0.0550	0.2884	0.0743	27.7910	7.5924
10	0.2546	0.0442	0.3439	0.0573	44.0270	11.9254

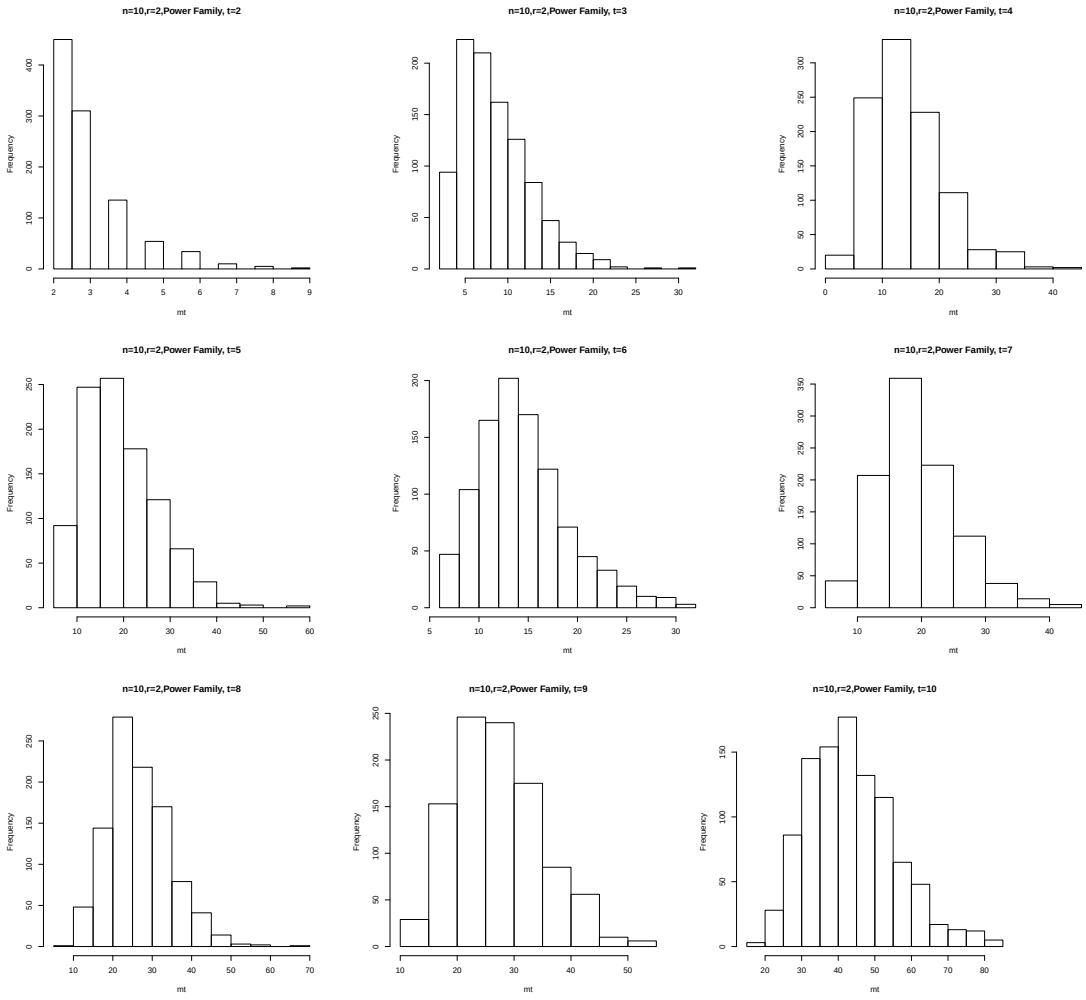

Figure 5.16: $\lambda_1 = 0.30$,$E(R) = 3$, $n = 10$, $r = 2$, $t = 2, \ldots, 10$

Table 5.18: $\lambda_1 = 0.30$, $\rho_2 = 0.439$, $E(R) = 3$, $r = 3$, $n = 10$

t	$mean(\hat{\lambda}_1)$	$se(\hat{\lambda}_1)$	$mean(\hat{\rho}_2)$	$se(\hat{\rho}_2)$	$mean(m_t)$	$se(m_t)$
2	0.4411	0.1540	0.2462	0.1529	5.1300	2.7416
3	0.2762	0.1039	0.1993	0.0853	11.2440	5.8772
4	0.2287	0.0660	0.1460	0.0604	17.1420	7.6305
5	0.2445	0.0657	0.2275	0.0722	20.1240	7.7572
6	0.2533	0.0666	0.3480	0.1010	25.1920	9.0302
7	0.2488	0.0578	0.3248	0.0955	29.9050	9.6860
8	0.2489	0.0515	0.3268	0.0730	36.1960	11.1369
9	0.2384	0.0442	0.2796	0.0615	41.7960	12.1931
10	0.2168	0.0368	0.2200	0.0704	50.0560	14.4593

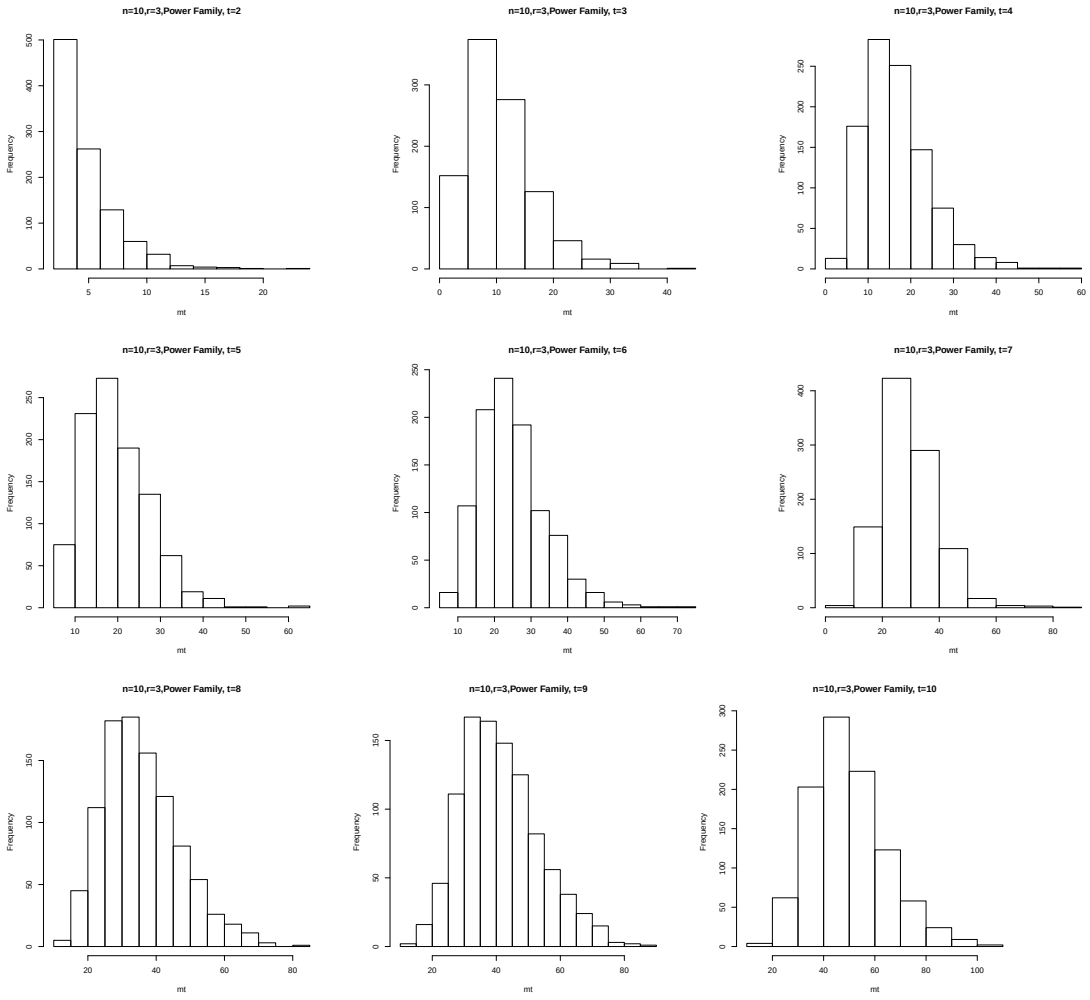

Figure 5.17: $\lambda_1 = 0.30$,$E(R) = 3$, $n = 10$, $r = 3$, $t = 2, \ldots, 10$

Table 5.19: $\lambda_1 = 0.30$, $\rho_2 = 0.439$, $E(R) = 3$, $r = 4$, $n = 10$

t	$mean(\hat{\lambda}_1)$	$se(\hat{\lambda}_1)$	$mean(\hat{\rho}_2)$	$se(\hat{\rho}_2)$	$mean(m_t)$	$se(m_t)$
2	0.3984	0.1110	0.0184	0.1041	3.0560	1.2516
3	0.5265	0.1419	0.2308	0.1824	4.9880	1.8186
4	0.3508	0.1071	0.3077	0.0842	10.7690	4.3585
5	0.3327	0.0856	0.2404	0.1053	11.8460	3.9099
6	0.3451	0.0826	0.2834	0.1082	15.2640	4.9027
7	0.3655	0.0821	0.4490	0.0689	21.3500	6.5137
8	0.2648	0.0521	0.2322	0.0612	29.1830	8.8985
9	0.2641	0.0536	0.3217	0.0701	34.5440	10.2186
10	0.2681	0.0485	0.3115	0.0604	38.5040	10.7225

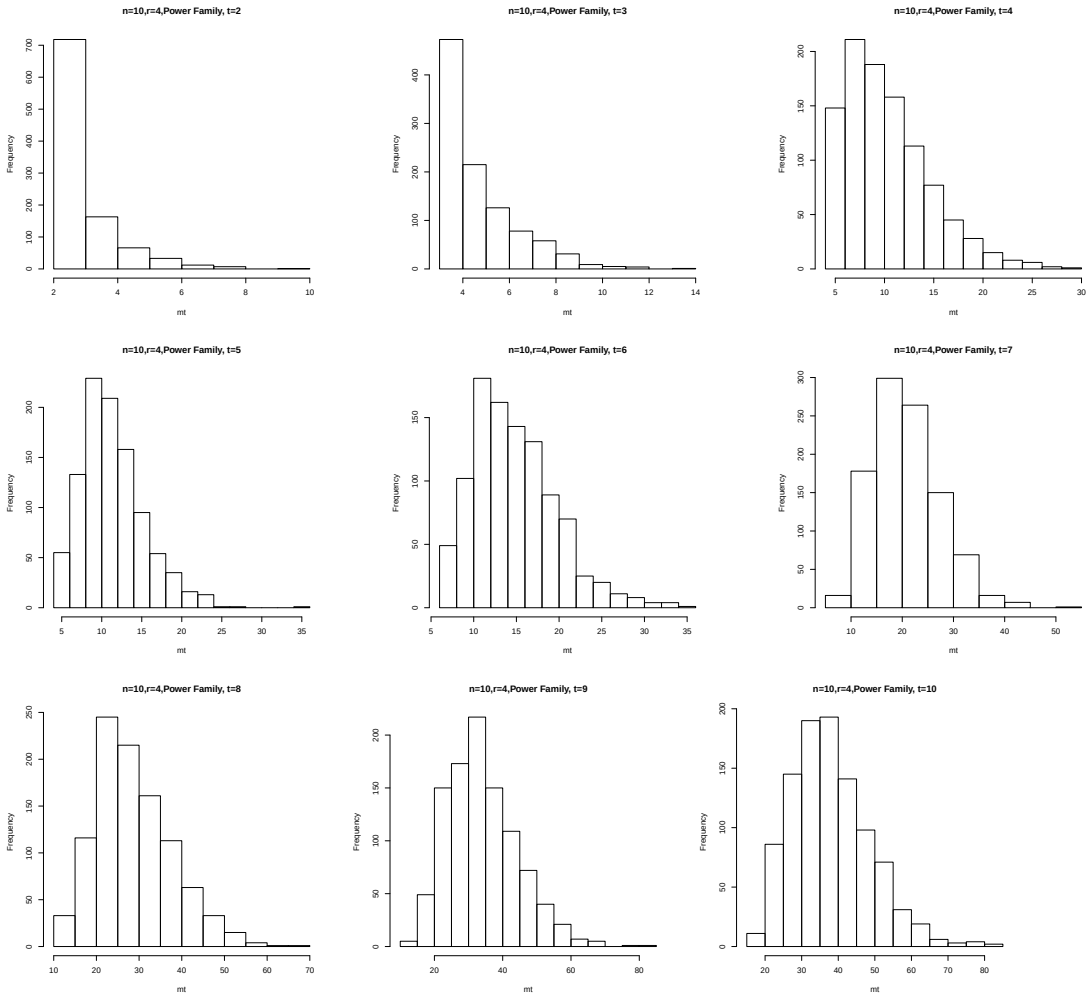

Figure 5.18: $\lambda_1 = 0.30$,$E(R) = 3$, $n = 10$, $r = 4$, $t = 2, \ldots, 10$

Table 5.20: $\lambda_1 = 0.85$, $\rho_2 = 0.566$, $E(R) = 4.25$, $r = 4$, $n = 5$

t	$mean(\hat{\lambda}_1)$	$se(\hat{\lambda}_1)$	$mean(\hat{\rho}_2)$	$se(\hat{\rho}_2)$	$mean(m_t)$	$se(m_t)$
2	0.8149	0.1879	0.6061	0.0687	2.9380	1.1788
3	0.9209	0.0752	0.1297	0.0473	3.9930	1.1306
4	0.7474	0.1471	0.3741	0.4113	5.0260	1.1192
5	0.8031	0.1205	0.4241	0.3582	6.9130	1.6261
6	0.8382	0.0967	0.2497	0.2755	7.0420	1.1382
7	0.7554	0.1090	0.4224	0.2871	8.9500	1.5710
8	0.7795	0.1157	0.6437	0.3293	9.9600	1.5572
9	0.8448	0.0747	0.3516	0.1686	12.0230	2.0031
10	0.8037	0.0856	0.5208	0.2147	13.9070	2.3744

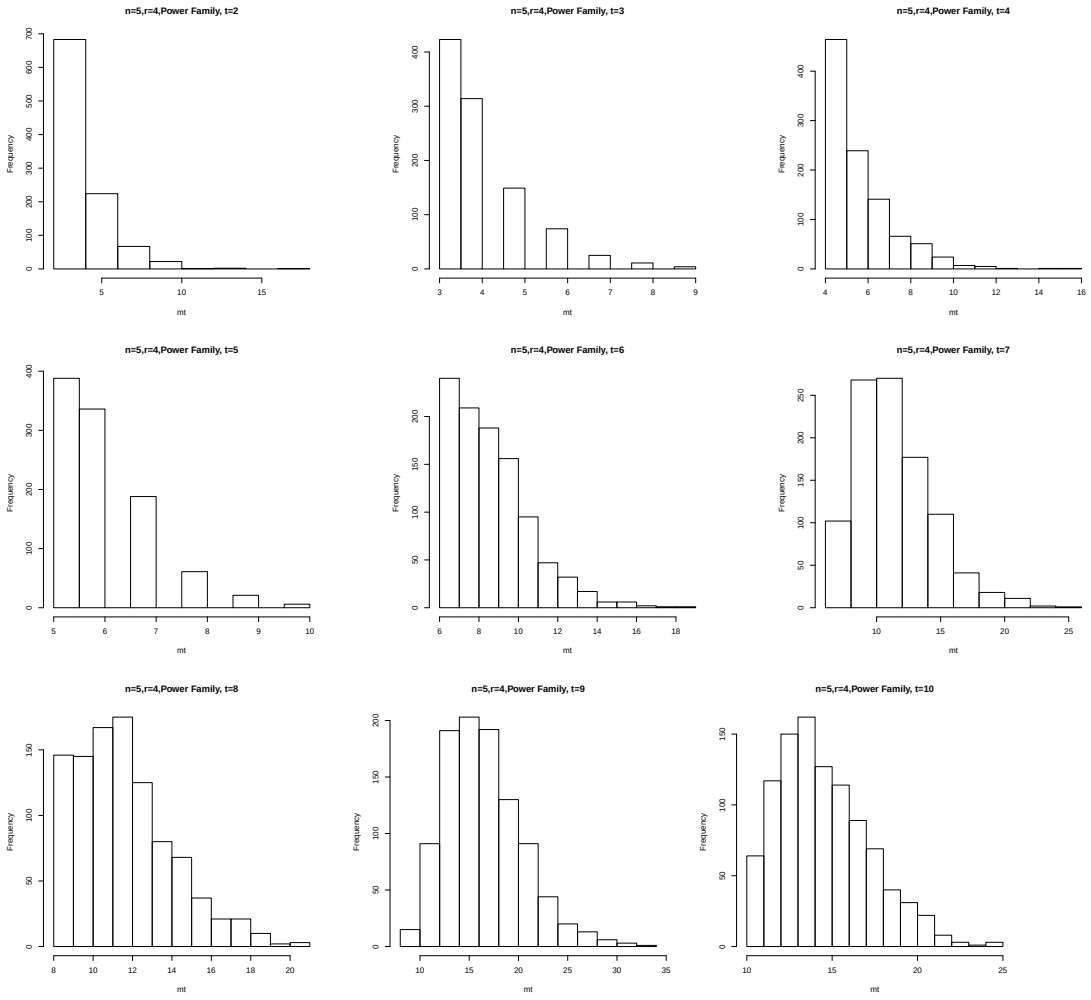

Figure 5.19: $\lambda_1 = 0.85$,$E(R) = 4.25$, $n = 5$, $r = 4$, $t = 2, \ldots, 10$

Table 5.21: $\lambda_1 = 0.85$, $\rho_2 = 0.566$, $E(R) = 4.25$, $r = 5$, $n = 5$

t	$mean(\hat{\lambda}_1)$	$se(\hat{\lambda}_1)$	$mean(\hat{\rho}_2)$	$se(\hat{\rho}_2)$	$mean(m_t)$	$se(m_t)$
2	0.8480	0.1445	0.3092	0.0782	3.0600	1.2814
3	0.8261	0.0864	0.0736	0.1233	7.1630	3.2253
4	0.6886	0.1318	0.5286	0.2053	8.9020	3.3147
5	0.7189	0.1179	0.5096	0.2438	11.0710	3.5845
6	0.7705	0.0818	0.3362	0.2131	14.1940	4.3136
7	0.8558	0.0525	0.0736	0.0939	13.8360	3.6689
8	0.7899	0.0692	0.3245	0.1976	16.8640	4.1503
9	0.7570	0.0834	0.4892	0.1831	18.1000	4.1513
10	0.6158	0.0899	0.7102	0.1131	25.0470	6.2552

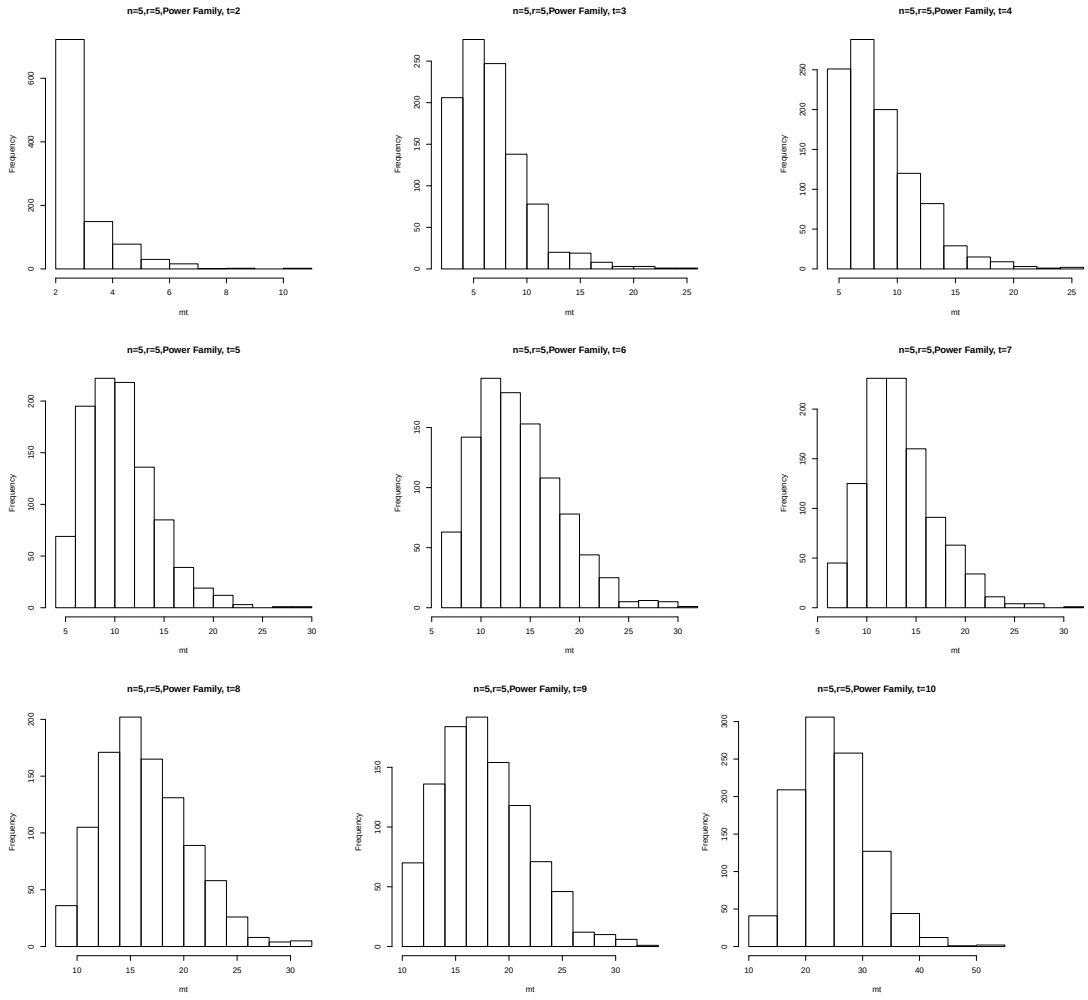

Figure 5.20: $\lambda_1 = 0.85$,$E(R) = 4.25$, $n = 5$, $r = 5$, $t = 2, \ldots, 10$

Table 5.22: $\lambda_1 = 0.85$, $\rho_2 = 0.566$, $E(R) = 8.5$, $r = 8$, $n = 10$

t	$mean(\hat{\lambda}_1)$	$se(\hat{\lambda}_1)$	$mean(\hat{\rho}_2)$	$se(\hat{\rho}_2)$	$mean(m_t)$	$se(m_t)$
2	0.9077	0.0951	0.1879	0.0523	2.9550	1.2484
3	0.7705	0.1859	0.5821	0.4520	4.0250	1.1531
4	0.8705	0.0638	0.0489	0.0816	5.0200	1.1467
5	0.8638	0.0911	0.2651	0.2433	6.0180	1.0931
6	0.7544	0.1151	0.5130	0.2507	9.0290	2.1323
7	0.8038	0.0941	0.4118	0.2584	10.1030	2.0906
8	0.7104	0.1280	0.7544	0.2475	11.0670	2.1069
9	0.8092	0.0965	0.6602	0.2122	11.9570	1.9491
10	0.8422	0.0833	0.4990	0.2781	12.0440	1.5801

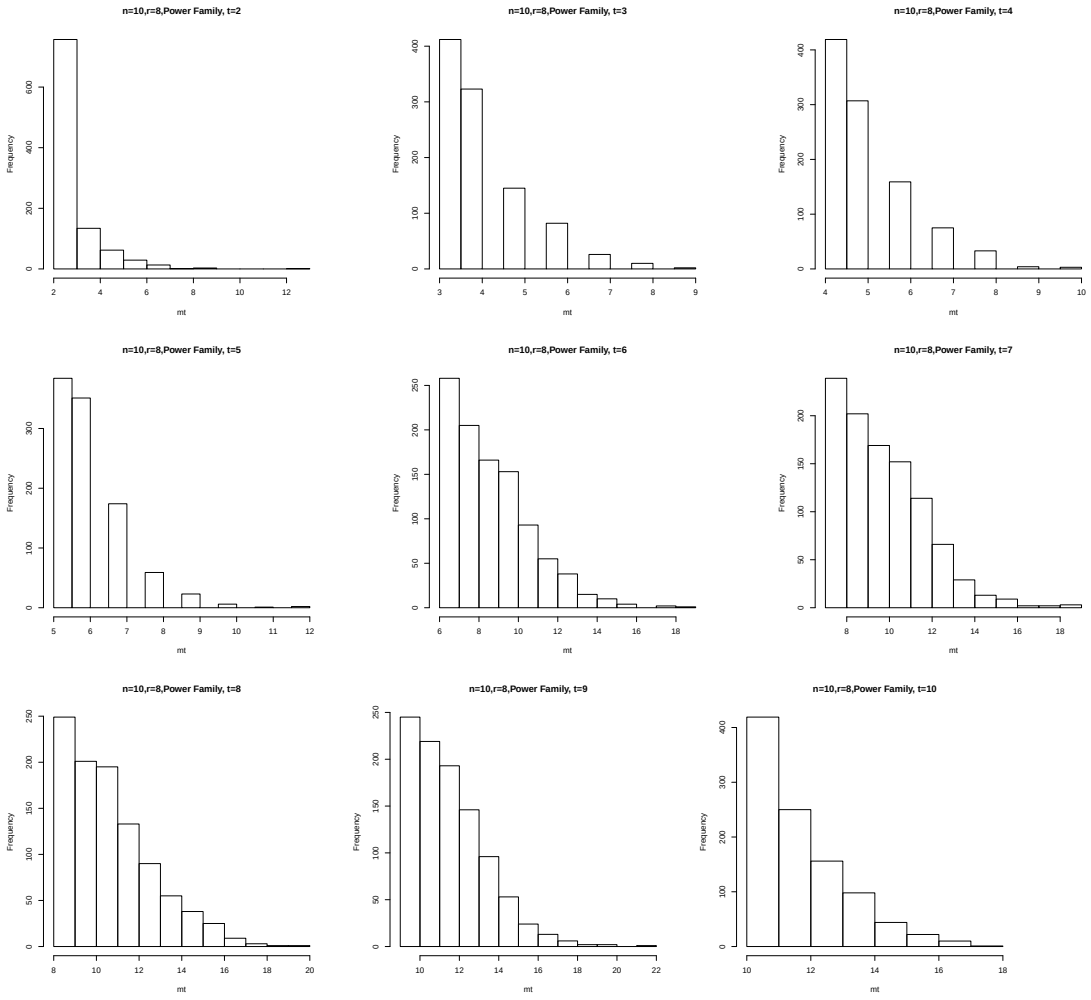

Figure 5.21: $\lambda_1 = 0.85$,$E(R) = 8.5$, $n = 10$, $r = 8$, $t = 2, \ldots, 10$

Table 5.23: $\lambda_1 = 0.85$, $\rho_2 = 0.566$, $E(R) = 8.5$, $r = 9$, $n = 10$

t	$mean(\hat{\lambda}_1)$	$se(\hat{\lambda}_1)$	$mean(\hat{\rho}_2)$	$se(\hat{\rho}_2)$	$mean(m_t)$	$se(m_t)$
2	0.7212	0.1878	0.3945	0.3856	3.9640	2.0201
3	0.7394	0.1440	0.4196	0.3017	6.0520	2.4424
4	0.7546	0.1350	0.4952	0.2651	6.8810	2.2521
5	0.8348	0.1016	0.3650	0.2291	7.0170	1.6479
6	0.8372	0.0766	0.2484	0.1363	8.9530	2.1376
7	0.8626	0.0926	0.4415	0.3840	8.8980	1.5310
8	0.8790	0.0578	0.1891	0.1142	9.9940	1.5983
9	0.8397	0.0700	0.3473	0.1798	13.0450	2.4684
10	0.8281	0.0939	0.6698	0.3139	13.0460	2.0134

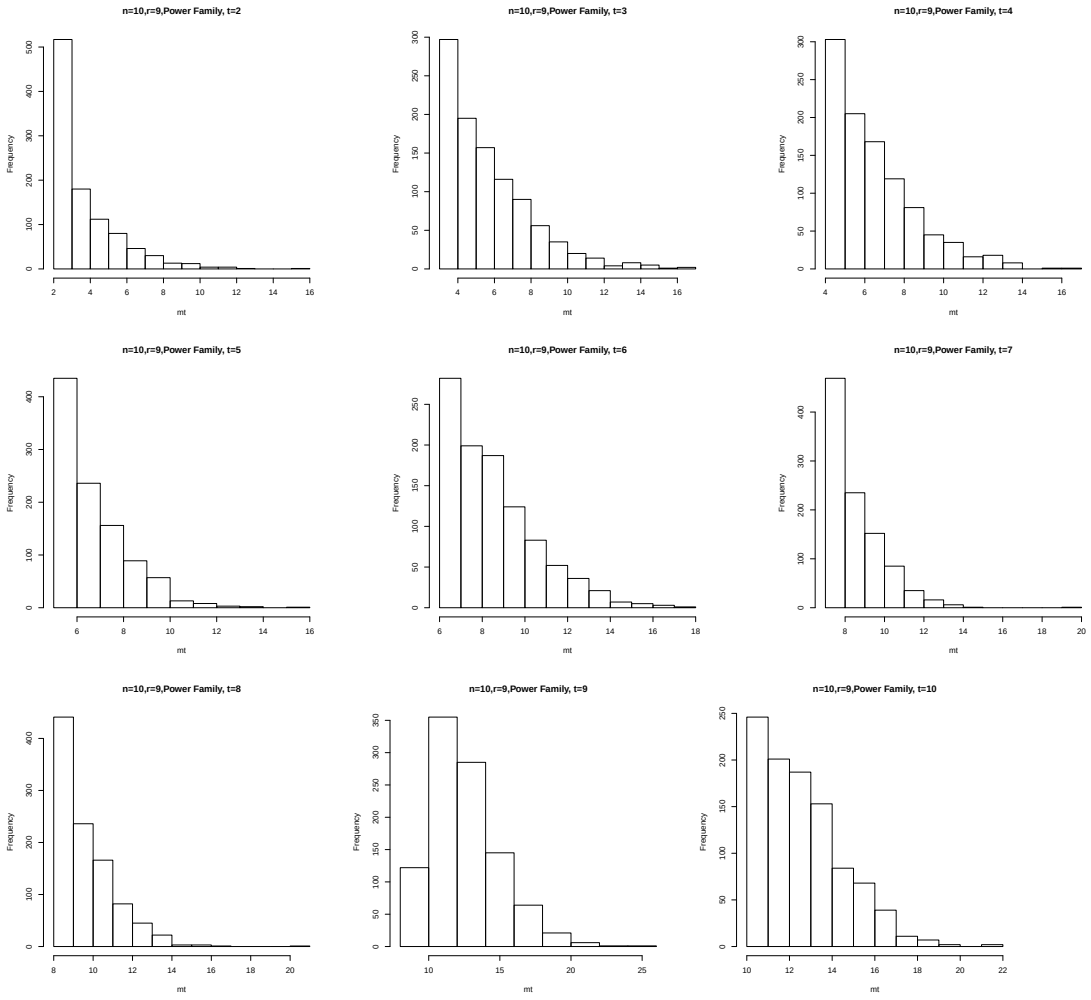

Figure 5.22: $\lambda_1 = 0.85$,$E(R) = 3$, $n = 10$, $r = 9$, $t = 2, \ldots, 10$

5.2 Clinical Trial Application

Bowman and George (1995) discuss an application to a clinical trial that compares two antibiotics for ear infections in children. Details of the study have been given in Mandel et al. 1982. Tables 5.24 and 5.25 show the data for the double-blind randomized clinical trial comparing cefaclor (CEF) and amoxicillian (AMO). These antibiotics are used for the treatment of acute otitis media (OME). Seventy five children have OME in both ears at the beginning of the study, and are randomly assigned to a 14-day treatment of CEF or AMO. X_1 is defined to be 1 if the right ear is clear at the 14^{th} day, 0 otherwise, and X_2 is similarly defined in terms of the left ear.

Table 5.24: # of ears cleared, CEF

0	1	2	Total
14	9	21	44

Table 5.25: # of ears cleared,AMO

0	1	2	Total
15	3	13	31

Without additional information, such as the effect of right ear- or left ear-specific covariates on the severity of ear infection, it is reasonable to assume X_1 and X_2 are exchangeable. For a child with bilateral OME infection, Mandel et al. (1982) estimated the probability that a specific ear (left or right) was infection free at 14 days to be 0.53. They also estimated the probability that this ear was infection free given that the other ear had no OME at 14 days to be 0.86. From these estimates, we get $P(X_1 = 0, X_2 = 0) = 0.456$ and $P(X_1 = 1, X_2 = 0) = P(X_1 = 0, X_2 = 1) = 0.014$, which not only indicate a high degree of dependence between X_1 and X_2, but also support an assumption of exchangeability between the two random variables. Therefore, in the analysis that follows we assume the CEF and AMO treatment groups consist of exchangeable pairs of binary observations.

There were a total of 44 children that were assigned CEF; 14 children had $r = 0$ ears clear after the 14 day treatment; 9 children had $r = 1$ ear clear; and 21 children had both ears clear, $r = 2$. I sampled a response from the CEF responses, if the response was 1 or 2 ($r \geq 1$) then $e = 1$ as described earlier.

Since there were 30 responses from the data set with one or more responses, I sampled from the CEF data one at a time until there were 30 clusters of size 2 with at least 1 response. The results are posted in Table 5.26. My $\hat{\lambda}_1$ after 100 such simulations averaged 0.5721 with standard error of 0.0729. Bowman and George (1995) estimated $\lambda_1 = 0.579$ with standard error 0.066. The number of clusters needed to get 30 clusters with one or more responses was 44.76 with standard error of 5.1701 compared to the number of clusters in the data of 44. If we lower the number of clusters from $t = 30$ to $t = 26$, because for 44 events with probability of success $\hat{\lambda} = 0.58$ would be $0.58 * 44 = 25.52$, the estimated value of λ_1 after 100 simulations is 0.5807 with standard error 0.0784. The number of clusters needed to get 26 clusters with one or more responses was 38.34 with standard error 4.2217. This shows our estimate gives approximately the same estimate for λ_1 and approximately the same standard error with fewer clusters needed.

For the AMO data, 31 children were assigned the treatment; 15 had $r = 0$, 3 had $r = 1$, and 13 had $r = 2$ responses. I sampled from the AMO responses, if the responses was $r \geq 1$ then $e = 1$. Since the data set had 16 responses of one or more from 31 clusters, I sampled one at a time until I had $t = 16$ clusters with $r \geq 1$ responses. The results are posted in Table 5.27. My $\hat{\lambda}_1$ after 100 such simulations averaged 0.4963 with standard error of 0.0902. Bowman and George (1995) estimated $\lambda_1 = 0.486$ with standard error 0.085. The average number of clusters needed to get 16 clusters with one or more responses was 30.3 with standard error of 5.9297 compared to the number of clusters in the data of 31.

To investigate the effect of lowering the number of clusters on our conditional estimator for λ_1, Tables 5.26 and 5.27 show how $\hat{\lambda}_1$ and it's standard error changes as the number of clusters decreased. Overall number of clusters needed to reach t of them having r or more responses is also recorded with it's standard error.

Table 5.26: Results of smaller cluster sizes, CEF treatment

t	$mean(\hat{\lambda}_1)$	$se(\hat{\lambda}_1)$	$mean(m_t)$	$se(m_t)$
30	0.5880	0.0654	44.0200	4.3992
26	0.5863	0.0696	38.1300	4.5185
24	0.5816	0.0727	35.7700	4.3620
22	0.5757	0.0753	32.8800	4.2218
20	0.5738	0.0793	30.0200	3.4786
18	0.5939	0.0830	26.3900	4.0098
16	0.5885	0.0884	23.3400	2.9889

Table 5.27: Results of smaller cluster sizes, AMO treatment

t	$mean(\hat{\lambda}_1)$	$se(\hat{\lambda}_1)$	$mean(m_t)$	$se(m_t)$
16	0.4852	0.0853	30.5600	5.0339
14	0.4761	0.0894	27.8100	6.1112
12	0.4928	0.0982	23.1900	5.3479
10	0.4844	0.1057	19.6700	4.6058

We can see that a small reduction of clusters did not affect the estimation of λ_1 or the standard error of $\hat{\lambda}_1$ for either treatment group. However, if reduced too much, the standard error begins to increase. The overall number of clusters needed also decreased without loss of information.

5.3 Conclusion

From the result of my research on an alternate design for estimating exchangeable binary data, our estimator is unbiased and a reliable estimator to estimate the probability of a response. It has the same or smaller standard error than previous estimators. This new design shows that a reduction could be made in the number of clusters needed to get the same results as previous methods. The motivation for this research was finding new methods that reduce the number of animals needed in developmental toxicity studies and finding estimators with smaller standard error than previous methods. Evidence has been shown that both of these could be possible by taking one cluster at a time, and observing clusters until you have a certain number with the desired number of responses (or more).

Further work is to be done in comparing the simulation results with the exchangeable binary data in Bowman and George (1995). The clinical data set did a fraction of the comparison, and showed favorable results for the estimator shown in this paper. Futher work needs to be done to see the full extent of the capabilities for this method.

Further work will also be done to extend this estimator to be used with unequal cluster sizes. As discussed, in actual developmental toxicity studies, cluster sizes are not equal. Estimation for unequal cluster sizes of exchangeable binary data can be extended to this situation. Then, if the number of clusters needed is reduced, the motivation of this paper will in fact be reached.

Bibliography

"About SOT." www.toxicology.org.The Society of Toxicology, May 23, 2002 (revised August 2, 2010). Web. Jan. 2017.

Bowman, D. and George E.O. (1995). A Saturated Model for Analyzing Exchangeable Binary Data: Applications to Clinical and Developmental Toxicity Studies. *Journal of the American Statistical Association* **90**, 871-879.

Bickel, P. and Doksum, K. (1977) *Mathematical Statistics: Basic Ideas and Selected Topics*. Holden-Day.

Chen,J.J. and Kodell, R.L.(1989). Quantitative Risk assessment for teratological effects. *Journal of the American Statistical Association* **84**,966 - 971.

Dette, H., Neumeyer, N., and Pilz, K. (2005). A Note on Nonparametric Estimation of the Effective Dose in Quantal Bioassay. *Journal of the American Statistical Association* **100**, 503-510.

Dette, H. and Scheder, R. (2010). A finite sample comparison of nonparametric estimates of the effective dose in quantal bioassay. *Journal of Statistical Computation and Simulation* **80**, 527-544.

Dette, H., Pepelyshev, A., and Wong W.K. (2009). Optimal designs for dose-finding experiments in toxicity studies. *Bernoulli* **15**, 124-145.

Emmeke A., Verhage, M., Veenvliet, J., Dolan, C., Van Der Sluis, S. (2014). A solution to dependency: using multilevel analysis to accommodate nested data. *Nature neuroscience* **17**, 491 - 496.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications, Vol. II, Second edition*. Wiley. New York.

Finney, D. J. (1964). *Statistical Method in Biological Assay 2nd ed.*. Griffin. London.

Freedman, D. and Diaconis, P. (1982). De Finetti's Theorem for Symmetric Location Families. *The Annals of Statistics* **10**, 184-189.

Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Cambridge, New York.

George, E.O. and Bowman, D.(1995). A Full Likelihood Procedure for Analysing Exchangeable Binary Data. *Biometrics* **51**, 521-523.

George, E.O., Cheon, K., Yuan, Y., and Szabo, A. (2016). On exchangeable multinomial distributions. *Biometrika* **103**, 397-408.

Hilbe, J.M. (2007). *Negative binomial regression*. Cambridge University Press, Cambridge, UK.

Kuk, A.Y.C, (2004). A litter-based approach to risk assessment in developmental toxicity studies via a power family of completely monotone functions. *Journal of the Royal Statistical Society. Series C(Applied Statistics)* **53**, 369-386.

Mandel, E.M., Bluestone, C.D., Rockette, H.E., Blatter, M.M., Reisinger, K.S., Wucher, F.P., and Harper.J. (1982). Duration of Effusion After Antibiotic Treatemtn for Acute Otitis Media: Comparison of Cefaclor adn Amoxicillin. *Pediatric Infectious Diseases* **1**, 310-316.

Morgan, B.J.T (1992). *Analysis of Quantal Response Data*. Chapman and Hall. London, U.K.

Mukhopadhyay, S.(2000). Bayesian Nonparametric Inference on the Dose Level with Specified Response Rate. *Biometrics* **56**, 220-226.

Pang, Z., and Kuk, A.Y.C. (2007). Test of Marginal Compatibility and Smoothing Methods for Exchangeable Binary Data With Unequal Cluster Sizes. *Biometrics* **63**, 218-227.

Parasuraman, S. (2011). Toxicological screening. *Journal of Pharmacology and Pharmacotherapeutics* **2**, 74-79.

Prentice, R.L. (1976). A Generalization of the Probit and Logit Methods for Dose Response Curves. *Biometrics* **32**, 761-768.

Prentice, R.L. (1986). Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by covariate measurement errors. *Journal of the American Statistical Association* **81**, 321-327.

Ryan,L. (1992). Quantitative risk assessment for developmental toxicity. *Biometrics* **48**, 163-174.

Stefanescu, C. and Turnbull, B. (2003). Likelihood Inference for Exchangeable Binary Data with Varying Cluster Sizes. *Biometrics* **59**, 18-24.

Szabo, A. and George, E.O. (2010). On the use of stochastic ordering to test for trend with clustered binary data. *Biometrika* **97**, 95-108.

Tan, F. Rayner, G.J., Wang, X., and Peng, H. (2010). A full likelihood procedure of exchangeable negative binomials for modelling correlated and overdispersed count data. *Journal of Statistical Planning and Inference* **140**, 2849-2859.

U.S. Food and Drug Administration. (1993). Toxicological Principles for the Safety Assessment of Direct Food Additives and Color Additives Used in Food, Redbook 2000: IV.C.9.b Guidelines for Developmental Toxicity Studies. U. S. Food and Drug Administration, Center for Food Safety and Applied Nutrition, Washington, DC.

Williams, D.A. (1975). The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. *Biometrics* **31**, 949-952.

Williams, D.A. (1988). Estimation bias using the beta-binomial distribution in teratology. *Biometrics* **44**, 305-309.

Xu, J.-L. and Prorok, P.C. (2003). Modeling and analyzing exchangeable binary data with random cluster sizes. *Statistics in Medicine* **22**, 2401–2416.

Yuan, Y. and Yin, G. (2011). Dose-Response Curve Estimation: A Semiparametric Mixture Approach. *Biometrics* **67**, 1543-1554.

Zeger, S.L. and Liang, K. (1986). Longitundinal data analysis for discrete and continuous outcomes. *Biometrics* **42**, 121-130.