A Statistical Test for a Parameter Describing the Association in Correlated Binary Data

by

Jordan David Bernick, B. Math (Carleton)

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfillment of
the requirement for the degree of
Master of Science

School of Mathematics and Statistics
Ottawa-Carleton Institute for Mathematics and Statistics
Carleton University
Ottawa, Ontario
August, 2007

© Copyright
2007, Jordan David Bernick
NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.
Abstract

In this thesis we develop a test of statistical hypothesis for the association parameters in the non-linear dynamic conditional probability model (NLDPCP) suggested by Farrell and Sutradhar (2006) for longitudinal correlated binary data. In the development of such a test we consider the simplest case for the above model where the probability associated with the response at a particular time is only conditionally related to the previous response. We then derive the Rao-Fisher scoring test statistic as well as both its asymptotic and non-asymptotic probability distributions, and derive a power function for a decision rule based on the above test.
Acknowledgments

I would like to express my thanks to my supervisor, Dr Patrick Farrell. Without his help and guidance this paper would have been nothing but a figment of my imagination. Dr Farrell has enriched my understanding of the field of statistics and has introduced me to the wonderful adventure of academic research. For all of that I am for ever thankful. I would also like to thank my parents for their help, support and understanding throughout the years since if it were not for them I doubt I would have made it this far.
Table of Contents

Abstract ii

Acknowledgments iii

1 Introduction 1

2 Correlation for the General Lag Model 8

3 Correlation for the "Lag 1" Model 11

4 Derivation of the Rao Score Test 17

5 Recursive Relation of the Marginal Means 30

6 Power Functions: Asymptotic Distribution Test Statistic
   6.1 Power Function: Two Time Periods 33
   6.2 Power Function: Three Time Periods 41
7 Small Sample Inference

7.1 Derivation of the Distribution of the Sum of n Independent Non Identically Distributed Bernoulli Random Variables

7.2 Derivation of the Distribution of the Rao Score Statistic: Two Time Periods

7.3 Methodology for the Derivation of the Distribution of the Rao Score Statistic: Three Time Periods

8 A Rao Score Test for the Model Containing a Covariate

9 Conclusion and Discussion

10 References

11 Appendix

11.1 Minitab Code to Calculate Power of the Test Based on $\sum_{i=1}^{n} Y_{1i} Y_{2i}$ for the Case of $T = 2$, $n = 25$, $\gamma_0 = 0$

11.2 Macro for the Power of the Rao Score Test when $T=2$

11.3 Macro for the Correlation for the Lag 1 Model where $T=2$
Introduction

There has been great interest and much attention paid to longitudinally correlated binary data for quite some time. Correlated longitudinal binary data occurs in many important empirical studies in which a certain binary outcome is measured on the same experimental unit over time. In many cases it is assumed that these data occur in independent clusters, where within each cluster the measurements are correlated. One such example is a clinical trial where an independent sample of patients (clusters) may be selected and after which each patient would be tested for the presence or absence of a certain disease at different points in time. Such an illustration can be found in Ware et al (1984) where the effect of a mother’s smoking behavior on wheezing problems in children was examined. In this study an independent sample of 537 mother-child pairs was selected, and data signifying whether or not the mother smoked, and whether or not the child wheezed was collected annually over a four-year period. Other examples include Sashegyi et al. (2000) and Sutradhar and Farrell (2004), who considered the effect over time of intervention programs on the smoking status of elementary and high school students.

The development and estimation of parameters in models for longitudinally correlated binary data have been dealt with by several authors who proposed several different techniques and approaches. Fitzmaurice and Laird (1993) suggested a likelihood-based approach for the estimation of the model parameters, while Shults and Chaganty (1998) used a quasi-least squares technique. Sutradhar and Das (1999) and Sutradhar (2003) proposed an estimation procedure based on generalized quasi-likelihood (GQL), while Chaganty and Joe (2004) employed an approach that is based on generalized estimation equations (GEE).
Due to the complexity of these models and the approaches for their estimation, it is quite
difficult to derive the properties of the estimators for the model parameters. One way to try and
investigate these properties is to use Monte Carlo simulation studies. However since these
simulations require a mechanism for generating longitudinally correlated binary response data,
several authors have focused on developing such data generating methods. One of the problems
encountered in such data generating methods is that of achieving full range for the correlation
between the binary responses, since in many cases the permitted range for the correlation parameters
is highly restricted.

Recently Farrell and Rogers-Stewart (2007) have presented a review that summarizes methods
for generating longitudinally correlated binary data. We will briefly describe some of these methods
here. In what follows, let $\tilde{Y} = \begin{bmatrix} Y_1 & Y_2 & \ldots & Y_T \end{bmatrix}'$ represent a $T$-dimensional vector of binary
responses with mean $\vec{\mu} = \begin{bmatrix} \mu_1 & \mu_2 & \ldots & \mu_T \end{bmatrix}'$, assume also that $\rho_{t,t'}$ reflects the correlation
between $Y_t$ and $Y_{t'}$ where $t \neq t'$.

One method for generating longitudinally correlated binary data is based on the work of
Bahadur (1961) who proposed a representation of a multivariate binary distribution. When all
coefficients of order three or higher are ignored, the representation reduces to

$$f(y_1, \ldots, y_T) = \left\{ \prod_{t=1}^{T} \mu_t^i (1 - \mu_t)^{1-y_t} \right\} \left\{ 1 + \sum_{t \neq t'} \rho_{t,t'} \tilde{y}_t \tilde{y}_{t'} \right\},$$

where $\tilde{y}_t = (y_t - \mu_t) / \sqrt{\mu_t (1 - \mu_t)}$. 
Unfortunately, Bahadur’s (1961) representation is computationally difficult to manipulate when the number of time periods considered gets large. Also note that the parameters $\rho_{it}$ are constrained since $f(y_1, \ldots, y_T)$ can only range between zero and one.

Kanter (1975) proposed a model designed to only generate correlated binary responses according to a stationary autoregressive process of an order $p$. Initially, a value for the first binary response is generated from a Bernoulli distribution with parameter $\mu$ and then the next value is generated according to an autoregressive process of an order $\min(t-1, p)$ that depends on the time $t-1$ of the binary response $Y_{t-1}$. For example, after a value of $Y_1$ is obtained, the generation of AR(1) binary data with mean $\mu$ and $\rho_{it*} = \rho$ for all $t, t^* = 1, \ldots, T$ and $t \neq t^*$ would proceed according to the model

$$Y_t = U_t(Y_{t-1} \oplus W_t) + (1 - U_t)W_t$$

where $\oplus$ represents addition modulo 2. In addition, $U_t$ and $W_t$ would be generated as Bernoulli random variables with parameters $\xi$ and $\eta$ where these parameters are linked to parameters $\mu$ and $\rho$ by the formulae $\rho = [(\eta - \mu)(1 - 2\mu)(1 - \eta)/(2\eta - 1)\mu(1 - \mu)]$ and $\xi = (\mu - \eta)/[\eta(1 - 2\eta)]$. Note that this again imposes constraints on the possible values which $\rho$ can take on, since $\eta$ and $\xi$ are bounded between zero and one.
McKenzie (1981) extended the approach of Kanter (1975) to stationary ARMA models. However both the procedure of Kanter (1975) and McKenzie's (1981) extension are not appropriate for non-stationary processes and in addition the ranges permitted for the longitudinal association parameters are quite restrictive.

Emrich and Piedmonte (1991) proposed a method in which a multivariate normal random vector is first generated, and subsequently transformed in order to obtain a binary sequence, while Gange (1995) suggests the use of an interactive proportional fitting algorithm to generate a correlated binary sequence. Park et al. (1996) developed a method where the binary sequence is generated by making use of a linear combination of \( m \) independent Poisson variables. The method is computationally efficient even when dealing with large time periods, and can accommodate non-stationary, unpatterned correlation. However, only positive correlations are permitted.

Lunn and Davis (1998) introduce an efficient model that generates stationary binary response data that are correlated either as a stationary AR(1) or an exchangeable process with parameter \( p \). The approach is computationally efficient for large time periods, but unfortunately it is extremely restrictive in terms of the types of correlation it can accommodate, as it is only appropriate for a few stationary structures and positive association.

Qaqish (2003) proposed generating a binary sequence by using the conditional distribution for the response at time \( t \), \( Y_t \) given the previous responses \( Y_{t-1}, ..., Y_1 \):

\[
\lambda_t = P(Y_t = 1|Y_{t-1}, ..., Y_1) = \mu_t + \sum_{r=1}^{t-1} b_r (Y_r - \mu_r)
\]
where \( b_{n\tau} \) represents \( b_{n\tau} = [\text{Cov}(\vec{Y}_{t-1})]^{-1}\text{Cov}(Y_t, \vec{Y}_{t-1}) \) with

\[
\text{Cov}(\vec{Y}_{t-1}) = \begin{bmatrix}
\text{Var}(Y_1) & \text{Cov}(Y_{t-1}, Y_1) \\
\text{Cov}(Y_1, Y_2) & \text{Var}(Y_2) \\
\vdots & \vdots \\
\text{Cov}(Y_1, Y_{t-1}) & \text{Var}(Y_{t-1})
\end{bmatrix}
\]

being the covariance matrix for the vector \( \vec{Y}_{t-1} \) and \( \text{Cov}(Y_t, \vec{Y}_{t-1}) = \begin{bmatrix}
\text{Cov}(Y_1, Y_t) & \text{Cov}(Y_2, Y_t) & \ldots & \text{Cov}(Y_{t-1}, Y_t)
\end{bmatrix}^T \) is the \((t-1)\) dimensional vector of covariances of \( Y_t \). This method permits unpatterned correlation and often allows for wider ranges for the correlation parameters than other approaches. However due to the linear relationship between the values of \( b_{n\tau} \) and the conditional probability statement for \( \lambda_t \), which must naturally take on values between zero and one, the values for \( b_{n\tau} \) and accordingly \( \rho_{n\tau} \) are still constrained. For example, for a non stationary AR(1) process having correlation parameter \( \rho \), Qaqish (2003) showed that

\[
b_{n\tau} = \frac{\rho/[1 + (t - 2)\rho]}{\sqrt{\nu_{n\tau}}} = \nu_{n\tau}, \quad \nu_{n\tau} = \mu_t(1 - \mu_t).\]

Therefore, since the values of \( b_{n\tau} \) are constrained as \( 0 < \lambda_t < 1 \), it follows that the value of \( \rho \) will be constrained as well.

Recently Farrell and Sutradhar (2006) proposed a method for generating non-stationary longitudinal binary data that allows for full ranges of the association parameters, as well as accommodating unpatterned correlation structures. This method is appropriate for data measured over a small number of time periods. Rather than the linear model of Qaqish (2003), their model describes the conditional probability \( \lambda_t \) as a function of previous responses and covariates with the use of a logit link.
\[ P(Y_t = 1|y_{t-1}, ..., y_1) = \log \frac{\tilde{y}_t^{\tilde{\beta}}}{1 + \tilde{y}_t^{\tilde{\beta}}} + \gamma_1 y_{t-1} + ... + \gamma_{t-1} y_1 \]

where \( \tilde{x}_t = \left[ \begin{array}{c} x_{t1} & x_{t2} & \ldots & x_{tr} \end{array} \right]' \) is a vector of \( r \) covariates and \( \tilde{\beta} = \left[ \begin{array}{c} \beta_1 & \beta_2 & \ldots & \beta_r \end{array} \right]' \) is a vector of regression parameters. Using a simpler case of their model where the conditional mean of \( Y_t \) is given by

\[ P(Y_t = 1|y_{t-1}) = \log \frac{\tilde{y}_t^{\tilde{\beta}}}{1 + \tilde{y}_t^{\tilde{\beta}}} \]

Farrell and Sutradhar (2006) derive a general expression for the correlation between \( Y_t \) and \( Y_{t-1} \), showing that full range between -1 and +1 is permitted. We henceforth refer to this expression as a lag 1 model.

In this thesis we further investigate the model proposed in Farrell & Sutradhar (2006). Initially we develop formulae for the general lag model where the conditional mean of \( Y_t \) is described as

\[ P(Y_t = 1|y_{t-1}, ..., y_1) = \log \frac{\tilde{y}_t^{\tilde{\beta}}}{1 + \tilde{y}_t^{\tilde{\beta}}} + \gamma_1 y_{t-1} + ... + \gamma_{t-1} y_1 \]

and then demonstrate the derivation of the moments of \( Y_t \) for the "lag 1 model". With the use of mathematical induction we also verify the results concerning the covariance and correlation of the "lag 1 model" stated in Farrell & Sutradhar (2006). In addition, with the use of a Rao-Fisher scoring test of hypothesis we draw inference about the parameter \( \gamma \) when dealing with the simplest case of the model where conditional mean of the binary response \( Y_t \) depends only on the previous response \( Y_{t-1} \), i.e. we develop several different variations of a test of statistical hypothesis for \( \gamma \) based on various versions of the logistic regression model for \( P(Y_t = 1|y_{t-1}) \). In addition a power function is
derived for the decision rule based on the asymptotic distribution of the derived test statistic for the cases where repeated measurements are taken over two or three time periods. For small sample inference, the actual distribution of the test statistic is also derived for the case where two time periods are considered, and a methodology for deriving the distribution of the statistic when three time periods are considered is explained as well.
2 Correlation for the General Lag Model

One possible model for expressing the probability distribution of a time dependent sequence of binary random responses has been proposed by Farrell and Sutradhar (2006). Specifically, consider \( i \) independent clusters where the \( i^{th} \) cluster is composed of \( T \) time dependent binary responses. In other words the random vector \( \vec{Y}_i = [Y_{i1}, Y_{i2}, \ldots, Y_{iT-1}, Y_{iT}]' \) will represent the \( i^{th} \) cluster such that \( Y_{it} \) is independent of \( Y_{ij} \) for any two differing clusters \( i \) and \( j \). Within a given \( i^{th} \) cluster however we will use the following conditional probability model proposed by Farrell and Sutradhar (2006) as a possible way of expressing the dependency between binary responses within a given structure. Thus by letting:

\[
P(Y_{it} = 1|Y_{i,t-1}, \ldots, Y_{i1}) = \log i(\chi_{it}'\hat{\beta} + \gamma_1Y_{i,t-1} + \ldots + \gamma_{t-1}Y_{i1}) = \lambda_{it}
\]

we let \( \lambda_{it} \) be the conditional probability that within cluster \( i \), the \( t^{th} \) binary response \( Y_{it} \) takes on the value 1, given all previous random responses within the cluster.

Farrell and Sutradhar (2006) only investigated a simpler form of this model where

\[
P(Y_{it} = 1|Y_{i,t-1}) = \log i(\chi_{it}'\hat{\beta} + \gamma_1Y_{i,t-1}).
\]

Thus we derive here the general form for the correlation within a given cluster. Since we begin by doing so for a particular cluster, we suppress the notation for cluster, dropping the \( i \) subscript.

First let us write down the expression for the joint distribution of the binary responses within a particular cluster.
\[ P(y_T, \ldots, y_1) = P(y_T|y_{T-1}, \ldots, y_1)P(y_{T-1}|y_{T-2}, \ldots, y_1) \ldots P(y_2|y_1)P(y_1) \]

\[ = \lambda_T^{y_T}(1 - \lambda_T)^{1-y_T} \lambda_{T-1}^{y_{T-1}}(1 - \lambda_{T-1})^{1-y_{T-1}} \lambda_{T-2}^{y_{T-2}}(1 - \lambda_{T-2})^{1-y_{T-2}} \ldots \lambda_2^{y_2}(1 - \lambda_2)^{1-y_2} \lambda_1^{y_1}(1 - \lambda_1)^{1-y_1} \]

\[ = \prod_{j=1}^{T} \lambda_j^{y_j}(1 - \lambda_j)^{1-y_j} \]

since

\[ P(y_t|y_{t-1}, \ldots, y_1) = \lambda_t^{y_t}(1 - \lambda_t)^{1-y_t} \]

and

\[ P(Y_1 = 1) = \mu_1 = \lambda_1. \]

Now, the correlation within a given cluster between binary random responses \( Y_t \) and \( Y_u \) that we wish to determine is given by

\[ Corr(Y_t, Y_u) = \frac{Cov(Y_t, Y_u)}{sd(Y_t)sd(Y_u)} \]

where \( sd(Y_t) \) represents the standard deviation of the binary response \( Y_t \), which is written as

\[ sd(Y_t) = \sqrt{\mu_t(1 - \mu_t)} \]

since \( Y_t \) is marginally distributed as a Bernoulli random variable with parameter \( \mu_t \) corresponding to the probability \( \mu_t = P(Y_t = 1) \). Note also that

\[ Cov(Y_t, Y_u) = E(Y_tY_u) - E(Y_t)E(Y_u) = P(Y_t = 1, Y_u = 1) - \mu_t\mu_u \]

is the covariance of \( Y_t \) and \( Y_u \).
To write out an alternative form for the expression $Cov(Y_t, Y_u)$, note that the marginal probability $P(Y_t = 1)$ can be obtained by taking the summation of the joint probability $P(Y_t = 1, y_{t-1},..., y_1)$ over all other random responses $Y_{t-1},..., Y_1$ preceding $t$ in the clusters.

Specifically we do so in the following manner:

$$
\mu_t = P(Y_t = 1) = \sum_{y_{t-1}=0}^{1} \sum_{y_1=0}^{1} P(Y_t = 1, y_{t-1},..., y_1) = \sum_{y_{t-1}=0}^{1} \sum_{y_1=0}^{1} \prod_{j=1}^{t} \lambda_j^{y_j} (1 - \lambda_j)^{1-y_j} \big|_{Y_1=1}
$$

Next, for $t \geq u$, we calculate the expectation of the cross product $Y_t Y_u$ in the same way that we do for the marginal probability above:

$$
E(Y_t Y_u) = P(Y_t = 1, Y_u = 1) = \sum_{y_{t-1}=0}^{1} \sum_{y_u=0}^{1} P(Y_t = 1, y_{t-1},..., y_u = 1, y_1)
$$

$$
= \sum_{y_{t-1}=0}^{1} \sum_{y_u=0}^{1} \prod_{j=1}^{t} \lambda_j^{y_j} (1 - \lambda_j)^{1-y_j} \big|_{Y_1=1, Y_u=1}
$$

where the notation "|_{Y_1=1, Y_u=1}" means that we substitute in $Y_t = 1$ and $Y_u = 1$ anywhere they appear in the above expression. Therefore the covariance between $Y_u$ and $Y_t$ is thus defined, and we can express the correlation between these two binary responses as

$$
Corr(Y_t, Y_u) = \frac{\mu_t (1 - \mu_t) \mu_u (1 - \mu_u)}{\frac{1}{2} \left[ \sum_{y_{t-1}=0}^{1} \sum_{y_u=0}^{1} \prod_{k=1}^{t} \lambda_k^{y_k} (1 - \lambda_k)^{1-y_k} \big|_{Y_1=1, Y_u=1} - \mu_t \mu_u \right]}
$$

where $t \geq u$. 
3. Correlation for the "Lag 1" Model

This general form for the correlation specified at the end of Chapter 2 should simplify to the expression derived by Farrell and Sutradhar (2006) where the conditional mean of $Y_t$ is only a function of a vector of covariates, and the previous binary response $Y_{t-1}$. Specifically, letting:

$$\lambda_t = P(Y_t = 1|Y_{t-1}) = \log \mu(x_t \beta + \gamma y_{t-1})$$

we get the following conditional probabilities

$$p_{11} = P(Y_t = 1|Y_{t-1} = 1) = e^{x_t \beta + \gamma y_{t-1}}/(1 + e^{x_t \beta + \gamma y_{t-1}})$$

$$p_{10} = P(Y_t = 1|Y_{t-1} = 0) = e^{x_t \beta}/(1 + e^{x_t \beta})$$

Generally speaking the marginal probability of $Y_t = 1$ at time $t = 1$ is defined as

$$\mu_t = P(Y_t = 1).$$

similar to Farrell and Sutradhar (2006), if we let

$$\mu_1 = P(Y_1 = 1) = e^{x_1 \beta}/(1 + e^{x_1 \beta})$$

it is possible to compute the marginal probability $\mu_2$, that $Y_2 = 1$ under the above model as follows

$$\mu_2 = P(Y_2 = 1) = P(Y_2 = 1|Y_1 = 1)P(Y_1 = 1) + P(Y_2 = 1|Y_1 = 0)P(Y_1 = 0)$$

$$= p_{21} \mu_1 + p_{20}(1 - \mu_1) = p_{20} + \mu_1(p_{21} - p_{20}).$$

In general, using the same approach, it can be shown that

$$\mu_t = p_{t0} + \mu_{t-1}(p_{t1} - p_{t0})$$

for $t = 2, ..., T$. 

11
To derive an expression for the covariance between \( Y_u \) and \( Y_t \) \((t > u)\), note that

\[
Cov(Y_u, Y_t) = E(Y_u, Y_t) - E(Y_u)E(Y_t) = P(Y_t = 1, Y_u = 1) - \mu_u \mu_t
\]

where

\[
E(Y_u, Y_t) = \sum_{y_{t-1}=0}^{1} \cdots \sum_{y_{u+1}=0}^{1} P(Y_t = 1, y_{t-1}, \ldots, y_{u+1}, Y_u = 1)
\]

\[
= \sum_{y_{t-1}=0}^{1} \cdots \sum_{y_{u+1}=0}^{1} P(Y_t = 1|y_{t-1})P(y_{t-1}|y_{t-2})\ldots P(y_{u+1}|y_{u})P(y_{u}) = 1)P(Y_u = 1)
\]

\[
= \sum_{y_{t-1}=0}^{1} \cdots \sum_{y_{u+1}=0}^{1} P(Y_t = 1|y_{t-1})P(y_{t-1}|y_{t-2})\ldots P(y_{u+1}|y_{u})P(y_{u}) = 1)\mu_u
\]

\[
= \mu_u \sum_{y_{t-1}=0}^{1} \cdots \sum_{y_{u+1}=0}^{1} P(Y_t = 1|y_{t-1})P(y_{t-1}|y_{t-2})\ldots P(y_{u+1}|y_{u}) = 1)
\]

\[
= \mu_u A
\]

where

\[
A = \sum_{y_{t-1}=0}^{1} \cdots \sum_{y_{u+1}=0}^{1} P(Y_t = 1|y_{t-1})P(y_{t-1}|y_{t-2})\ldots P(y_{u+1}|y_{u}) = 1).
\]

Substituting, this suggests that

\[
Cov(Y_u, Y_t) = \mu_u \sum_{y_{t-1}=0}^{1} \cdots \sum_{y_{u+1}=0}^{1} P(Y_t = 1|y_{t-1})P(y_{t-1}|y_{t-2})\ldots P(y_{u+1}|y_{u}) = 1)
\]

\[-\mu_u \mu_t
\]

\[
= \mu_u A - \mu_u \mu_t = \mu_u (A - \mu_t)
\]
and so
\[ \text{Cov}(Y_u, Y_i) = \mu_u(A - \mu_{i+1}). \]

**THEOREM 3.1:**

Under the above model

\[ \text{Cov}(Y_u, Y_{w+j}) = \mu_u(1 - \mu_u) \prod_{k=w+1}^{w+j} (p_{t1} - p_{t0}). \]

**PROOF:**

We shall prove the above result using mathematical induction. First, we prove the result for \(j=1\), where we consider

\[ \text{Cov}(Y_u, Y_{u+1}) = P(Y_{u+1} = 1, Y_u = 1) - \mu_u \mu_u = p_{u+1,1} \mu_u - \mu_u = \mu_u (p_{u+1,1} - \mu_u) \]

\[ = \mu_u [p_{u+1,1} - p_{u+1,0} - \mu_u (p_{u+1,1} - p_{u+1,0})] \]

\[ = \mu_u (1 - \mu_u) (p_{u+1,1} - p_{u+1,0}) = \mu_u (1 - \mu_u) \prod_{k=w+1}^{w+1} (p_{t1} - p_{t0}). \]

Hence, the result is proven for \(j=1\).

Now, assume that \( \text{Cov}(Y_u, Y_{w+j}) = \mu_u (1 - \mu_u) \prod_{k=w+1}^{w+j} (p_{t1} - p_{t0}) \) is true and show that based on this fact

\[ \text{Cov}(Y_u, Y_{w+j+1}) = \mu_u (1 - \mu_u) \prod_{k=w+1}^{w+j+1} (p_{t1} - p_{t0}) \]

is also true. This covariance between \(Y_u\) and \(Y_{w+j+1}\) is given by

\[ \text{Cov}(Y_u, Y_{w+j+1}) = \mu_u \sum_{y_{w+j}}^{1} \sum_{y_{w+j-1}}^{0} P(Y_{w+j+1} = 1|y_{w+j})P(y_{w+j}|y_{w+j-1})...P(y_{w+1}|Y_u = 1) \]
where

\[ \mu_{w+j+1} = \mu_{w+j+1,1} \mu_{w+j} + \mu_{w+j+1,0} (1 - \mu_{w+j}). \]

If we let

\[ \sum_{\omega=0}^{1} \sum_{\nu=0}^{1} P(Y_{w+j} = 1 | Y_{w+j-1}) P(Y_{w+j-1} | Y_{w+j-2}) \ldots P(Y_{w+1} | Y_u = 1) \]

expanding

\[ \sum_{\omega=0}^{1} \sum_{\nu=0}^{1} P(Y_{w+j} = 1 | Y_{w+j-1}) P(Y_{w+j-1} | Y_{w+j-2}) \ldots P(Y_{w+1} | Y_u = 1) \]

over \( Y_{w+j} \) yields

\[ \sum_{\omega=0}^{1} \sum_{\nu=0}^{1} P(Y_{w+j} = 1 | Y_{w+j-1}) P(Y_{w+j-1} | Y_{w+j-2}) \ldots P(Y_{w+1} | Y_u = 1) \]

\[ = \mu_{w+j+1,1} \mu_{w+j} + \mu_{w+j+1,0} \sum_{\omega=0}^{1} \sum_{\nu=0}^{1} P(Y_{w+j} = 0 | Y_{w+j-1}) P(Y_{w+j-1} | Y_{w+j-2}) \ldots P(Y_{w+1} | Y_u = 1) \]

\[ = \mu_{w+j+1,1} \mu_{w+j} + \mu_{w+j+1,0} \sum_{\omega=0}^{1} \sum_{\nu=0}^{1} (1 - P(Y_{w+j} = 1 | Y_{w+j-1})) P(Y_{w+j-1} | Y_{w+j-2}) \ldots P(Y_{w+1} | Y_u = 1). \]

Thus

\[ \text{Cov}(Y_{w+j+1}, Y_{w+j+1}) \]

\[ = \mu_{w+j+1,1} \mu_{w+j} + \mu_{w+j+1,0} \sum_{\omega=0}^{1} \sum_{\nu=0}^{1} (1 - P(Y_{w+j} = 1 | Y_{w+j-1})) P(Y_{w+j-1} | Y_{w+j-2}) \ldots P(Y_{w+1} | Y_u = 1) \]

\[ - \mu_{w+j+1,0} \mu_{w+j+1,1} \mu_{w+j} - \mu_{w+j+1,0} \mu_{w+j+1,0}. \]
\[ = \mu_u [A p_{w+j+1,1} - p_{w+j+1,0} A + p_{w+j+1,0} \sum_{y_{w+j-1}=0}^{1} \ldots \sum_{y_{w+1}=0}^{1} P(y_{w+j-1}|y_{w+j-2}) P(y_{w+j}|Y_u = 1) \\
- p_{w+j,1,0} - \mu_{w+j}(p_{w+j+1,1} - p_{w+j+1,0})]. \]

However
\[
\sum_{y_{w+j-1}=0}^{1} \ldots \sum_{y_{w+1}=0}^{1} P(y_{w+j-1}|y_{w+j-2}) \ldots P(y_{w+1}|Y_u = 1) = 1
\]
since we are taking the sum over the support of the density. Hence
\[
Cov(Y_u, Y_{w+j}) = \mu_u [A (p_{w+j+1,1} - p_{w+j+1,0}) - (p_{w+j+1,1} - p_{w+j+1,0}) \mu_{w+j}]
\]
\[
= \mu_u (p_{w+j+1,1} - p_{w+j+1,0}) (A - \mu_{w+j}) = (p_{w+j+1,1} - p_{w+j+1,0}) Cov(Y_u, Y_{w+j}).
\]

However we assumed that
\[
Cov(Y_u, Y_{w+j}) = \mu_u (1 - \mu_u) \prod_{t=1}^{w+j} (p \varphi - p \theta)
\]
and as a result
\[
Cov(Y_u, Y_{w+j+1}) = (p_{w+j+1,1} - p_{w+j+1,0}) \mu_u (1 - \mu_u) \prod_{t=1}^{w+j} (p \varphi - p \theta) = Cov(Y_u, Y_{w+j+1})
\]
\[
= \mu_u (1 - \mu_u) \prod_{t=1}^{w+j+1} (p \varphi - p \theta)
\]
which is what we wanted to show. Therefore, Theorem 3.1 is true for all \( j \) by the principle of mathematical induction. Thus, a much simpler expression arises for the covariance between the binary responses within a cluster under the "lag 1" model than under the generic form. Finally, note
that since $\text{Var}(Y_u) = \mu_u(1 - \mu_u)$,

$$\text{Cov}(Y_u, Y_{uj}) = \text{Var}(Y_u) \prod_{j=1}^{u-1} (p_{ij} - p_{i0})$$

Therefore

$$\frac{[\text{Cov}(Y_u, Y_{uj})] \text{Var}(Y_u)}{\text{Var}(Y_u)} = \prod_{j=1}^{u-1} (p_{ij} - p_{i0}) \leq 1$$

which leads to the conclusion that:

$$\text{Cov}(Y_u, Y_{uj}) \leq \text{Var}(Y_u)$$
4. Derivation of the Rao Score Test

Continuing on with the "lag-1" model where for the $i^{th}$ cluster $P(Y_{it} = l|y_{i,t-1}) = \logit(x_{it} + \gamma y_{i,t-1}) = \lambda_{it}$ imagine that of central interest is the extent or significance of the association between the binary responses within a particular cluster. We can assess the magnitude of this association through the parameter $\gamma$ in the above model. We shall now explore this further.

However, in doing so we begin by dropping the $x_{it}$ term in the model for reasons of simplification so that we can focus on the $\gamma$ term. Thus, the model we initially consider is

$$P(Y_{it} = l|y_{i,t-1}) = e^{\gamma y_{i,t-1}}/(1 + e^{\gamma y_{i,t-1}}) = \lambda_{it}. \tag{4.1}$$

We consider a test of hypothesis of $H_0: \gamma = \gamma_0$ versus $H_a: \gamma = \gamma_1$. One particular hypothesis of interest would be $\gamma = 0$ under the null, as this would suggest a lack of significant association among the binary responses in a cluster.

As a motivating example of such a model and test of hypothesis on the association parameter $\gamma$, consider a matched pairs experiment in which a binary response is observed on each of $n$ units at each of $T = 2$ distinct points in time. Of interest, say, is whether or not the probability of observing one of the two outcomes at the first time point is the same as the probability of observing one of the two outcomes at the second time point. As an illustration of such a scenario, Agresti (2002) presents the results of a poll designed to assess the extent of changes in the views of British citizens towards the performance of the Prime Minister over time. Specifically, a sample of $n$ citizens was selected at a particular point in time and each person indicated whether they approved or disapproved of the
performance of the Prime Minister. Six months later, the same $n$ citizens were surveyed as to whether they approved or disapproved. The responses at the two different time periods were cross-classified into a 2x2 contingency table as follows:

<table>
<thead>
<tr>
<th></th>
<th>Time 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Approve</td>
<td>$n_{11}$</td>
<td>$n_{10}$</td>
</tr>
<tr>
<td>Disapprove</td>
<td>$n_{01}$</td>
<td>$n_{00}$</td>
</tr>
<tr>
<td></td>
<td>$n_{+1}$</td>
<td>$n_{+0}$</td>
</tr>
</tbody>
</table>

Of particular interest here is perhaps whether the approval rating for the Prime Minister has changed over time. Therefore, on the population level suppose we let

\[ P(Y_{i1} = 1, Y_{i2} = 1) = \pi_{11} \]
be the joint probability that the $i^{th}$ citizen in the population approves of the Prime Minister at time 1 and also approves at time 2.

\[ P(Y_{i1} = 1, Y_{i2} = 0) = \pi_{10} \]
be the joint probability that the $i^{th}$ citizen in the population approves of the Prime Minister at time 1 and disapproves at time 2.

\[ P(Y_{i1} = 0, Y_{i2} = 1) = \pi_{01} \]
be the joint probability that the $i^{th}$ citizen in the population disapproves of the Prime Minister at time 1 and approves at time 2.

\[ P(Y_{i1} = 0, Y_{i2} = 0) = \pi_{11} \]
be the joint probability that the $i^{th}$ citizen in the population disapproves of the Prime Minister at time 1 and also disapproves at time 2.

Analogous to the table above then, on the population level, we have the following
where $\pi_{1+} = P(Y_1 = 1)$ is the marginal probability of approving the Prime Minister at time 1, and $\pi_{2+} = P(Y_2 = 1)$ is the marginal probability of approving the Prime Minister at time 2.

Using this notation, if we wish to test if there is a change in the approval rating of the Prime Minister we test

$$H_0 : \pi_{1+} = \pi_{2+} \quad \text{vs} \quad H_a : \pi_{1+} \neq \pi_{2+}.$$ 

Note that the equation $\pi_{1+} = \pi_{2+}$ under $H_0$ can be rewritten as $\pi_{10} + \pi_{11} = \pi_{10} + \pi_{01}$ or equivalently as $\pi_{10} = \pi_{01}$ where $\pi_{ij} = P(Y_1 = i, Y_2 = j)$ and so the test can be rewritten as

$$H_0 : \pi_{10} = \pi_{01} \quad \text{vs} \quad H_a : \pi_{10} \neq \pi_{01}$$

or equivalently:

$$H_0 : P(Y_1 = 1, Y_2 = 0) = P(Y_1 = 0, Y_2 = 1)$$

$$H_a : P(Y_1 = 1, Y_2 = 0) \neq P(Y_1 = 0, Y_2 = 1).$$

The above statement can be translated into a statement in terms of $\gamma$ in our model as follows.

Note that
\[ P(Y_1 = 1, Y_2 = 0) = P(Y_1 = 0, Y_2 = 1) \]

is equivalent to
\[ P(Y_2 = 0|Y_1 = 1)P(Y_1 = 1) = P(Y_2 = 1|Y_1 = 0)P(Y_1 = 0) = \frac{1}{2} [1 - P(Y_1 = 1)] \]

which is equivalent to:
\[ [1 - P(Y_2 = 1|Y_1 = 1)]P(Y_1 = 1) = \frac{1}{2} [1 - P(Y_1 = 1)] \]

or
\[ P(Y_1 = 1) - P(Y_2 = 1|Y_1 = 1)P(Y_1 = 1) = \frac{1}{2} [1 - P(Y_1 = 1)] \]

which implies that
\[ e^r/(1 + e^r) = \{P(Y_1 = 1) - \frac{1}{2} [1 - P(Y_1 = 1)]\} [P(Y_1 = 1)]^{-1} \]

and so
\[ e^rP(Y_1 = 1) = (1 + e^r) \{P(Y_1 = 1) - \frac{1}{2} [1 - P(Y_1 = 1)]\} \]
\[ = P(Y_1 = 1) - \frac{1}{2} [1 - P(Y_1 = 1)] + e^r \{P(Y_1 = 1) - \frac{1}{2} [1 - P(Y_1 = 1)]\} \]

and
\[ 0 = \{P(Y_1 = 1) - \frac{1}{2} [1 - P(Y_1 = 1)]\} - \frac{1}{2} e^r(1 - P(Y_1 = 1)) \]

so
\[ \frac{1}{2} e^r(1 - P(Y_1 = 1)) = P(Y_1 = 1) - \frac{1}{2} [1 - P(Y_1 = 1)] \]

and equivalently
\[ e^r = \frac{P(Y_1 = 1) - \frac{1}{2} (1 - P(Y_1 = 1))}{\frac{1}{2} (1 - P(Y_1 = 1))} = 2 * \frac{P(Y_1 = 1)}{P(Y_1 = 0)} - 1 \]

and so
\[ \gamma = \ln(2 \cdot \frac{P(Y_{1=1})}{P(Y_{1=0})} - 1) \]

which depends on the probability \( P(Y_1 = 1) \). Thus if \( P(Y_1 = 1) \) is known we can test

\[ H_0 : \gamma = \ln(2 \cdot \frac{P(Y_{1=1})}{P(Y_{1=0})} - 1) \text{ v.s. } H_a : \gamma \neq \ln(2 \cdot \frac{P(Y_{1=1})}{P(Y_{1=0})} - 1) \]

which is equivalent to testing

\[ H_0 : \pi_{10} = \pi_{01} \text{ against } H_a : \pi_{10} \neq \pi_{01}. \]

Note that if we use the model where \( P(Y_n = 1) = e^{\gamma}/(1 + e^{\gamma}) \) we can rewrite the hypothesis test as:

\[ H_0 : \gamma = 0 \text{ against } H_a : \gamma \neq 0. \]

Suppose now that we were interested in testing the hypothesis

\[ H_0 : \frac{P(Y_{2=1}|Y_{1=1})}{P(Y_{2=1}|Y_{1=0})} = C \text{ against } H_a : \frac{P(Y_{2=1}|Y_{1=1})}{P(Y_{2=1}|Y_{1=0})} \neq C \]

which in the case of the rating of performance of the Prime Minister example can be considered as a test for the ratio between the probabilities of an individual approving of the Prime Minister on the second survey given that he approved on the first survey to the probability that an individual approves of the Prime Minister of the second survey given that he disapproved of the Prime Minister on the first survey. We can show that the above test of hypothesis is equivalent to:

\[ H_0 : \gamma = \ln(\frac{C}{2-C}) \text{ against } H_a : \gamma \neq \ln(\frac{C}{2-C}). \]
Equivalently if we were interested in testing the hypothesis of

\[ H_0 : \frac{P(Y_2 = 1 | Y_1 = 0)}{P(Y_2 = 0 | Y_1 = 1)} = C \text{ against } H_a : \frac{P(Y_2 = 1 | Y_1 = 0)}{P(Y_2 = 0 | Y_1 = 1)} \neq C \]

which in the case of the rating of performance of the Prime Minister example can be considered as a test for the ratio between the probabilities of an individual approving of the Prime Minister on the second survey given that he disapproved on the first survey to the probability that an individual disapproves of the Prime Minister of the second survey given that he approved of the Prime Minister on the first survey. It can also easily be shown that the hypothesis test is equivalent to:

\[ H_0 : \gamma = \ln(2C - 1) \text{ against } H_a : \gamma \neq \ln(2C - 1). \]

Thus we see that we can translate the above hypothesis tests which were initially about probabilities to tests in terms of the parameter \( \gamma \) for which we have a test statistic. In addition, as we shall see later, we can also conveniently obtain a corresponding power function as an added benefit.

Having justified the model in (4.1) through an application, we now return to the test on the association parameter \( \gamma \). Note that for this model

\[ p_{\beta 0} = P(Y_{it} = 1 | Y_{i,t-1} = 1) = e^\beta / (1 + e^\beta) \]

\[ p_{\beta 0} = P(Y_{it} = 1 | Y_{i,t-1} = 0) = 1/2 \]

implying that \( p_{\beta 0} \) is a fixed constant. Introducing a parameter \( \beta_0 \) representing a known fixed
constant into the model
\[ P(Y_{it} = 1|y_{it-1}) = \log it(\beta_0 + \gamma y_{it-1}) = \lambda_{it} \]

it is possible to control the values of the conditional probability \( P(Y_{it} = 1|y_{it-1} = 0) \) and let it take on values other than 1/2. In what follows, we will work with the model without \( \beta_0 \) for simplicity, but its introduction would not alter the results below, provided that \( \beta_0 \) is fixed and known.

Testing of statistical hypotheses with respect to statements about the parameter \( \gamma \) will play a central role in this paper. Thus we begin by investigating model (4.1) focusing on the parameter \( \gamma \). More specifically, we consider tests of hypothesis for this parameter making use of a methodology called the Rao Scoring test. The test statistic is obtained by taking the ratio of the score, which is the log of the likelihood of the joint distribution differentiated with respect to the parameter to be tested (in our case, \( \gamma \)), and the Fisher information corresponding to the parameter of interest.

So far as the development of the Rao scoring test is considered, we first derive the likelihood function as follows. Given that we take a random sample of \( n \) independent clusters i.e. \( \vec{Y}_{1}, \ldots, \vec{Y}_{n} \) independent random vectors for binary responses, we express the joint distribution of \( Y_{i1}, \ldots, Y_{iT} \) i.e. the \( T \) correlated binary responses within the \( i^{th} \) cluster as:

\[
P_i(y_{iT}, \ldots, y_{i1}) = P(y_{iT}|y_{i,T-1}, \ldots, y_{i1})P(y_{i,T-1}|y_{i,T-2}, \ldots, y_{i1}) \cdots P(y_{i2}|y_{i1})P(y_{i1})
= \lambda_{iT}^{y_{iT}} (1 - \lambda_{iT})^{1-y_{iT}} \lambda_{i,T-1}^{y_{i,T-1}} (1 - \lambda_{i,T-1})^{1-y_{i,T-1}} \lambda_{i,T-2}^{y_{i,T-2}} (1 - \lambda_{i,T-2})^{1-y_{i,T-2}} \cdots \lambda_{i2}^{y_{i2}} (1 - \lambda_{i2})^{1-y_{i2}} \lambda_{i1}^{y_{i1}} (1 - \lambda_{i1})^{1-y_{i1}}
= \prod_{t=1}^{T} \lambda_{it}^{y_{it}} (1 - \lambda_{it})^{1-y_{it}} \text{ where } y_{it} = 0 \text{ or } 1
\]

where \( \lambda_{it} = \mu_{it} \).
Since $\tilde{Y}_1, \ldots, \tilde{Y}_n$ is a random sample we only need to take the product of the joint distributions in order to get the likelihood of the data which is done as follows:

$$L = \prod_{i=1}^n P_i(y_i, \ldots, y_n) = \prod_{i=1}^n \prod_{j=1}^T \lambda_{ij}^{y_i}(1 - \lambda_{ij})^{1-y_i}.$$  

Note that we need not only consider a random sample for obtaining a likelihood function in this situation since a random sample requires that the random vectors be independent and identically distributed. A weaker requirement would be just the assumption of independence between the random vectors where their respective distributions need not be identical. This note will become clearer later on when we consider the distribution of a statistic that is obtained from the likelihood function of independent non-identically distributed random vectors $\tilde{Y}_1, \ldots, \tilde{Y}_n$.

Next we write the natural logarithm of the above function:

$$\ln(L) = \ln(\prod_{i=1}^n \prod_{j=1}^T \lambda_{ij}^{y_i}(1 - \lambda_{ij})^{1-y_i})$$

$$= \sum_{i=1}^n \sum_{j=1}^T \left[ y_i \ln(\lambda_{ij}) + (1 - y_i) \ln(1 - \lambda_{ij}) \right]$$

$$= \sum_{i=1}^n \left[ y_i \ln(\mu_i) + (1 - y_i) \ln(1 - \mu_i) \right] + \sum_{i=1}^n \sum_{j=2}^T \left[ y_i \ln(\lambda_{ij}) + (1 - y_i) \ln(1 - \lambda_{ij}) \right]$$

where the first summation is not a function of $\gamma$ but the second one is. Letting

$$\sum_{i=1}^n \left[ y_i \ln(\mu_i) + (1 - y_i) \ln(1 - \mu_i) \right] = C$$

we have:
\[ln(L) = C + \sum_{i=1}^{n} \sum_{j=2}^{T} [y_{ij} ln(\lambda_{ij}) + (1 - y_{ij}) ln(1 - \lambda_{ij})].\]

Also note that:

\[ln(\lambda_{ij}) = ln(-\frac{e^{y_{ij}-1}}{1 + e^{y_{ij}-1}}) - ln(1 + e^{y_{ij}-1}) = \gamma y_{ij-1} - ln(1 + e^{y_{ij}-1})\]

and that:

\[ln(1 - \lambda_{ij}) = ln(1 - \frac{e^{y_{ij}-1}}{1 + e^{y_{ij}-1}}) = ln(1 + e^{y_{ij}-1})\]

which gives us the following:

\[ln(L) = C + \sum_{i=1}^{n} \sum_{j=2}^{T} [y_{ij}(\gamma y_{ij-1} - ln(1 + e^{y_{ij}-1})) - (1 - y_{ij})ln(1 + e^{y_{ij}-1})]\]

\[= C + \sum_{i=1}^{n} \sum_{j=2}^{T} [\gamma y_{ij} y_{ij-1} - ln(1 + e^{y_{ij}-1})].\]

Next we differentiate with respect to \(\gamma\) to obtain the score as follows:

\[\frac{\partial ln(L)}{\partial \gamma} = \frac{\partial}{\partial \gamma}[C + \sum_{i=1}^{n} \sum_{j=2}^{T} [\gamma y_{ij} y_{ij-1} - ln(1 + e^{y_{ij}-1})]]\]

\[= \frac{\partial}{\partial \gamma} \sum_{i=1}^{n} \sum_{j=2}^{T} [\gamma y_{ij} y_{ij-1} - ln(1 + e^{y_{ij}-1})]\]

\[= \sum_{i=1}^{n} \sum_{j=2}^{T} [\gamma y_{ij} y_{ij-1} - \frac{\partial}{\partial \gamma}ln(1 + e^{y_{ij}-1})].\]

But

\[\frac{\partial}{\partial \gamma} ln(1 + e^{y_{ij}-1}) = \frac{\partial}{\partial \gamma} (1 + e^{y_{ij}-1})[1 + e^{y_{ij}-1}]^{-1} = y_{ij-1} e^{y_{ij}-1} / (1 + e^{y_{ij}-1}) = y_{ij-1} \lambda_{ij}\]

and so:

\[\frac{\partial ln(L)}{\partial \gamma} = \sum_{i=1}^{n} \sum_{j=2}^{T} [y_{ij} y_{ij-1} - y_{ij-1} \lambda_{ij}] = \sum_{i=1}^{n} \sum_{j=2}^{T} [y_{ij-1} (y_{ij} - \lambda_{ij})].\]
which is the score.

Now to find the Fisher information we need to get the negative expectation of the second
derivative of the natural logarithm of the likelihood with respect to \( \gamma \).

\[
I(\gamma) = -E\left( \frac{\partial^2 \ln(L)}{\partial \gamma^2} \right).
\]

First we obtain the second derivative as follows:

\[
\frac{\partial^2 \ln(L)}{\partial \gamma^2} = \frac{\partial}{\partial \gamma} \left( \frac{\partial \ln(L)}{\partial \gamma} \right) = \frac{\partial}{\partial \gamma} \left( a_n \right) = \frac{\partial}{\partial \gamma} \left\{ \sum_{j=1}^{n} \sum_{j=2}^{T} \left[ y_{ij} - \lambda y_{ij-1} \right] \right\} = \\
= \sum_{j=1}^{n} \sum_{j=2}^{T} \frac{\partial}{\partial \gamma} \left[ y_{ij} y_{ij-1} - y_{ij-1} \lambda y_{ij} \right]
\]

and

\[
\frac{\partial}{\partial \gamma} \left( \lambda y_{ij} \right) = \frac{\partial}{\partial \gamma} \left( \frac{e^{\gamma y_{ij-1}}}{1 + e^{\gamma y_{ij-1}}} \right) = \frac{\frac{\partial}{\partial \gamma} \left( e^{\gamma y_{ij-1}} \right) \left[ 1 + e^{\gamma y_{ij-1}} \right] - \frac{\partial}{\partial \gamma} \left( 1 + e^{\gamma y_{ij-1}} \right) \left[ e^{\gamma y_{ij-1}} \right]}{\left( 1 + e^{\gamma y_{ij-1}} \right)^2} \\
= y_{ij-1} \left[ \frac{e^{\gamma y_{ij-1}}}{1 + e^{\gamma y_{ij-1}}} \left( 1 - \frac{e^{\gamma y_{ij-1}}}{1 + e^{\gamma y_{ij-1}}} \right) \right] = y_{ij-1} \lambda y_{ij} \left( 1 - \lambda y_{ij} \right)
\]

which gives us the following:

\[
\frac{\partial^2 \ln(L)}{\partial \gamma^2} = -\sum_{j=1}^{n} \sum_{j=2}^{T} y_{ij-1}^2 \lambda y_{ij} \left( 1 - \lambda y_{ij} \right).
\]
Now all that remains is to take the expectation of the above which is easily done since the expectation of the sum is the sum of the expectations as shown below:

\[ I(\gamma) = E\left[ \sum_{i=1}^{n} \sum_{j=1}^{T} Y^2_{i,j-1} \lambda_{ij}(1 - \lambda_{ij}) \right] = \sum_{i=1}^{n} \sum_{j=1}^{T} E[Y^2_{i,j-1} \lambda_{ij}(1 - \lambda_{ij})]. \]

Since both \( Y^2_{i,j-1} \) and \( \lambda_{ij}(1 - \lambda_{ij}) \) are functions of the random binary response \( Y_{i,j-1} \) we can easily obtain the following expectation:

\[ E[Y^2_{i,j-1} \lambda_{ij}(1 - \lambda_{ij})] = \sum_{y_{i,j-1}=0}^{1} y^2_{i,j-1} \lambda_{ij}(1 - \lambda_{ij}) P(Y_{i,j-1} = y_{i,j-1}) \]

since \( P(Y_{i,j-1} = 1) = \mu_{i,j-1} \) and \( \lambda_{ij} = P(Y_{it} = 1|Y_{i,t-1}) \). Thus, we get:

\[
E[Y^2_{i,j-1} \lambda_{ij}(1 - \lambda_{ij})] = \sum_{y_{i,j-1}=0}^{1} y^2_{i,j-1} P(Y_{i,t} = 1|Y_{i,t-1} = y_{i,t-1})[1 - P(Y_{i,t} = 1|Y_{i,t-1} = y_{i,t-1})] P(Y_{i,j-1} = y_{i,j-1}) = P(Y_{it} = 1|Y_{i,t-1} = 1)[1 - P(Y_{it} = 1|Y_{i,t-1} = 1)] P(Y_{i,j-1} = 1) = p_{y1}(1 - p_{y1}) \mu_{i,j-1}
\]

where

\[ p_{y1} = P(Y_{ij} = 1|Y_{ij-1} = 1) = e^\gamma/(1 + e^\gamma). \]

As a result, the Fisher information is given by

\[ I(\gamma) = \sum_{i=1}^{n} \sum_{j=1}^{T} p_{y1}(1 - p_{y1}) \mu_{i,j-1}. \]

Note that since \( p_{y1} \) takes on the same value regardless of \( i \) and \( j \), i.e.,

\[ p_{y1} = e^\gamma/(1 + e^\gamma) = k \]
we can rewrite $I(\gamma)$ as

$$I(\gamma) = k(1-k) \sum_{j=1}^{n} \sum_{j'=2}^{T} H_{j,j'-1}$$

As for asymptotics for the Rao score statistic, we have

$$\frac{H_n^2}{I(\gamma)} \rightarrow^d \chi^2_1 \text{ or } \frac{H_n/ \sqrt{I(\gamma)}}{\rightarrow^d N(0,1)}$$

where $H_n = \sum_{j=1}^{n} A_j$ with $A_j = \sum_{j=2}^{T} [Y_{ij-1} (Y_{ij} - \mu_j)]$ obtained from $\frac{\partial \ln f(\gamma)}{\partial \gamma}$.

Note that in the case of independent non-identically distributed random vectors, this result can be shown via the Lindenberg Feller Central Limit Theorem. (Since $E(H_n) = 0$, and it is a mean taken from a sum of independent random variables with $E(A_i) = 0$ that are not necessarily identically distributed, and $I(\mu) = \theta^2_{(0)} = \sum_{j=1}^{n} \theta_j^2$ where $\theta_j^2$ can be shown to be the variance of $A_j = \sum_{j=2}^{T} [Y_{ij-1} (Y_{ij} - \mu_j)]$ and $H_n = \sum_{j=1}^{n} A_j$).

So setting up a hypothesis test:

$$H_0 : \gamma = \gamma_0 \text{ vs } H_a : \gamma = \gamma_1$$

we have the following rejection rule

Reject $H_0$ if $\left| \frac{H_n}{\sqrt{I(\gamma)}} \right| > c$
or equivalently

\[
\text{Reject } H_0 \text{ if } \left| \frac{\sum_{i=1}^n \sum_{j=1}^T \left[ X_{ij-1}(Y_{ij} - \lambda_{ij}) \right]}{\sqrt{k(1-k) \sum_{j=1}^n \sum_{j=2}^T \mu_{ij-1}}} \right| > c \quad \text{or if} \quad \left| \frac{\sum_{i=1}^n A_i}{\sqrt{k(1-k) \sum_{j=1}^n \sum_{j=2}^T \mu_{ij-1}}} \right| > c
\]

or if \( S > c \)

where \( A_i = \sum_{j=2}^T [Y_{ij-1}(Y_{ij} - \lambda_{ij})] \) and \( S = \left| \frac{\sum_{i=1}^n A_i}{\sqrt{k(1-k) \sum_{j=1}^n \sum_{j=2}^T \mu_{ij-1}}} \right| \).

The above can easily be used for setting up \( \alpha \) as the size of the test due to the asymptotic distribution of the statistic, but we will later on see that getting the power for this test is difficult due to the fact that \( H_n = \sum_{i=1}^n \sum_{j=2}^T [Y_{ij-1}(Y_{ij} - \lambda_{ij})] \) is a non-separable function of the parameter \( \gamma \).

Also note that the Fisher information is a function of the marginal probabilities \( \mu_{ij-1} \) and that we do not yet have an explicit form for \( \mu_{ij-1} \). Therefore, we consider next the derivation of a recursive relationship for the marginal probabilities \( \mu_{ij} \) based on the previous marginals \( \mu_{ij-1} \).
5. Recursive Relation of the Marginal Means

Since the Fisher information is a direct function of the marginal probability \( P(Y_{it} = 1) = \mu_{it} \), it would be convenient to have an explicit form to express any such given \( \mu_{it} \). The following is a recursive development of an expression for the marginal probabilities in the model where the conditional probability that \( Y_{it} = 1 \) is given by

\[
P(Y_{it} = 1|y_{it-1}) = \log it(y_{it-1}) = \lambda_{it}.
\]

We also introduce the following notation for the conditional probability that \( Y_{it} = 1 \) given that the previous binary response is one or zero respectively by letting it be either \( p_{it1} \) or \( p_{it0} \).

\[
p_{it1} = P(Y_{it} = 1|Y_{it-1} = 1) = \frac{e^\gamma}{1 + e^\gamma} = k
\]
\[
p_{it0} = P(Y_{it} = 1|Y_{it-1} = 0) = 1/2.
\]

Thus \( P(Y_{it} = 1|Y_{it-1} = 1) = f(\gamma) = k \) and \( P(Y_{it} = 1|Y_{it-1} = 0) = 1/2 \) for \( t = 2, \ldots, T \).

Also note that we must have a notation for the probability that the initial binary response is one since we do not have a \( Y_{i0} \) in the model upon which to condition. We will denote this probability as follows:

\[
P(Y_{i1} = 1) = \mu_{i1} = \lambda_{i1} = p_{i1}.
\]
Note that for the first time period, this expression for the marginal probability defined by $p_{it}$ does not require a third subscript, unlike $p_{ui}$ and $p_{i0}$ above, where conditioning is imposed based on the previous binary response being one or zero.

So now we are ready to develop the recursive relation. We start by using the definition of the marginal probability by summing out $Y_{it-1}$ for the bivariate distribution $P(Y_{it} = 1, Y_{it-1} = y_{it-1})$ as follows:

$$P(Y_{it} = 1) = \mu_{it} = \sum_{y_{it-1}=0}^{1} P(Y_{it} = 1, Y_{it-1} = y_{it-1}) =$$

$$P(Y_{it} = 1|Y_{it-1} = 1)P(Y_{it-1} = 1) + P(Y_{it} = 1|Y_{it-1} = 0)P(Y_{it-1} = 0) = k\mu_{it-1} + \frac{1}{2}(1 - \mu_{it-1}) = \mu_{it-1}(k - \frac{1}{2}) + \frac{1}{2}$$

and so we get the following recursion in terms of $t$:

$$\mu_{it} = \mu_{it-1}(k - \frac{1}{2}) + \frac{1}{2}.$$ 

So similarly

$$\mu_{it-1} = \mu_{it-2}(k - \frac{1}{2}) + \frac{1}{2},$$

which leads to

$$\mu_{it} = [\mu_{it-2}(k - \frac{1}{2}) + \frac{1}{2}](k - \frac{1}{2}) + \frac{1}{2} = \mu_{it-2}(k - \frac{1}{2})^2 + \frac{1}{2}[(k - \frac{1}{2}) + 1]$$

$$= \mu_{it-2}(k - \frac{1}{2})^2 + \frac{1}{2}[(k - \frac{1}{2})^1 + (k - \frac{1}{2})^0].$$

However, since

$$\mu_{it-2} = \mu_{it-3}(k - \frac{1}{2}) + \frac{1}{2}$$

we get

$$\mu_{it} = [\mu_{it-3}(k - \frac{1}{2}) + \frac{1}{2}](k - \frac{1}{2})^2 + \frac{1}{2}[(k - \frac{1}{2})^1 + (k - \frac{1}{2})^0]$$
Now at this point the pattern is starting to look repetitive and if we continue in the same fashion we will eventually get the following:

\[
\begin{align*}
\mu_{it} &= \mu_{it-1}(k - \frac{1}{2})^{t-1} + \frac{1}{2} \sum_{i=0}^{t-2} (k - \frac{1}{2})^i = \mu_{i1}(k - \frac{1}{2})^{t-1} + \frac{1}{2} \sum_{i=0}^{t-2} (k - \frac{1}{2})^i \\
\mu_{i1}(k - \frac{1}{2})^{t-1} + \frac{1}{2} \sum_{i=0}^{t-2} (k - \frac{1}{2})^i &= p_{it}(k - \frac{1}{2})^{t-1} + \frac{1}{2} \sum_{i=0}^{t-2} (k - \frac{1}{2})^i.
\end{align*}
\]

Thus we can express \(\mu_{it}\) as a relatively simple function of \(p_{it}\) for all \(t\). Note that if we choose to model \(p_{it}\) as

\[
p_{it} = P(Y_{it} = 1) = e^r/(1 + e^r) = k
\]

we obtain

\[
\mu_{it} = k(k - \frac{1}{2})^{t-1} + \frac{1}{2} \sum_{i=0}^{t-2} (k - \frac{1}{2})^i.
\]
6. Power Functions: 
The Asymptotic Distribution of the Test Statistic

6.1. Power Function for Two Time Periods

We now consider a test of hypothesis that would center on assessing the extent and significance of the association between two consecutive binary responses within the same cluster. Specifically, we focus on the test:

\[ H_0 : \gamma = \gamma_0 \quad \text{vs} \quad H_a : \gamma = \gamma_1. \]

We are particularly interested in the power of this test. To begin with, we derive an expression for the asymptotic power function for the situation when there are only two binary observations in each cluster; that is \( T = 2 \). Recall that for general \( T \), our rejection rule from Chapter 4 is

\[
\text{Reject } H_0 \text{ if } \left| \frac{\sum_{i=1}^{n} A_i}{\sqrt{T(1-T) \sum_{j=1}^{n} \sum_{t=2}^{T} \mu_{it-1}}} \right| > c
\]

where \( A_i = \sum_{t=2}^{T} [Y_{i,t-1}(Y_{i,t} - \lambda_{it})] \). For \( T = 2 \), this rejection rule simplifies to

\[
\text{Reject } H_0 \text{ if } \left| \frac{\sum_{i=1}^{n} A_i}{\sqrt{4 \sum_{j=1}^{n} \mu_{ii}}} \right| > c
\]

or equivalently
Reject \( H_0 \) if
\[
\left| \frac{\sum_{i=1}^{n} A_i}{\sqrt{k(1-k)\sum_{i=1}^{n} p_{ni}}} \right| > c
\]

where \( A_i = Y_i(Y_i - \lambda_i) \).

Therefore, if \( \alpha \) is the size of the test, that is
\[
P( \text{Reject } H_0 | H_0 \text{ is true}) = P(\text{Reject } H_0 | \gamma = \gamma_0) = \alpha
\]

then
\[
P( \text{Reject } H_0 | H_0 \text{ is true}) = P\left( \left| \frac{\sum_{i=1}^{n} A_i}{\sqrt{k_0(1-k_0)\sum_{i=1}^{n} p_{ni}}} \right| > c | \gamma = \gamma_0 \text{ or } k = k_0 = e^{\gamma_0}/(1 + e^{\gamma_0}) \right)
\]

But since \( \frac{H_n}{\sqrt{\lambda(\gamma)}} \to^d N(0, 1) \) which in our case is
\[
\frac{\sum_{i=1}^{n} A_i}{\sqrt{k_0(1-k_0)\sum_{i=1}^{n} p_{ni}}} \to^d N(0, 1) \text{ given that } \gamma = \gamma_0
\]

(equivalent to \( k = k_0 \)), we get \( c = z_{\alpha/2} \), which gives us a rejection rule of

Reject \( H_0 \) if
\[
\left| \frac{\sum_{i=1}^{n} A_i}{\sqrt{k_0(1-k_0)\sum_{i=1}^{n} p_{ni}}} \right| > z_{\alpha/2}
\]

or equivalently

Reject \( H_0 \) if
\[
\left| \sum_{i=1}^{n} A_i \right| > z_{\alpha/2} \sqrt{k_0(1-k_0)\sum_{i=1}^{n} p_{ni}}
\]
It is important to note that $\sum_{i=1}^{n} A_i$ is a function of the parameter $\gamma$. Thus the rejection rule would be:

$$\text{Reject } H_0 \text{ if } | \sum Y_i (Y_{i2} - \lambda_{i2\gamma_0}) | > z_{a/2} \sqrt{k_0(1 - k_0) \sum_{i=1}^{n} p_{i1}}$$

where $\lambda_{i2\gamma_0} = e^{\gamma_0 Y_i}(1 + e^{\gamma_0} \gamma)$. 

Since the general expression for the power is

$$\text{Power} = P(\text{Reject } H_0 | H_a \text{ is true})$$

in our situation we have

$$\text{Power} = P( | \sum_{i=1}^{n} Y_i (Y_{i2} - \lambda_{i2\gamma_0}) | > z_{a/2} \sqrt{k_0(1 - k_0) \sum_{i=1}^{n} p_{i1}} | \gamma = \gamma_1 \text{ or } k = k_1 = \frac{e^{\gamma_1}}{1 + e^{\gamma_1}})$$

$$= P( | \sum_{i=1}^{n} (Y_i Y_{i2} - Y_i \lambda_{i2\gamma_0}) | > z_{a/2} \sqrt{k_0(1 - k_0) \sum_{i=1}^{n} p_{i1} | k = k_1}).$$

To get a power function from the above probability statement we must first try to look at the asymptotics for

$$\sum_{i=1}^{n} (Y_i Y_{i2} - Y_i \lambda_{i2\gamma_0}).$$

We shall start by considering the expectation of the above summation:

$$E_{\gamma_1}[\sum_{i=1}^{n} (Y_i Y_{i2} - Y_i \lambda_{i2\gamma_0})] = \sum_{i=1}^{n} E[(Y_i Y_{i2} - Y_i \lambda_{i2\gamma_0})]$$
Next we determine the variance $V(\sum_{i=1}^{n}(Y_{i1}Y_{i2} - Y_{i1}\lambda_j\gamma_{i2}))$ as follows

$$V(\sum_{i=1}^{n}(Y_{i1}Y_{i2} - Y_{i1}\lambda_j\gamma_{i2})) = \sum_{i=1}^{n} V[(Y_{i1}Y_{i2} - Y_{i1}\lambda_j\gamma_{i2})]$$

$$= \sum_{i=1}^{n} [E[(Y_{i1}Y_{i2})^2] - 2E[Y_{i1}Y_{i2}\lambda_j\gamma_{i2}] + E[(Y_{i1}\lambda_j\gamma_{i2})^2] - E[(Y_{i1}Y_{i2} - Y_{i1}\lambda_j\gamma_{i2})^2]]$$

$$= \sum_{i=1}^{n} [P(Y_{i1} = 1, Y_{i2} = 1)\gamma_j - 2P(Y_{i1} = 1, Y_{i2} = 1)\gamma_j [e^{r_0}/(1 + e^{r_0})]]$$

$$= \sum_{i=1}^{n} [p_1(k_1 - k_0)^2 - \sum_{i=1}^{n} p_1k_1 - 2k_1k_0 + k_0^2 - p_1(k_1 - k_0)^2]$$

$$= \sum_{i=1}^{n} p_1[k_1 - 2k_1k_0 + k_0^2 - p_1(k_1 - k_0)^2].$$

Note that if it is assumed that $p_1 = P(Y_{i1} = 1) = e^{r_0}/(1 + e^{r_0}) = k$ for all $i = 1, ..., n$ clusters, under the alternative hypothesis the mean and variance above simplify to

$$E(\sum_{i=1}^{n}(Y_{i1}Y_{i2} - Y_{i1}\lambda_j\gamma_{i2})) = (k_1 - k_0)nk_1$$

and

$$V(\sum_{i=1}^{n}(Y_{i1}Y_{i2} - Y_{i1}\lambda_j\gamma_{i2}))$$

$$= \sum_{i=1}^{n} p_1[k_1 - 2k_1k_0 + k_0^2] - \sum_{i=1}^{n} p_1^2(k_1 - k_0)^2]$$
With both the mean and the variance of this sum of independent random variables we are in a position to apply the Lindeberg-Feller Central Limit Theorem in order to get an asymptotic distribution that will be of use for calculating the power function above. Specifically, we can state that

\[
\frac{\sum_{i=1}^{n}(Y_{ni} - \bar{Y} - \lambda \bar{\sigma}_0) - n(1 - k_0)\sum_{i=1}^{n}p_{ni}}{\sqrt{n}\sum_{i=1}^{n}p_{ni}(1 - 2k_0 + k_0^2)p_{ni}(1 - k_0)^2}} \rightarrow N(0, 1)
\]

which implies that

\[
\frac{\sum_{i=1}^{n}(Y_{ni} - \bar{Y} - \lambda \bar{\sigma}_0) - (k_1 - k_0)\sum_{i=1}^{n}p_{ni}}{\sqrt{\sum_{i=1}^{n}p_{ni}(1 - 2k_0 + k_0^2)p_{ni}(1 - k_0)^2}}} \rightarrow N(0, 1).
\]

Thus, returning to the above power function we have the following:

\[
P \left( \left| \sum_{i=1}^{n}(Y_{ni} - \bar{Y} - \lambda \bar{\sigma}_0) \right| > z_{a/2} \sqrt{k_0(1 - k_0)\sum_{i=1}^{n}p_{ni}} \mid k = k_1 \right)
\]

\[
= P(\sum_{i=1}^{n}(Y_{ni} - \bar{Y} - \lambda \bar{\sigma}_0) > z_{a/2} \sqrt{k_0(1 - k_0)\sum_{i=1}^{n}p_{ni}} \mid k = k_1)
\]

\[
+ P(\sum_{i=1}^{n}(Y_{ni} - \bar{Y} - \lambda \bar{\sigma}_0) < -z_{a/2} \sqrt{k_0(1 - k_0)\sum_{i=1}^{n}p_{ni}} \mid k = k_1)
\]
Finally, note that under the assumption that \( p_{\gamma} = p(Y_1 = 1) = \frac{e^{\gamma}}{1 + e^{\gamma}} = k \), we obtain the power function

\[
1 - \Phi\left[ -z_{\alpha/2} \frac{\sqrt{\sum_{i=1}^{n} p_{i1} - (k_1-k_0) k_1}}{\sqrt{\sum_{i=1}^{n} p_{i1} (k_1-2k_1k_0 + k_0^2) - p_{i1} (k_1-k_0)^2}} \right] + \Phi\left[ z_{\alpha/2} \frac{\sqrt{\sum_{i=1}^{n} p_{i1} - (k_1-k_0) k_1}}{\sqrt{\sum_{i=1}^{n} p_{i1} (k_1-2k_1k_0 + k_0^2) - p_{i1} (k_1-k_0)^2}} \right]
\]

\[
1 - \Phi\left[ -z_{\alpha/2} \frac{\sqrt{\sum_{i=1}^{n} p_{i1} - (k_1-k_0) k_1}}{\sqrt{\sum_{i=1}^{n} p_{i1} (k_1-2k_1k_0 + k_0^2) - p_{i1} (k_1-k_0)^2}} \right] - \frac{\sqrt{\sum_{i=1}^{n} p_{i1} (k_1-2k_1k_0 + k_0^2) - p_{i1} (k_1-k_0)^2}}{\sqrt{\sum_{i=1}^{n} p_{i1} (k_1-2k_1k_0 + k_0^2) - p_{i1} (k_1-k_0)^2}}
\]

As a numerical illustration of this formula, Table 6.1 below gives the power for the test

\( H_0 : \gamma = 0 \) against \( H_a : \gamma \neq 0 \)

for different values of \( \gamma \) under \( H_a \) for sample sizes of \( n = 25 \) and \( n = 50 \) and \( T = 2 \). However, note that \( P(Y_1 = 1) = P(Y_2 = 1|Y_1 = 1) = \frac{e^{\gamma}}{1 + e^{\gamma}} = k \), which suggests that a value of \( k \) can be computed for any value of \( \gamma \) specified under \( H_a \). In addition, it can also be shown that for the model under
consideration that the correlation between $Y_1$ and $Y_2$ can be written in terms of $k$ only, suggesting that, in this case, there is a one-to-one relationship between $\gamma$ under $H_a$, and the correlation between $Y_1$ and $Y_2$. Specifically, we can derive an expression for $Corr(Y_1, Y_2)$ as a function of $k$ as follows.

Table 6.1: Rao score power for $T=2$

<table>
<thead>
<tr>
<th>Value of $\gamma$ under $H_a$</th>
<th>$k$ under $H_a$</th>
<th>Correlation under $H_a$</th>
<th>Power for $n = 25$</th>
<th>Power for $n = 50$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.75</td>
<td>0.148047</td>
<td>-0.251358</td>
<td>0.264859</td>
<td>0.481442</td>
</tr>
<tr>
<td>-1.50</td>
<td>0.182426</td>
<td>-0.246955</td>
<td>0.265577</td>
<td>0.482725</td>
</tr>
<tr>
<td>-1.25</td>
<td>0.222700</td>
<td>-0.232526</td>
<td>0.250218</td>
<td>0.454937</td>
</tr>
<tr>
<td>-1.00</td>
<td>0.268941</td>
<td>-0.206508</td>
<td>0.216914</td>
<td>0.392367</td>
</tr>
<tr>
<td>-0.75</td>
<td>0.320821</td>
<td>-0.168395</td>
<td>0.168411</td>
<td>0.296281</td>
</tr>
<tr>
<td>-0.50</td>
<td>0.377541</td>
<td>-0.119241</td>
<td>0.113992</td>
<td>0.183524</td>
</tr>
<tr>
<td>-0.25</td>
<td>0.437823</td>
<td>-0.0617855</td>
<td>0.068777</td>
<td>0.0887318</td>
</tr>
<tr>
<td>0.25</td>
<td>0.562177</td>
<td>0.0618452</td>
<td>0.0741712</td>
<td>0.0999682</td>
</tr>
<tr>
<td>0.50</td>
<td>0.622459</td>
<td>0.120134</td>
<td>0.156976</td>
<td>0.272921</td>
</tr>
<tr>
<td>0.75</td>
<td>0.679179</td>
<td>0.172465</td>
<td>0.306620</td>
<td>0.553426</td>
</tr>
<tr>
<td>1.00</td>
<td>0.731059</td>
<td>0.217707</td>
<td>0.506782</td>
<td>0.818003</td>
</tr>
<tr>
<td>1.25</td>
<td>0.777300</td>
<td>0.255728</td>
<td>0.710817</td>
<td>0.956982</td>
</tr>
<tr>
<td>1.50</td>
<td>0.817574</td>
<td>0.287024</td>
<td>0.867238</td>
<td>0.994861</td>
</tr>
<tr>
<td>1.75</td>
<td>0.851953</td>
<td>0.312398</td>
<td>0.954971</td>
<td>0.999735</td>
</tr>
</tbody>
</table>

$$Corr(Y_1, Y_2) = \frac{E[\gamma_1 \gamma_2] - E[\gamma_1]E[\gamma_2]}{\sqrt{V[\gamma_1]} \sqrt{V[\gamma_2]}} = \frac{E[\gamma_1 \gamma_2] - E[\gamma_1]E[\gamma_2]}{\sqrt{V[\gamma_1]} \sqrt{V[\gamma_2]}}$$
since

\[ P(Y_1 = 1) = P(Y_2 = 1|Y_1 = 1) = e^{\gamma}(1 + e^{\gamma}) = k \]

and

\[ P(Y_2 = 1) = \mu_2 = k^2 - \frac{1}{2}k + \frac{1}{2} = k^2 + \frac{1}{2}(1 - k) \quad \text{(Chapter 5)} \]

Thus Table 6.1 also shows the value of \( Corr(Y_1, Y_2) \) corresponding to \( \gamma \) indicated under \( H_a \). In addition, note that using the above equation yields a correlation value of zero for the case where \( \gamma = 0 \). Thus the test

\[ H_0 : \gamma = 0 \quad \text{against} \quad H_a : \gamma \neq 0 \]

can be considered a test for a lack of correlation. From the above table we can observe that the Rao Score test attains high power at the points where the alternative hypothesis is one of positive correlation for a sample size of \( n = 50 \) or more. However, when the alternative hypothesis is a point of negative correlation the above test does not perform well. This however is not really a big problem considering the scarcity of examples where correlated binary data is negatively correlated.
6.2 Power Function for Two Time Periods with a Cross Product Statistic

To this point we have been using Rao scoring test statistics in order to make a decision regarding the test of hypothesis:

\[ H_0 : \gamma = \gamma_0 \text{ against } H_a : \gamma = \gamma_1. \]

Another statistic that might be a possible candidate for testing the hypothesis regarding the parameter \( \gamma \) for the case where \( T = 2 \) is the sum \( \sum_{i=1}^{n} Y_i Y_{i+1} \).

We can obtain the mean and variance of this statistic in order to standardize it and eventually get its asymptotic distribution with the use of the Central Limit Theorem. Specifically, we consider the model where the conditional probability distribution for \( Y_{i+1} \) is only a function of \( Y_i \) according to

\[ P(Y_{i+1} = 1 | Y_i) = e^{\gamma Y_i} / (1 + e^{\gamma Y_i}) \]

and

\[ P(Y_i = 1) = e^\gamma / (1 + e^\gamma) = k. \]

It is straightforward to derive the expectation of the sum \( \sum_{i=1}^{n} Y_i Y_{i+1} \) as follows

\[
E(\sum_{i=1}^{n} Y_i Y_{i+1}) = \sum_{i=1}^{n} E(Y_i Y_{i+1}) = \sum_{i=1}^{n} P(Y_i = 1, Y_{i+1} = 1)
\]

\[= \sum_{i=1}^{n} P(Y_{i+1} = 1 | Y_i = 1) P(Y_i = 1) = \sum_{i=1}^{n} k^2 = nk^2. \]

When calculating the variance of the sum, note that \( \tilde{Y}_1, \ldots, \tilde{Y}_n \) are independent random vectors and so:
\[ \text{Var} [\sum_{i=1}^n Y_i Y_{i2}] = \sum_{i=1}^n \text{Var}[Y_i Y_{i2}] = \sum_{i=1}^n [\mathbb{E}[(Y_i Y_{i2})^2] - \mathbb{E}[Y_i Y_{i2}]]^2 \]
\[ = \sum_{i=1}^n [k^2 - (k^2)^2] = nk^2(1 - k^2). \]

Next, according to the Central Limit Theorem:
\[ \frac{\sum_{i=1}^n Y_i Y_{i2} - \sum_{i=1}^n Y_i Y_{i2}^2}{\sqrt{\text{Var} [\sum_{i=1}^n Y_i Y_{i2}]}} = \frac{\sum_{i=1}^n Y_i Y_{i2} - nk^2}{\sqrt{nk^2(1 - k^2)}} \to N(0, 1). \]

Therefore, if we wish to base the test of hypothesis for \( H_0 : \gamma = \gamma_0 \) against \( H_a : \gamma = \gamma_1 \) on the sum \( \sum_{i=1}^n Y_i Y_{i2} \) we could consider the rejection rule

\[ \text{Reject } H_0 \text{ if } \left| \frac{\sum_{i=1}^n Y_i Y_{i2} - nk^2}{\sqrt{nk^2(1 - k^2)}} \right| > c. \]

where \( c \) is chosen to control the significance level of the test. In particular, if the size of the test is set to \( \alpha \), we could write

\[ P(\text{Reject } H_0 | H_0 \text{ is true}) = P\left( \left| \frac{\sum_{i=1}^n Y_i Y_{i2} - nk^2}{\sqrt{nk^2(1 - k^2)}} \right| > c | \gamma = \gamma_0, k = k_0 \right) = \alpha \]

where \( k_0 = e^{\gamma_0}/(1 + e^{\gamma_0}) \).

For this probability statement to hold, it must be the case that \( c = z_\alpha \), so the rejection region is:
Reject $H_0$ if 

$$\left| \frac{\sum_{i=1}^{n} y_i \gamma - nk_0^2}{\sqrt{\gamma} \sqrt{(1-k_0^2)}} \right| > z_{a/2}.$$ 

Note that we can also set up an expression for the power function for this test when $H_a$ is true and $\gamma = \gamma_1$ or equivalently $k = k_1 = e^{\gamma_1}/(1 + e^{\gamma_1})$. Specifically,

$$P(\text{Reject } H_0|H_a \text{ is true}) =$$

$$P\left( \left| \frac{\sum_{i=1}^{n} y_i \gamma - nk_0^2}{\sqrt{\gamma} \sqrt{(1-k_0^2)}} \right| > z_{a/2} \middle| \gamma = \gamma_1, k = k_1 \right)$$

$$= P\left(\frac{\sum_{i=1}^{n} y_i \gamma - nk_0^2}{\sqrt{\gamma} \sqrt{(1-k_0^2)}} > z_{a/2} \middle| \gamma = \gamma_1, k = k_1 \right)$$

$$+ P\left(\frac{\sum_{i=1}^{n} y_i \gamma - nk_0^2}{\sqrt{\gamma} \sqrt{(1-k_0^2)}} < -z_{a/2} \middle| \gamma = \gamma_1, k = k_1 \right)$$

$$= P\left(\sum_{i=1}^{n} y_i \gamma - nk_0^2 > z_{a/2} \frac{\sqrt{\gamma} \sqrt{(1-k_0^2)}}{\sqrt{\gamma} \sqrt{(1-k_0^2)}} + \sqrt{n} \frac{(k_0^2 - k_1^2)}{\sqrt{\gamma} \sqrt{(1-k_0^2)}} \middle| \gamma = \gamma_1, k = k_1 \right)$$

$$+ P\left(\sum_{i=1}^{n} y_i \gamma - nk_0^2 < -z_{a/2} \frac{\sqrt{\gamma} \sqrt{(1-k_0^2)}}{\sqrt{\gamma} \sqrt{(1-k_0^2)}} + \sqrt{n} \frac{(k_0^2 - k_1^2)}{\sqrt{\gamma} \sqrt{(1-k_0^2)}} \middle| \gamma = \gamma_1, k = k_1 \right)$$

$$= P\left(Z > z_{a/2} \frac{\sqrt{\gamma}}{\sqrt{(1-k_0^2)}} + \sqrt{n} \frac{(k_0^2 - k_1^2)}{\sqrt{\gamma} \sqrt{(1-k_0^2)}} \middle| \gamma = \gamma_1, k = k_1 \right)$$

$$+ P\left(Z < -z_{a/2} \frac{\sqrt{\gamma}}{\sqrt{(1-k_0^2)}} + \sqrt{n} \frac{(k_0^2 - k_1^2)}{\sqrt{\gamma} \sqrt{(1-k_0^2)}} \middle| \gamma = \gamma_1, k = k_1 \right)$$

$$= 1 - \Phi(z_{a/2} \frac{\sqrt{\gamma}}{\sqrt{(1-k_0^2)}} + \sqrt{n} \frac{(k_0^2 - k_1^2)}{\sqrt{\gamma} \sqrt{(1-k_0^2)}}) + \Phi(-z_{a/2} \frac{\sqrt{\gamma}}{\sqrt{(1-k_0^2)}} + \sqrt{n} \frac{(k_0^2 - k_1^2)}{\sqrt{\gamma} \sqrt{(1-k_0^2)}}).$$
Note that, as would intuitively be expected, the above power increases as the sample size $n$ increases, and also as the difference $k_2 - k_1$ gets larger. As a numerical illustration of this formula, Table 6.2 below gives the power for the test

$$H_0: \gamma = 0 \quad \text{against} \quad H_a: \gamma \neq 0$$

for different values of $\gamma$ and $\text{Corr}(Y_1, Y_2)$ under $H_a$ for sample sizes of $n = 25$ and $n = 50$.

<table>
<thead>
<tr>
<th>Value of $\gamma$ under $H_a$</th>
<th>$k$ under $H_a$</th>
<th>Correlation under $H_a$</th>
<th>Power for $n = 25$</th>
<th>Power for $n = 50$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.75</td>
<td>0.148047</td>
<td>-0.251358</td>
<td>0.976831</td>
<td>1.00000</td>
</tr>
<tr>
<td>-1.50</td>
<td>0.182426</td>
<td>-0.246955</td>
<td>0.904838</td>
<td>0.999931</td>
</tr>
<tr>
<td>-1.25</td>
<td>0.222700</td>
<td>-0.232526</td>
<td>0.759964</td>
<td>0.995577</td>
</tr>
<tr>
<td>-1.00</td>
<td>0.268941</td>
<td>-0.206508</td>
<td>0.560825</td>
<td>0.942211</td>
</tr>
<tr>
<td>-0.75</td>
<td>0.320821</td>
<td>-0.168395</td>
<td>0.354580</td>
<td>0.735469</td>
</tr>
<tr>
<td>-0.50</td>
<td>0.377541</td>
<td>-0.119241</td>
<td>0.186581</td>
<td>0.399719</td>
</tr>
<tr>
<td>-0.25</td>
<td>0.437823</td>
<td>-0.0617855</td>
<td>0.0803577</td>
<td>0.134475</td>
</tr>
<tr>
<td>0.25</td>
<td>0.562177</td>
<td>0.0618452</td>
<td>0.137992</td>
<td>0.208147</td>
</tr>
<tr>
<td>0.50</td>
<td>0.622459</td>
<td>0.120134</td>
<td>0.370996</td>
<td>0.599958</td>
</tr>
<tr>
<td>0.75</td>
<td>0.679179</td>
<td>0.172465</td>
<td>0.661609</td>
<td>0.902251</td>
</tr>
<tr>
<td>1.00</td>
<td>0.731059</td>
<td>0.217707</td>
<td>0.874887</td>
<td>0.990118</td>
</tr>
<tr>
<td>1.25</td>
<td>0.777300</td>
<td>0.255728</td>
<td>0.970349</td>
<td>0.999645</td>
</tr>
<tr>
<td>1.50</td>
<td>0.817574</td>
<td>0.287024</td>
<td>0.995870</td>
<td>1.00000</td>
</tr>
<tr>
<td>1.75</td>
<td>0.851953</td>
<td>0.312398</td>
<td>0.999699</td>
<td>1.00000</td>
</tr>
</tbody>
</table>

Note that the values for the power are larger when $n = 50$ than when $n = 25$ as we would expect. Also note that the values for the power are larger, when the absolute distance from $\gamma = 0$ is larger which is also what we would expect in this case. Finally, note that the above statistic performs much better than the Rao score statistic since it attains higher power for smaller samples sizes and
attains higher power at the points where the correlation is negative where as the Rao Score test preforms well only when the alternative hypothesis is one of positive correlation.
6.3 Power Function for Three Time Periods

For the same test of hypothesis

\[ H_0 : \gamma = \gamma_0 \quad \text{vs} \quad H_a : \gamma = \gamma_1 \]

we now consider the derivation of an approximate power function for the case of binary responses collected over many independent clusters for \( T = 3 \) consecutive time periods. Returning to the results derived in Chapter 4, the rejection rule is now:

\[
\text{Reject } H_0 \quad \text{if} \quad \frac{\sum_{i=1}^{n} A_i}{\sqrt{\frac{1}{2}(1-\lambda) \sum_{i=1}^{n} \sum_{t=2}^{3} \lambda_{it-1}}} > c
\]

where \( A_i = \sum_{t=2}^{3} [Y_{i,t-1} (Y_{it} - \lambda_{it})] \).

In addition since

\[ \mu_{ii} = \mu_{i1} + \frac{1}{2} \sum_{t=0}^{r-2} (k-\frac{1}{2})^t \]

we get

\[ \mu_{i2} = \mu_{i1} + \frac{1}{2} \sum_{t=0}^{r-2} (k-\frac{1}{2})^t = \frac{1}{2} (k-\frac{1}{2})^0 = \mu_{i1} + \frac{1}{2} \]

and \( \mu_{ii} = \mu_{i1} \)

Therefore, \( \sum_{t=2}^{3} \mu_{it-1} = \mu_{i1} + \mu_{i2} = p_{i1} + p_{i1}(k-\frac{1}{2}) + \frac{1}{2} = p_{i1}(k+\frac{1}{2}) + \frac{1}{2} \)
and so eventually we have

\[ \sum_{t=1}^{n} \sum_{t=2}^{3} \mu_{i, t-1} = \sum_{t=1}^{n} [p_{i1}(k + \frac{1}{2}) + \frac{1}{2}] = (k + \frac{1}{2}) \sum_{t=1}^{n} p_{i1} + \frac{n}{2}. \]

Using this result, our rejection rule for \( T = 3 \) would be

\[
\text{Reject } H_0 \text{ if } \left| \frac{\sum_{t=1}^{n} A_t}{\sum_{t=1}^{n} \sum_{t=2}^{3} p_{i1} + \frac{n}{2}} \right| > c
\]

where

\[ A_t = Y_t(Y_{t2} - \lambda_{t2}) + Y_{t2}(Y_{t3} - \lambda_{t3}). \]

Now, since the significance level, \( \alpha \) for a test is defined generally as

\[ P(\text{Reject } H_0 | H_0 \text{ is true}) = \alpha \]

then, incorporating the above rejection rule into this definition yields

\[
P(\text{Reject } H_0 | H_0 \text{ is true}) = P\left( \left| \frac{\sum_{t=1}^{n} A_t}{\sum_{t=1}^{n} \sum_{t=2}^{3} p_{i1} + \frac{n}{2}} \right| > c | \gamma = \gamma_0 \text{ or } k = k_0 \right) = P\left( \left| \frac{\sum_{t=1}^{n} A_t}{\sum_{t=1}^{n} \sum_{t=2}^{3} p_{i1} + \frac{n}{2}} \right| > c | \gamma = \gamma_0 \text{ or equivalently } k = k_0 \right).
\]

Based on the expression above, note that \( A_t \) is a function of \( k_0 \) under this probability statement. In
addition, since \( -\frac{H_n}{f(0)} \to^d N(0, 1) \) which in our case is

\[
\frac{\sum_{i=1}^n A_i}{\sqrt{k_0(1-k_0)\sum_{i=1}^n p_{i1} + \frac{3}{2}}} \to^d N(0, 1)
\]

given that \( \gamma = \gamma_0 \) (or \( k = k_0 \)) then \( c = z_{a/2} \) and the rejection rule becomes

\[
\text{Reject } H_0 \text{ if } \left| \frac{\sum_{i=1}^n A_i}{\sqrt{k_0(1-k_0)\sum_{i=1}^n p_{i1} + \frac{3}{2}}} \right| > z_{a/2}
\]

or equivalently

\[
\text{Reject } H_0 \text{ if } \left| \sum_{i=1}^n A_i \right| > z_{a/2} \sqrt{k_0(1-k_0)[(k_0 + 2)\sum_{i=1}^n p_{i1} + \frac{3}{2}]}
\]

It is extremely important to note that \( \left| \sum_{i=1}^n A_i \right| \) is a function of \( \gamma_0 \). More specifically, note that the rejection rule can be written as

\[
\text{Reject } H_0 \text{ if } \left| \sum_{i=1}^n Y_{i2} - \lambda_{i2\gamma_0} + Y_{i3} - \lambda_{i3\gamma_0} \right| > z_{a/2} \sqrt{k_0(1-k_0)[(k_0 + 2)\sum_{i=1}^n p_{i1} + \frac{3}{2}]}
\]

or alternatively as

\[
\left| \sum_{i=1}^n Y_{i1}(Y_{i2} - \frac{\gamma_0}{1 + \gamma_0 \rho_{i2}}) + Y_{i2}(Y_{i3} - \frac{\gamma_0}{1 + \gamma_0 \rho_{i3}}) \right| > z_{a/2} \sqrt{k_0(1-k_0)[(k_0 + 2)\sum_{i=1}^n p_{i1} + \frac{3}{2}]}
\]

Since the general definition for power is

\[
\text{Power} = P(\text{Reject } H_0|H_a \text{ is true})
\]

for the particular situation at hand with have
In order to apply this result, it is necessary to derive the mean, variance and asymptotic
distribution of a conveniently standardized random variable, in the same manner as the case with
\( T = 2 \). As before with \( T = 2 \) , it is assumed that all \( p_{ni} \) are equal ( \( p_{ni} = k \) for all \( i = 1, \ldots, n \)) for
maintaining a relative simplicity of calculation. The same steps can however be done without that
assumption. For the case of \( T = 3 \), the mean of interest, then is in fact

\[
E_{\gamma_1}[\sum_{i=1}^{n} Y_{i1}(Y_{i2} - \frac{e^{\gamma_2}Y_{i1}}{1+e^{\gamma_2}Y_{i1}}) + Y_{i2}(Y_{i3} - \frac{e^{\gamma_2}Y_{i2}}{1+e^{\gamma_2}Y_{i2}})]
\]

\[
= \sum_{i=1}^{n} E_{\gamma_1}[Y_{i1}(Y_{i2} - \frac{e^{\gamma_2}Y_{i1}}{1+e^{\gamma_2}Y_{i1}})] + \sum_{i=1}^{n} E_{\gamma_1}[Y_{i2}(Y_{i3} - \frac{e^{\gamma_2}Y_{i2}}{1+e^{\gamma_2}Y_{i2}})].
\]

Note that the first part of this expression has already been obtained in Section 6.1 when deriving an
equation for the power when \( T = 2 \). Incorporating the results into this equation for the mean gives

\[
E_{\gamma_1}[\sum_{i=1}^{n} Y_{i1}(Y_{i2} - \frac{e^{\gamma_2}Y_{i1}}{1+e^{\gamma_2}Y_{i1}}) + Y_{i2}(Y_{i3} - \frac{e^{\gamma_2}Y_{i2}}{1+e^{\gamma_2}Y_{i2}})]
\]

\[
= (k_1 - k_0)n p_1 + \sum_{i=1}^{n} [E_{\gamma_1}[Y_{i2}Y_{i3}] - E_{\gamma_1}[Y_{i2}] - \frac{e^{\gamma_2}Y_{i2}}{1+e^{\gamma_2}Y_{i2}}]
\]

\[
= (k_1 - k_0)n p_1 + \sum_{i=1}^{n} [P_{\gamma_1}[Y_{i1} = 1, Y_{i3} = 1] - P_{\gamma_1}[Y_{i2} = 1] - \frac{e^{\gamma_2}Y_{i2}}{1+e^{\gamma_2}Y_{i2}}]
\]

\[
= (k_1 - k_0)n p_1 + \sum_{i=1}^{n} \left[ -\frac{e^{\gamma_1}}{1+e^{\gamma_1}} P_{\gamma_1}[Y_{i2} = 1] - P_{\gamma_1}[Y_{i2} = 1] - \frac{e^{\gamma_2}Y_{i2}}{1+e^{\gamma_2}Y_{i2}} \right]
\]
\[(k_1 - k_0)np_1 + \sum_{j=1}^{n} P_{ij} [Y_{ij} = 1] (k_1 - k_0) = (k_1 - k_0)np_1 + n\mu_{2,\gamma_1} (k_1 - k_0) = n(k_1 - k_0)(p_1 + \mu_{2,\gamma_1})\]

where \(\mu_{2,\gamma_1} = k_1 k_1 + \frac{1}{2} (1 - k_1)\).

So now that we have obtained the mean of \(\sum_{i=1}^{n} [Y_{i1}(Y_{i2} - \frac{e^{2T_{i1}}}{1 + e^{2T_{i1}}}) + Y_{i2}(Y_{i3} - \frac{e^{2T_{i2}}}{1 + e^{2T_{i2}}})]\) under the alternative hypothesis \(H_a : \gamma = \gamma_1\), next we need to obtain its variance

\[
V_{\gamma_1} \left( \sum_{i=1}^{n} Y_{i1}(Y_{i2} - \frac{e^{2T_{i1}}}{1 + e^{2T_{i1}}}) + Y_{i2}(Y_{i3} - \frac{e^{2T_{i2}}}{1 + e^{2T_{i2}}}) \right) = \sum_{i=1}^{n} \text{Var}_{\gamma_1} \left[ Y_{i1}(Y_{i2} - \frac{e^{2T_{i1}}}{1 + e^{2T_{i1}}}) + Y_{i2}(Y_{i3} - \frac{e^{2T_{i2}}}{1 + e^{2T_{i2}}}) \right]
\]

\[
= \sum_{i=1}^{n} E_{\gamma_1} \left[ (Y_{i1}(Y_{i2} - \frac{e^{2T_{i1}}}{1 + e^{2T_{i1}}}) + Y_{i2}(Y_{i3} - \frac{e^{2T_{i2}}}{1 + e^{2T_{i2}}}))^2 \right] - \sum_{i=1}^{n} [(k_1 - k_0)(k_1 + \mu_{2,\gamma_1})]^2
\]

\[
= \sum_{i=1}^{n} E_{\gamma_1} \left[ (Y_{i1}(Y_{i2} - \frac{e^{2T_{i1}}}{1 + e^{2T_{i1}}}))^2 \right]
+ 2 \sum_{i=1}^{n} E \left[ Y_{i1}(Y_{i2} - \frac{e^{2T_{i1}}}{1 + e^{2T_{i1}}})Y_{i2}(Y_{i3} - \frac{e^{2T_{i2}}}{1 + e^{2T_{i2}}}) \right]
+ \sum_{i=1}^{n} E \left[ (Y_{i2}(Y_{i3} - \frac{e^{2T_{i2}}}{1 + e^{2T_{i2}}}))^2 \right] - n[(k_1 - k_0)(k_1 + \mu_{2,\gamma_1})]^2.
\]

Note again that the first term of the above equation has already been obtained in the previous section when we obtained the power function for the case where \(T = 2\). Thus

\[
V_{\gamma_1} \left( \sum_{i=1}^{n} Y_{i1}(Y_{i2} - \frac{e^{2T_{i1}}}{1 + e^{2T_{i1}}}) + Y_{i2}(Y_{i3} - \frac{e^{2T_{i2}}}{1 + e^{2T_{i2}}}) \right) = nk_1[k_1 - 2k_1k_0 + k_0^2]
+ 2 \sum_{i=1}^{n} E_{\gamma_1} \left[ Y_{i1}(Y_{i2} - \frac{e^{2T_{i1}}}{1 + e^{2T_{i1}}}) \right] Y_{i2}(Y_{i3} - \frac{e^{2T_{i2}}}{1 + e^{2T_{i2}}})]
\]
\[ + \sum_{t=1}^{n} E_{\gamma_t} \left[ \left( Y_{\alpha} - \frac{e^{70/2}}{1+e^{70/2}} \right) Y_{\alpha} \right] \left( Y_{\beta} - \frac{e^{70/2}}{1+e^{70/2}} \right) \right] n \left[ (k_1 - k_0)(k_1 + \mu_{2,\gamma_t}) \right]^2. \]

This simply leaves the necessity of evaluating the above two expectations in order to get the variance. We then consider next

\[
E_{\gamma_t}[Y_{\alpha}(Y_{\alpha} - \frac{e^{70/2}}{1+e^{70/2}})Y_{\beta}(Y_{\beta} - \frac{e^{70/2}}{1+e^{70/2}})] = E_{\gamma_t}[Y_{\alpha}Y_{\beta} Y_{\beta} - Y_{\alpha} \frac{e^{70/2}}{1+e^{70/2}} - Y_{\beta} \frac{e^{70/2}}{1+e^{70/2}} + \frac{e^{70/2}}{(1+e^{70/2})^2}] = E_{\gamma_t}[Y_{\alpha} Y_{\beta} Y_{\beta} - E_{\gamma_t}[Y_{\alpha} Y_{\beta} Y_{\beta} - \frac{e^{70/2}}{1+e^{70/2}}]] + E_{\gamma_t}[Y_{\alpha} Y_{\beta} \frac{e^{70/2}}{1+e^{70/2}} \frac{e^{70/2}}{1+e^{70/2}}] = P_{\gamma_t}[Y_{\alpha} = 1, Y_{\beta} = 1, Y_{\beta} = 1] - P_{\gamma_t}[Y_{\alpha} = 1, Y_{\beta} = 1, Y_{\beta} = 1](e^{70/2}) - P_{\gamma_t}[Y_{\alpha} = 1, Y_{\beta} = 1, Y_{\beta} = 1](\frac{e^{70/2}}{1+e^{70/2}}) + P_{\gamma_t}[Y_{\alpha} = 1, Y_{\beta} = 1, Y_{\beta} = 1](1 - k_0) - P_{\gamma_t}[Y_{\alpha} = 1, Y_{\beta} = 1](1 - k_0) = (1 - k_0)[P_{\gamma_t}[Y_{\alpha} = 1, Y_{\beta} = 1, Y_{\beta} = 1] - k_0 P_{\gamma_t}[Y_{\alpha} = 1, Y_{\beta} = 1]].
\]

Now, if we realize that

\[ P_{\gamma_t}[Y_{\alpha} = 1, Y_{\beta} = 1] P_{\gamma_t}[Y_{\alpha} = 1, Y_{\beta} = 1] P_{\gamma_t}[Y_{\alpha} = 1] = k_1 k_1 \]

and similarly, that

\[ P_{\gamma_t}[Y_{\alpha} = 1, Y_{\beta} = 1] = k_1 k_1 \]

we get

\[ E_{\gamma_t}[Y_{\alpha}(Y_{\alpha} - \frac{e^{70/2}}{1+e^{70/2}})Y_{\alpha} - \frac{e^{70/2}}{1+e^{70/2}})] \]
Next we find the other expectation:

\[
E_{Y_i}[[Y_2(Y_3 - \frac{e^{0.7 a}}{1+e^{0.7 a}})]^2] = E_{Y_i}[Y_2^2(Y_3 - \frac{e^{0.7 a}}{1+e^{0.7 a}})^2]
= E_{Y_i}[Y_2^2(Y_3 - 2Y_3(\frac{e^{0.7 a}}{1+e^{0.7 a}}) + (\frac{e^{0.7 a}}{1+e^{0.7 a}})^2)]
= E_{Y_i}[Y_2^2(Y_3)^2] - 2E_{Y_i}[Y_2^2Y_3(\frac{e^{0.7 a}}{1+e^{0.7 a}})] + E_{Y_i}[Y_2^2((\frac{e^{0.7 a}}{1+e^{0.7 a}})^2)]
= P_{Y_3}[Y_2 = 1, Y_3 = 1] - 2P_{Y_3}[Y_2 = 1, Y_3 = 1](\frac{e^{0.7 a}}{1+e^{0.7 a}}) + P_{Y_3}[Y_2 = 1](\frac{e^{0.7 a}}{1+e^{0.7 a}})^2
\]

With the equations for these two expectations, it is straightforward to obtain the variance of the statistic of interest:

\[
Var_{Y_i}[[\sum_{j=1}^n Y_j(Y_2 - \frac{e^{0.7 a}}{1+e^{0.7 a}}) + Y_2(Y_3 - \frac{e^{0.7 a}}{1+e^{0.7 a}})]]
= nk_1[k_1 - 2k_1k_0 + k_0^2] + nk_1\mu_{2,Y_1}(1 - k_0)(k_1 - k_0) + nk_1\mu_{2,Y_1}(k_1 - 2k_0)^2 + n\mu_{2,Y_1}[k_1(1 - 2k_0) + k_0^2] - n[(k_1 - k_0)(p_1 + \mu_{2,Y_1})]^2
\]
Since we have the mean and variance we are ready to look at the asymptotic distribution of the standardized form of the statistic. We can write this as:

\[
\frac{\sum_{i=1}^{n} Y_{ni}(Y_{i2} - \frac{\epsilon_{i1} b_{n}}{\epsilon_{i1} + \epsilon_{i2} b_{0}}) + Y_{ni}(Y_{i2} - \frac{\epsilon_{i1} b_{n}}{\epsilon_{i1} + \epsilon_{i2} b_{0}})}{\sqrt{\sum_{i=1}^{n} Y_{ni}(Y_{i2} - \frac{\epsilon_{i1} b_{n}}{\epsilon_{i1} + \epsilon_{i2} b_{0}}) + Y_{ni}(Y_{i2} - \frac{\epsilon_{i1} b_{n}}{\epsilon_{i1} + \epsilon_{i2} b_{0}})} - n(k_{1} - k_{0})(k_{1} + \mu_{2,1})}{((k_{1} - k_{0})(k_{1} + \mu_{2,1}))^{2}}} \to N(0, 1)
\]

or alternatively,

\[
\frac{\sum_{i=1}^{n} Y_{ni}(Y_{i2} - \frac{\epsilon_{i1} b_{n}}{\epsilon_{i1} + \epsilon_{i2} b_{0}}) + Y_{ni}(Y_{i2} - \frac{\epsilon_{i1} b_{n}}{\epsilon_{i1} + \epsilon_{i2} b_{0}})}{\sqrt{\sum_{i=1}^{n} Y_{ni}(Y_{i2} - \frac{\epsilon_{i1} b_{n}}{\epsilon_{i1} + \epsilon_{i2} b_{0}}) + Y_{ni}(Y_{i2} - \frac{\epsilon_{i1} b_{n}}{\epsilon_{i1} + \epsilon_{i2} b_{0}})} - n(k_{1} - k_{0})(k_{1} + \mu_{2,1})}{((k_{1} - k_{0})(k_{1} + \mu_{2,1}))^{2}}} \to N(0, 1).
\]

We can now write the power function as:

\[
\text{Power} = P\left(\sum_{i=1}^{n} Y_{ni}(Y_{i2} - \frac{\epsilon_{i1} b_{n}}{\epsilon_{i1} + \epsilon_{i2} b_{0}}) + Y_{ni}(Y_{i2} - \frac{\epsilon_{i1} b_{n}}{\epsilon_{i1} + \epsilon_{i2} b_{0}}) \mid \gamma = \gamma_{1} \text{ i.e. } k = k_{1}\right) > z_{\alpha/2} \sqrt{k_{0}(1 - k_{0})((k_{0} + 2) \sum_{i=1}^{n} P_{ni} + \frac{\alpha}{2})}
\]

\[
= P\left(\sum_{i=1}^{n} Y_{ni}(Y_{i2} - \frac{\epsilon_{i1} b_{n}}{\epsilon_{i1} + \epsilon_{i2} b_{0}}) + Y_{ni}(Y_{i2} - \frac{\epsilon_{i1} b_{n}}{\epsilon_{i1} + \epsilon_{i2} b_{0}}) \mid \gamma = \gamma_{1} \text{ i.e. } k = k_{1}\right) > z_{\alpha/2} \sqrt{k_{0}(1 - k_{0})((k_{0} + 2) \sum_{i=1}^{n} P_{ni} + \frac{\alpha}{2})}
\]

\[
\text{Power} = P\left(\sum_{i=1}^{n} Y_{ni}(Y_{i2} - \frac{\epsilon_{i1} b_{n}}{\epsilon_{i1} + \epsilon_{i2} b_{0}}) + Y_{ni}(Y_{i2} - \frac{\epsilon_{i1} b_{n}}{\epsilon_{i1} + \epsilon_{i2} b_{0}}) \mid \gamma = \gamma_{1} \text{ i.e. } k = k_{1}\right) < -z_{\alpha/2} \sqrt{k_{0}(1 - k_{0})((k_{0} + 2) \sum_{i=1}^{n} P_{ni} + \frac{\alpha}{2})}
\]

\[
= P\left(\sum_{i=1}^{n} Y_{ni}(Y_{i2} - \frac{\epsilon_{i1} b_{n}}{\epsilon_{i1} + \epsilon_{i2} b_{0}}) + Y_{ni}(Y_{i2} - \frac{\epsilon_{i1} b_{n}}{\epsilon_{i1} + \epsilon_{i2} b_{0}}) \mid \gamma = \gamma_{1} \text{ i.e. } k = k_{1}\right) < -z_{\alpha/2} \sqrt{k_{0}(1 - k_{0})((k_{0} + 2) \sum_{i=1}^{n} P_{ni} + \frac{\alpha}{2})}
\]

\[
\text{Power} = P\left(\sum_{i=1}^{n} Y_{ni}(Y_{i2} - \frac{\epsilon_{i1} b_{n}}{\epsilon_{i1} + \epsilon_{i2} b_{0}}) + Y_{ni}(Y_{i2} - \frac{\epsilon_{i1} b_{n}}{\epsilon_{i1} + \epsilon_{i2} b_{0}}) \mid \gamma = \gamma_{1} \text{ i.e. } k = k_{1}\right) < -z_{\alpha/2} \sqrt{k_{0}(1 - k_{0})((k_{0} + 2) \sum_{i=1}^{n} P_{ni} + \frac{\alpha}{2})}
\]

\[
= P\left(\sum_{i=1}^{n} Y_{ni}(Y_{i2} - \frac{\epsilon_{i1} b_{n}}{\epsilon_{i1} + \epsilon_{i2} b_{0}}) + Y_{ni}(Y_{i2} - \frac{\epsilon_{i1} b_{n}}{\epsilon_{i1} + \epsilon_{i2} b_{0}}) \mid \gamma = \gamma_{1} \text{ i.e. } k = k_{1}\right) < -z_{\alpha/2} \sqrt{k_{0}(1 - k_{0})((k_{0} + 2) \sum_{i=1}^{n} P_{ni} + \frac{\alpha}{2})}
\]

\[
= P\left(\sum_{i=1}^{n} Y_{ni}(Y_{i2} - \frac{\epsilon_{i1} b_{n}}{\epsilon_{i1} + \epsilon_{i2} b_{0}}) + Y_{ni}(Y_{i2} - \frac{\epsilon_{i1} b_{n}}{\epsilon_{i1} + \epsilon_{i2} b_{0}}) \mid \gamma = \gamma_{1} \text{ i.e. } k = k_{1}\right) < -z_{\alpha/2} \sqrt{k_{0}(1 - k_{0})((k_{0} + 2) \sum_{i=1}^{n} P_{ni} + \frac{\alpha}{2})}
\]
Bearing in mind that the asymptotic distribution of the statistic is normal, we re-write the formula for power as

\[
\text{Power} = P(Z > \frac{z_{\alpha/2} \sqrt{n(k_1-k_0)(k_0+2)} \sum_{i=1}^{n} P_{i1} + \frac{1}{2} - m(k_1-k_0)(k_1+\mu_{21})}{\sqrt{n(k_1-2k_1k_0+k_0^2)(k_1+\mu_{21}) + 2k_1(1-k_0)(k_1-k_0) - (k_1-k_0)(k_1+\mu_{21})^2}} \mid \gamma = \gamma_i, i.e. k = k_1) \]

Note that when \( p_{i1} = k \) are equal for all \( i = 1, ..., n \) clusters, the power formula simplifies to

\[
\text{Power} = 1 - \Phi\left(\frac{z_{\alpha/2} \sqrt{n(k_1-k_0)(k_0+2)} \sum_{i=1}^{n} P_{i1} + \frac{1}{2} - m(k_1-k_0)(k_1+\mu_{21})}{\sqrt{n(k_1-2k_1k_0+k_0^2)(k_1+\mu_{21}) + 2k_1(1-k_0)(k_1-k_0) - (k_1-k_0)(k_1+\mu_{21})^2}} \right)
\]
\[
\frac{\sqrt{\pi}(k_1-k_0)(k_1+\mu_2\gamma)}{\sqrt{[(k_1-2k_1\kappa_0+k_0^2)(k_1+\mu_2\gamma_2\tau_2)-2k_1k_1(1-k_0)(k_1-k_0)-\{(k_1-k_0)(k_1+\mu_2\gamma_2\tau_2)\}^2]}}
\]

\[+
\Phi\left(\frac{-\pi\mu\sqrt{k_0(1-k_0)(\mu_0+2k_1^{\frac{1}{2}})}}{\sqrt{[(k_1-2k_1\kappa_0+k_0^2)(k_0+\mu_2\gamma_2\tau_2)+2k_1k_1(1-k_0)(k_1-k_0)-\{(k_1-k_0)(k_1+\mu_2\gamma_2\tau_2)\}^2]}}\right).
\]

\[-\frac{\sqrt{\pi}(k_1-k_0)(k_1+\mu_2\gamma)}{\sqrt{[(k_1-2k_1\kappa_0+k_0^2)(k_1+\mu_2\gamma_2\tau_2)-2k_1k_1(1-k_0)(k_1-k_0)-\{(k_1-k_0)(k_1+\mu_2\gamma_2\tau_2)\}^2]}}\cdot
\]

Note that the power increases with \(n\), which is as we would expect intuitively.
7 Small Sample Inference

7.1 Derivation of the Distribution of the Sum of \( n \) Independent Non-Identically Distributed Bernoulli Random Variables

Note that to this point we have presented illustrations for the derivation of the power for the test \( H_0 : \gamma = \gamma_0 \) versus \( H_a : \gamma = \gamma_1 \) given that we use the asymptotic distribution of a suitable statistic.

For example, the test statistic introduced in Chapter 4 relies on \( \sum_{i=1}^{n} A_i \). However we may encounter situations where the sample size is relatively small and we are not able to apply the Central Limit Theorem. For such situations, the exact distribution of the test statistic would be of interest. In order to develop this exact distribution, it is necessary to first consider a derivation of the distribution of the independent sum \( \sum_{i=1}^{n} A_i \). We present here three different approaches for deriving the distribution. Two are based on the transformation method, while the other is based on moment generating functions. We show that complications arise when using the transformation method. We demonstrate this first.

In order to use the transformation method, the joint distribution of the independent random variables of interest is transformed to a joint distribution of the same dimensionality where one of the original random variables is replaced by the sum of all variables. From this transformed joint distribution, we derive the marginal distribution of the sum by summing out the variables from the transformed joint distribution that are not needed. We shall clarify this below.

In particular we are interested in the distribution of \( H_n = \sum_{i=1}^{n} A_i \) where \( A_i = Y_{1i}(Y_{2i} - \lambda_{12}) \)
under $H_0 : \gamma = 0$. We can first investigate the distribution of $A_i$, which for $T = 2$ can be enumerated as

<table>
<thead>
<tr>
<th>$Y_{i1}$</th>
<th>$Y_{i2}$</th>
<th>$A_i$</th>
<th>$P(a_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>$\frac{1}{2}$</td>
<td>$P_{\gamma=0}(Y_{i2} = 1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>$-\frac{1}{2}$</td>
<td>$P_{\gamma=0}(Y_{i2} = 0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>$P_{\gamma=0}(Y_{i2} = 1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>$P_{\gamma=0}(Y_{i2} = 0</td>
</tr>
</tbody>
</table>

This result can be simplified, as there are two different basic outcomes for the binary responses $Y_{i1}$ and $Y_{i2}$ that will produce $a_i = 0$. Combining gives the distribution

$A_i : \begin{pmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \end{pmatrix}$

$P(a_i) : \begin{pmatrix} \frac{1}{2}p_{11} & (1-p_{11}) & \frac{1}{2}p_{11} \end{pmatrix}$

Note that we can express the distribution of a given $A_i$ above with the use of indicator variables

$$P(A_i = x_i) = \frac{1}{2}p_{11} * I(x_i = \frac{1}{2}) + (1-p_{11}) * I(x_i = 0) + \frac{1}{2}p_{11} * I(x_i = -\frac{1}{2})$$

$$= \frac{1}{2}p_{11} * \{I(x_i = \frac{1}{2}) + I(x_i = -\frac{1}{2})\} + (1-p_{11}) * I(x_i = 0)$$

$$= \frac{1}{2}p_{11} * [1-I(x_i = 0)] + (1-p_{11}) * I(x_i = 0)$$

$$= \frac{1}{2}p_{11} + (1-\frac{3}{2}p_{11}) * I(x_i = 0)$$

where $I(X = x) = 1$ if $X = x$

0 otherwise
Thus, the joint distribution of the independent variables $A_1 \ldots A_n$ can be written as

$$P(a_1 \ldots a_n) = \prod_{i=1}^{n} P(a_i) = \prod_{i=1}^{n} [\frac{1}{2} p_{ii} + (1 - \frac{3}{2} p_{ii}) \times I(a_i = 0)]$$

Then, in order to derive the distribution of the sum $H_n = \sum_{i=1}^{n} A_i$ we use the transformation method.

Letting

$$Z = \sum_{i=1}^{n} A_i$$ with $W_i = A_i$ for all $i = 1, \ldots, n - 1$, that is $A_n = Z - \sum_{i=1}^{n-1} W_i$ and $A_i = W_i$ for all $i = 1, \ldots, n - 1$.

we can obtain the joint distribution of $W_1, \ldots, W_{n-1}, Z$, which we shall write as:

$$P(w_1, \ldots, w_{n-1}, z) =$$

$$\prod_{i=1}^{n-1} \left[ \frac{1}{2} p_{ii} + (1 - \frac{3}{2} p_{ii}) \times I(w_i = 0) \right] \times \left[ \frac{1}{2} p_{ii} + (1 - \frac{3}{2} p_{ii}) \times I(z - \sum_{j=1}^{n-1} w_i = 0) \right]$$

$$= P(w_1, \ldots, w_{n-1}) P(z|w_1, \ldots, w_{n-1})$$

where henceforth for simplicity

$$P(Z = z|W_1 = w_1) = (z|w_1),$$

$$P(w_1, \ldots, w_{n-1}) = P(W_1 = w_1, \ldots, W_{n-1} = w_{n-1})$$

$$P(z|w_1, \ldots, w_{n-1}) = P(Z = z|W_1 = w_1, \ldots, W_{n-1} = w_{n-1}).$$

Therefore

$$P(z) = \sum_{w_1} \cdots \sum_{w_{n-1}} P(w_1, \ldots, w_{n-1}) P(z|w_1, \ldots, w_{n-1})$$

$$= \sum_{w_1} \cdots \sum_{w_{n-1}} \left\{ \prod_{i=1}^{n-1} \left[ \frac{1}{2} p_{ii} + (1 - \frac{3}{2} p_{ii}) \times I(w_i = 0) \right] \right\}$$

$$\times \left[ \frac{1}{2} p_{ii} + (1 - \frac{3}{2} p_{ii}) \times I(z - \sum_{j=1}^{n-1} w_i = 0) \right].$$

This is notably difficult to compute for large values of $n$. For example if we let $n = 2$, the
\[ Z = H_2 = \sum_{i=1}^{2} A_i = A_1 + A_2, \text{ and} \]

\[ P(w_1, z) = \left[ \frac{1}{2} p_{11} + \left(1 - \frac{3}{2} p_{11}\right) \times I(w_1 = 0) \right] \]

\[ \times \left[ \frac{1}{2} p_{21} + \left(1 - \frac{3}{2} p_{21}\right) \times I(z - w_1 = 0) \right] \]

\[ = P(w_1)P(z|w_1) \]

where

\[ P(z|w_1) = \frac{1}{2} p_{21} + \left(1 - \frac{3}{2} p_{21}\right) \times I(z - w_1 = 0). \]

Note that since \( Z = \sum_{i=1}^{2} A_i = A_1 + A_2 \) if \( W_1 = 0 \) then \( Z \) cannot be 1 or -1 and so

\[ P(Z = 1|W_1 = 0) = P(Z = -1|W_1 = 0) = 0. \]

Alternatively, if \( W_1 = -\frac{1}{2} \) then \( Z \) cannot be 1 or \( \frac{1}{2} \)

so \( P(Z = 1|W_1 = -\frac{1}{2}) = P(Z = \frac{1}{2}|W_1 = -\frac{1}{2}) = 0, \) and if \( W_1 = \frac{1}{2} \) then \( Z \) cannot be -1 or \( -\frac{1}{2} \)

and so \( P(Z = -1|W_1 = \frac{1}{2}) = P(Z = -\frac{1}{2}|W_1 = \frac{1}{2}) = 0. \) Since \( P(z) = \sum_{w_1} P(w_1)P(z|w_1), \) these results give

\[ P(1) = \sum_{w_1} P(w_1)P(1|w_1) = P(-\frac{1}{2})P(1 - \frac{1}{2}) + P(0)P(1|0) + P(\frac{1}{2})P(1|\frac{1}{2}) = 0 + 0 \]

\[ + \left[ \frac{1}{2} p_{11} + \left(1 - \frac{3}{2} p_{11}\right) \times I(w_1 = \frac{1}{2}) \right] \times \left[ \frac{1}{2} p_{21} + \left(1 - \frac{3}{2} p_{21}\right) \times I(1 - \frac{1}{2} = 0) \right] = \frac{1}{2} p_{11} \cdot \frac{1}{2} p_{21} \]

\[ - \frac{1}{4} p_{11} p_{21} = \frac{1}{2} \prod_{j=1}^{2} p_{j1}, \]

\[ P(\frac{1}{2}) = P(-\frac{1}{2})P(\frac{1}{2}|1 - \frac{1}{2}) + P(0)P(\frac{1}{2}|0) + P(\frac{1}{2})P(\frac{1}{2}|\frac{1}{2}) = 0 + P(0)P(\frac{1}{2}|0) + P(\frac{1}{2})P(\frac{1}{2}|\frac{1}{2}) \]

\[ = \left[ \frac{1}{2} p_{11} + \left(1 - \frac{3}{2} p_{11}\right) \times I(w_1 = 0) \right] \times \left[ \frac{1}{2} p_{21} + \left(1 - \frac{3}{2} p_{21}\right) \times I(\frac{1}{2} - 0 = 0) \right] \]

\[ + \left[ \frac{1}{2} p_{11} + \left(1 - \frac{3}{2} p_{11}\right) \times I(w_1 = \frac{1}{2}) \right] \times \left[ \frac{1}{2} p_{21} + \left(1 - \frac{3}{2} p_{21}\right) \times I(\frac{1}{2} - \frac{1}{2} = 0) \right] \]

\[ - \frac{1}{4} p_{21}(1 - p_{11}) + \frac{1}{2} p_{11}(1 - p_{21}) = \frac{1}{2} \sum_{j=1}^{2} p_{j1}(1 - p_{k1}), \]
\[ P(0) = P(-\frac{1}{2})P(0| -\frac{1}{2}) + P(0)P(0|0) + P(\frac{1}{2})P(0|\frac{1}{2}) = \]

\[ \left[ \frac{1}{2} p_{11} + (1 - \frac{3}{2} p_{11}) \cdot I(w_1 = -\frac{1}{2}) \right] \cdot \left[ \frac{1}{2} p_{21} + (1 - \frac{3}{2} p_{21}) \cdot I(0 + \frac{1}{2} = 0) \right] \]

\[ + \left[ \frac{1}{2} p_{11} + (1 - \frac{3}{2} p_{11}) \cdot I(w_1 = 0) \right] \cdot \left[ \frac{1}{2} p_{21} + (1 - \frac{3}{2} p_{21}) \cdot I(0 - 0 = 0) \right] \]

\[ + \left[ \frac{1}{2} p_{11} + (1 - \frac{3}{2} p_{11}) \cdot I(w_1 = \frac{1}{2}) \right] \cdot \left[ \frac{1}{2} p_{21} + (1 - \frac{3}{2} p_{21}) \cdot I(0 - \frac{1}{2} = 0) \right] \]

\[ = \frac{1}{2} p_{11} \cdot \frac{1}{2} p_{21} + (1 - p_{11})(1 - p_{21}) + \frac{1}{2} p_{11} \cdot \frac{1}{2} p_{21} = \frac{1}{2} p_{11} p_{21} + (1 - p_{11})(1 - p_{21}) \]

\[ P(-\frac{1}{2}) = P(-\frac{1}{2})P(-\frac{1}{2} - \frac{1}{2}) + P(0)P(-\frac{1}{2}|0) + P(\frac{1}{2})P(-\frac{1}{2}|\frac{1}{2}) \]

\[ = P(-\frac{1}{2})P(-\frac{1}{2} - \frac{1}{2}) + P(0)P(-\frac{1}{2}|0) \]

\[ - \left[ \frac{1}{2} p_{11} + (1 - \frac{3}{2} p_{11}) \cdot I(w_1 = -\frac{1}{2}) \right] \cdot \left[ \frac{1}{2} p_{21} + (1 - \frac{3}{2} p_{21}) \cdot I(-\frac{1}{2} + \frac{1}{2} = 0) \right] \]

\[ + \left[ \frac{1}{2} p_{11} + (1 - \frac{3}{2} p_{11}) \cdot I(w_1 = 0) \right] \cdot \left[ \frac{1}{2} p_{21} + (1 - \frac{3}{2} p_{21}) \cdot I(-\frac{1}{2} - 0 = 0) \right] \]

\[ = \frac{1}{2} p_{11} (1 - p_{21}) + (1 - p_{11}) \cdot \frac{1}{2} p_{21} = \frac{1}{2} \sum_{j=1}^{2} p_{1j} (1 - p_{2j}) \]

and finally:

\[ P(-1) = P(-\frac{1}{2})P(-1 - \frac{1}{2}) + P(0)P(-1|0) + P(\frac{1}{2})P(-1|\frac{1}{2}) = P(-\frac{1}{2})P(-1 - \frac{1}{2}) \]

\[ = \left[ \frac{1}{2} p_{11} + (1 - \frac{3}{2} p_{11}) \cdot I(w_1 = -\frac{1}{2}) \right] \cdot \left[ \frac{1}{2} p_{21} + (1 - \frac{3}{2} p_{21}) \cdot I(-1 + \frac{1}{2} = 0) \right] \]

\[ = \frac{1}{2} p_{11} \cdot \frac{1}{2} p_{21} = \frac{1}{2} \sum_{j=1}^{2} p_{1j} p_{2j}. \]

Note that it can be verified that \( \sum_z P(Z = z) = 1 \). In addition, the distribution obtained is symmetric about \( Z = 0 \).

Although the above does give us a form for deriving the distribution of the sum, the form is unfortunately too complicated to work with. In order to get a relatively simple functional
representation of the distribution of \( H_n \) we require a different method than that used above.

Specifically, instead of writing \( P(A_t = x_t) \) as a function of indicator variables, we can express the distribution of a given \( A_i \) as follows:

\[
P(A_i = x_i) = \left( \frac{p_{it}}{2} \right)^{4x_i^2}(1 - p_{it})^{1 - 4x_i^2} \quad \text{where } x_i = -\frac{1}{2}, 0, \frac{1}{2}
\]

\[
0 \quad \text{otherwise}.
\]

Thus, the joint distribution of the independent variables \( a_1 \ldots a_n \) can be written as

\[
P(a_1 \ldots a_n) = \prod_{i=1}^n P(a_i) = \prod_{i=1}^n \left( \frac{p_{it}}{2} \right)^{4x_i^2}(1 - p_{it})^{1 - 4x_i^2}.
\]

Then, in order to derive the distribution of the sum \( H_n = \sum_{i=1}^n A_i \), we use the transformation method.

Letting

\[
Z = \sum_{i=1}^n A_i \quad \text{with} \quad W_i = A_i \quad \text{for all } i = 1, ..., n - 1, \quad \text{that is} \quad A_n = Z - \sum_{i=1}^{n-1} W_i \quad \text{and} \quad A_i = W_i \quad \text{for all } i = 1, ..., n - 1
\]

we can obtain the joint distribution of \( W_1, \ldots, W_{n-1}, Z \), which we shall write as:

\[
P(w_1, \ldots, w_{n-1}, z) = \left\{ \prod_{i=1}^{n-1} \left( \frac{p_{it}}{2} \right)^{4w_i^2}(1 - p_{it})^{1 - 4w_i^2} \right\}
\]

\[
\times \left( \frac{p_{nt}}{2} \right)^{4r^2-\sum_{i=1}^{n-1} w_i^2 - \sum_{i=1}^{n-1} w_i} \left( 1 - p_{nt} \right)^{r^2 - \sum_{i=1}^{n-1} w_i^2}
\]

\[
= P(w_1, \ldots, w_{n-1})P(z|w_1, \ldots, w_{n-1}).
\]

where \( z \) can take on the values \( k \) for \( k = 0, \ldots, n \) and \( w_i \) can take on the values \( -\frac{1}{2}, 0, \frac{1}{2} \) for \( i = 1, \ldots, n - 1 \). Thus we obtain the marginal distribution of \( Z \) by summing the above joint distribution over all values of \( w_1, \ldots, w_{n-1} \):

\[
P(z) = \sum_{w_1} \ldots \sum_{w_{n-1}} P(w_1, \ldots, w_{n-1})P(z|w_1, \ldots, w_{n-1})
\]
This again is notably difficult to compute for large values of \( n \). For example if we let \( n = 2 \),

\[ Z = H_2 = \sum_{i=1}^{2} A_i = A_1 + A_2, \]  

and

\[ P(w_1, z) = \left( \frac{p_{i1}}{2} \right)^{4w_1^2} (1 - p_{i1})^{1 - 4w_1^2} \times \left( \frac{p_{i2}}{2} \right)^{4(z - w_1)^2} (1 - p_{i2})^{1 - 4(z - w_1)^2} \]

\[ = P(w_1)P(z|w_1) \]

where

\[ P(z|w_1) = \left( \frac{p_{i1}}{2} \right)^{4(z - w_1)^2} (1 - p_{i2})^{1 - 4(z - w_1)^2} \]

Note that since \( Z = \sum_{i=1}^{2} A_i = A_1 + A_2 \), and \( W_1 = 0 \) then \( Z \) cannot be 1 or -1 and so

\[ P(Z = 1|W_1 = 0) = P(Z = -1|W_1 = 0) = 0. \]

Alternatively, if \( W_1 = -\frac{1}{2} \) then \( Z \) cannot be 1 or \( \frac{1}{2} \)

so

\[ P(Z = 1|W_1 = -\frac{1}{2}) = P(Z = \frac{1}{2}|W_1 = -\frac{1}{2}) = 0, \]

and if \( W_1 = \frac{1}{2} \) then \( Z \) cannot be -1 or \( \frac{1}{2} \)

and so

\[ P(Z = -1|W_1 = \frac{1}{2}) = P(Z = -\frac{1}{2}|W_1 = \frac{1}{2}) = 0. \]

Thus, using the fact that

\[ P(z) = \sum_{w_1} P(w_1)P(z|w_1). \]

gives us

\[ P(1) = \sum_{w_1} P(w_1)P(1|w_1) = P(-\frac{1}{2})P(1|\frac{1}{2}) + P(0)P(1|0) + P(\frac{1}{2})P(1|\frac{1}{2}) = 0 + 0 \]

\[ + \left( \frac{p_{i1}}{2} \right)^{4\frac{1}{2}} (1 - p_{i1})^{1 - 4\frac{1}{2}} \times \left( \frac{p_{i2}}{2} \right)^{4(0 - \frac{1}{2})^2} (1 - p_{i2})^{1 - 4(0 - \frac{1}{2})^2} = \frac{1}{2} p_{i1} \frac{1}{2} p_{i2} = \frac{1}{4} p_{i1} p_{i2} \]

\[ - \frac{1}{2} \prod_{i=1}^{2} P_{i1} \]

\[ P\left(\frac{1}{2}\right) = P(-\frac{1}{2})P(\frac{1}{2}|-\frac{1}{2}) + P(0)P(\frac{1}{2}|0) + P(\frac{1}{2})P(\frac{1}{2}|\frac{1}{2}) = 0 + P(0)P(\frac{1}{2}|0) + P(\frac{1}{2})P(\frac{1}{2}|\frac{1}{2}) \]
\[
\begin{align*}
&= \left( \left( \frac{p_{11}}{2} \right)^{4+0^2} (1 - p_{11})^{1-4+0^2} \right) \cdot \left( \frac{p_{21}}{2} \right)^{4\frac{1}{2}-0^2} (1 - p_{21})^{1-4\frac{1}{2}-0^2} \\
&\quad + \left( \left( \frac{p_{11}}{2} \right)^{4\frac{1}{2}+0^2} (1 - p_{11})^{1-4\frac{1}{2}+0^2} \right) \cdot \left( \frac{p_{21}}{2} \right)^{4\frac{1}{2}-\frac{1}{2}^2} (1 - p_{21})^{1-4\frac{1}{2}-\frac{1}{2}^2} \\
&= \frac{1}{2} p_{21} (1 - p_{11}) + \frac{1}{2} p_{11} (1 - p_{12}) = \frac{1}{2} \sum_{j=1}^{2} p_{j1} (1 - p_{ki}) \\
& P(0) = P(0)P(0|0) + P(0)P(0|0) + P(0)P(0|0) = \\
&\quad \left( \left( \frac{p_{11}}{2} \right)^{4(-\frac{1}{2})^2} (1 - p_{11})^{1-4(-\frac{1}{2})^2} \right) \cdot \left( \frac{p_{21}}{2} \right)^{4\left(0-\frac{1}{2}\right)^2} (1 - p_{21})^{1-4\left(0-\frac{1}{2}\right)^2} \\
&\quad + \left( \left( \frac{p_{11}}{2} \right)^{4^2} (1 - p_{11})^{1-4^2} \right) \cdot \left( \frac{p_{21}}{2} \right)^{4\left(0-\frac{1}{2}\right)^2} (1 - p_{21})^{1-4\left(0-\frac{1}{2}\right)^2} \\
&= \frac{1}{2} p_{11} \frac{1}{2} p_{21} + (1 - p_{11})(1 - p_{21}) + \frac{1}{2} p_{11} \frac{1}{2} p_{21} = \frac{1}{2} p_{11} p_{21} + (1 - p_{11})(1 - p_{21}) \\
& P(-\frac{1}{2}) = P(-\frac{1}{2})P(-\frac{1}{2}|-\frac{1}{2}) + P(0)P(-\frac{1}{2}|0) + P(\frac{1}{2})P(-\frac{1}{2}|\frac{1}{2}) = \\
&\quad \left( \left( \frac{p_{11}}{2} \right)^{4(-\frac{1}{2})^2} (1 - p_{11})^{1-4(-\frac{1}{2})^2} \right) \cdot \left( \frac{p_{21}}{2} \right)^{4\left(0-\frac{1}{2}\right)^2} (1 - p_{21})^{1-4\left(0-\frac{1}{2}\right)^2} \\
&+ \left( \left( \frac{p_{11}}{2} \right)^{4^2} (1 - p_{11})^{1-4^2} \right) \cdot \left( \frac{p_{21}}{2} \right)^{4\left(0-\frac{1}{2}\right)^2} (1 - p_{21})^{1-4\left(0-\frac{1}{2}\right)^2} \\
&= \frac{1}{2} P(1) (1 - p_{21}) + (1 - p_{11}) \frac{1}{2} p_{21} = \frac{1}{2} \sum_{j=1}^{2} p_{j1} (1 - p_{ki}) \\
\text{and finally:} \\
& P(-1) = P(-\frac{1}{2})P(-\frac{1}{2}|-\frac{1}{2}) + P(0)P(-\frac{1}{2}|0) + P(\frac{1}{2})P(-\frac{1}{2}|\frac{1}{2}) = P(-\frac{1}{2})P(-\frac{1}{2}|-\frac{1}{2}) = \\
&\quad \left( \left( \frac{p_{11}}{2} \right)^{4(-\frac{1}{2})^2} (1 - p_{11})^{1-4(-\frac{1}{2})^2} \right) \cdot \left( \frac{p_{21}}{2} \right)^{4\left(-1+\frac{1}{2}\right)^2} (1 - p_{21})^{1-4\left(-1+\frac{1}{2}\right)^2} \\
&= \frac{1}{2} p_{11} \frac{1}{2} p_{21} \frac{1}{2} \prod_{j=1}^{2} p_{j1} \\
\text{where } \sum_{z} P(Z = z) = 1. \text{ Note again that it can be verified that } \sum_{z} P(Z = z) = 1. \text{ In addition, the distribution obtained is symmetric about } Z = 0. \text{ However, as before, this method is not practical. In order to avoid the above complications which seem to stem from the use of the transformation}
method, we shall consider an approach that takes advantage of a particular form of the mgf of $A_t$.
Since the derivation of the exact distribution of the sum $\sum_i^n A_i$ is quite complicated we shall start
with the derivation of the sum of independent nonidentically distributed Bernoulli random variable.
This will serve as a guideline for the derivation of the distribution of the test statistic of interest. The
test statistic distribution will therefore follow. Note that the notation used in the remainder of this
chapter have no relationship to any of the notation that has been introduced up to this point.

Let $X_1, \ldots, X_n$ be independent random variables such that $X_i$ is distributed as a Bernoulli($p_i$) for
all $i = 1, \ldots, n$. The probability density function (pdf) of $X_i$ is given by

$$P(X_i = x_i) = p_i^{x_i}(1 - p_i)^{1-x_i} \text{ where } x_i = 0, 1$$

$$0 \text{ otherwise}$$

while the moment generating function (mgf) is:

$$M_{X_i}(t) = (1 - p_i) + p_i e^t.$$  

**THEOREM 7.1:**

Let $X_1, \ldots, X_n$ be independent random variables such that $X_i$ is distributed as a Bernoulli($p_i$) for
all $i = 1, \ldots, n$. Then the pdf of $\sum_{i=1}^n X_i$ is given by:

$$P(\sum_{i=1}^n X_i = n - j) = \sum_{k=0}^j (-1)^k \binom{n-j+k}{k} \sum_{n-j+k}^n \prod_{i=n-j-k}^n p_i \text{ where } n - j = 1, \ldots, n$$

$$0 \text{ otherwise}.$$  

**PROOF:**
Since \( X_1, \ldots, X_n \) are independent, the mgf of their sum gives us the following:

\[
M_{\sum_{i=1}^{n} X_i}(t) = \prod_{i=1}^{n} M_{X_i}(t) = \prod_{i=1}^{n} [(1 - p_i) + p_i e^t] = \prod_{i=1}^{n} [1 + p_i (e^t - 1)] = \prod_{i=1}^{n} [1 + b p_i]
\]

where \( b = (e^t - 1) \). Therefore

\[
M_{\sum_{i=1}^{n} X_i}(t) = 1 + b \sum_{i=1}^{n} p_i + b^2 \sum_{i=1}^{n} \prod_2 p_i + b^3 \sum_{i=1}^{n} \prod_3 p_i + \ldots + b^k \sum_{i=1}^{n} \prod_k p_i + \ldots + b^n \sum_{i=1}^{n} \prod_n p_i
\]

where the notation \( \sum_{i=1}^{n} \prod_k p_i \) symbolizes the sum of all possible combinations of \( k \) possible \( p \)s from \( p_1, \ldots, p_n \). For example, \( \sum_{i=1}^{n} \prod_1 p_i = \sum_{i=1}^{n} p_i \) is the sum of all possible products of one of the \( p_i \)s. Similarly, \( \sum_{i=1}^{n} \prod_2 p_i \) is the sum of all possible combinations \( p_i p_j \) where \( i \neq j \) from \( p_1, \ldots, p_n \).

Note again that \( k \) has no relation to the conditional distribution investigated in previous sections.

The expression \( b^k = (e^t - 1)^k \) can be explicitly obtained through a binomial expansion

\[
b^k = (e^t - 1)^k = \sum_{t=0}^{k} \binom{k}{t} e^{kt} (-1)^{k-t}.
\]

Therefore substituting this expression for \( b^k \) for all \( k = 1, \ldots, n \) gives

\[
M_{\sum_{i=1}^{n} X_i}(t) = 1 + (e^t - 1) \sum_{i=1}^{n} p_i + [\sum_{t=0}^{2} \binom{2}{t} e^{2t} (-1)^{t}] \sum_{i=1}^{n} \prod_2 p_i + \ldots + [\sum_{t=0}^{3} \binom{3}{t} e^{3t} (-1)^{t}] \sum_{i=1}^{n} \prod_3 p_i + \ldots + [\sum_{t=0}^{k} \binom{k}{t} e^{kt} (-1)^{t}] \sum_{i=1}^{n} \prod_k p_i + \ldots + [\sum_{t=0}^{n} \binom{n}{t} e^{nt} (-1)^{t}] \sum_{i=1}^{n} \prod_n p_i
\]
\[= 1 + (e^t - 1) \sum_{i=1}^{n} p_i + \left[ \sum_{j=0}^{2} \binom{2}{j} e^{(2-j)t} (-1)^j \sum_{i=1}^{n} \prod_{i=1}^{n} p_i \right]
\]
\[+ \left[ \sum_{j=0}^{1} \binom{1}{j} e^{(1-j)t} (-1)^j \sum_{i=1}^{n} \prod_{i=1}^{n} p_i \right] + \ldots \]
\[+ \left[ \sum_{j=0}^{n-2} \binom{n-2}{j} e^{(n-2-j)t} (-1)^j \sum_{i=1}^{n} \prod_{i=1}^{n} p_i \right] + \ldots + \left[ \sum_{j=0}^{n-3} \binom{n-3}{j} e^{(n-3-j)t} (-1)^j \sum_{i=1}^{n} \prod_{i=1}^{n} p_i \right]
\]
\[+ \left[ \sum_{j=0}^{n-4} \binom{n-4}{j} e^{(n-4-j)t} (-1)^j \sum_{i=1}^{n} \prod_{i=1}^{n} p_i \right] + \ldots + \left[ \sum_{j=0}^{n-5} \binom{n-5}{j} e^{(n-5-j)t} (-1)^j \sum_{i=1}^{n} \prod_{i=1}^{n} p_i \right]
\]
\[+ \ldots + \left[ \sum_{j=0}^{n-6} \binom{n-6}{j} e^{(n-6-j)t} (-1)^j \sum_{i=1}^{n} \prod_{i=1}^{n} p_i \right]
\]

To get a clearer picture regarding the mgf, it is necessary to expand the following four terms in the expression:

\[
\left[ \sum_{j=0}^{n} \binom{n}{j} e^{(n-j)t} (-1)^j \sum_{i=1}^{n} \prod_{i=1}^{n} p_i \right]
\]
\[
= \binom{n}{0} e^{(n-0)t} (-1)^0 \sum_{i=1}^{n} \prod_{i=1}^{n} p_i + \binom{n}{1} e^{(n-1)t} (-1)^1 \sum_{i=1}^{n} \prod_{i=1}^{n} p_i + \binom{n}{2} e^{(n-2)t} (-1)^2 \sum_{i=1}^{n} \prod_{i=1}^{n} p_i + \ldots + \binom{n}{n} e^{(n-n)t} (-1)^n \sum_{i=1}^{n} \prod_{i=1}^{n} p_i
\]

and

\[
\left[ \sum_{j=0}^{n-1} \binom{n-1}{j} e^{(n-1-j)t} (-1)^j \sum_{i=1}^{n-1} \prod_{i=1}^{n-1} p_i \right]
\]
\[
= \binom{n-1}{0} e^{(n-1-0)t} (-1)^0 \sum_{i=1}^{n-1} \prod_{i=1}^{n-1} p_i + \binom{n-1}{1} e^{(n-1-1)t} (-1)^1 \sum_{i=1}^{n-1} \prod_{i=1}^{n-1} p_i + \ldots + \binom{n-1}{n-1} e^{(n-1-n+1)t} (-1)^{n-1} \sum_{i=1}^{n-1} \prod_{i=1}^{n-1} p_i
\]

as well as

\[
\left[ \sum_{j=0}^{n-2} \binom{n-2}{j} e^{(n-2-j)t} (-1)^j \sum_{i=1}^{n-2} \prod_{i=1}^{n-2} p_i \right]
\]
\[67\]

\[
= (s^{-2})e^0 \sum_{i=1}^{n} \prod_{j=1}^{n-1} P_i + (s^{-2})e^0 \sum_{i=1}^{n} \prod_{j=1}^{n-1} P_i \\
+ [\sum_{i=2}^{n-1} (s^{-2})e^0 \prod_{j=1}^{n-1} P_i] 
\]

and

\[
[\sum_{i=0}^{n-3} (s^{-3})e^0 \prod_{j=1}^{n-1} P_i] 
\]

Note that the above terms were considered in an order that is opposite to their order of appearance in
the equation for \(M \sum_{i=1}^{n} x_i(t)\). Nevertheless, substituting these expansions into \(M \sum_{i=1}^{n} x_i(t)\) and
rearranging the terms yields the following, which shows an emerging pattern.

\[
M \sum_{i=1}^{n} x_i(t) = \\
\sum_{i=1}^{n} \prod_{j=1}^{n} P_i + \sum_{i=1}^{n} \prod_{j=1}^{n-1} P_i + (s^{-1}) \sum_{i=1}^{n} \prod_{j=1}^{n-2} P_i \\
+ [\sum_{i=2}^{n} (s^{-2}) \prod_{j=1}^{n-2} P_i] 
\]

and which gives us the mgf of $Y = \sum_{i=1}^{n} X_i$. Note that since $Y = \sum_{i=1}^{n} X_i$ is a discrete random variable (in particular we know that the support of $Y = \sum_{i=1}^{n} X_i$ is on the set $1, \ldots, n$) then its mgf is of the form

$$M_Y(t) = E[e^{itY}] = \sum_{i=1}^{n} e^{it} P(Y = i) =$$

$$e^{0P(Y = 0)} + e^{iP(Y = 1)} + e^{2P(Y = 2)} + \ldots + e^{kP(Y = k)} + \ldots + e^{nP(Y = n)}.$$

Thus if we have $M_Y(t)$ in the above form we can now for instance get the probability $P(Y = k)$ directly from the mgf since it is going to be the coefficient of $e^{it}$ when the mgf is factorized in the manner above.

Using the same approach we can now get the probability $P(\sum_{i=1}^{n} X_i = n - j)$ by picking the coefficient associated with $e^{i(n-j)}$. Therefore, we have the following probability distribution function for $\sum_{i=1}^{n} X_i$:

$$P(\sum_{i=1}^{n} X_i = n - j) = \sum_{k=0}^{j} (-1)^{k(n-j+k)} \sum_{i=1}^{n} \prod_{i\neq j+k} p_i \text{ where } n-j = 1, \ldots, n$$

0 otherwise,

which is what we were trying to prove.
7.2 Derivation of the Distribution of the Rao Score Statistic: Two Time Periods

Now that we have seen the method for deriving the distribution of the sum of independent non-identically distributed Bernoulli random variables it should be easier to follow the following derivation of the distribution of the test statistic given by \( S \) in Chapter 4.

We start by considering the exact probability distribution for \( \sum_{i=1}^{n} A_i \) in the simple case where \( T = 2 \). The mgf for any given \( A_i \) can be written as:

\[
M_{A_i}(t) = 1 + p_{11}[(1-k)e^{-kt} + ke^{(1-k)t} - 1].
\]

Now if we set \( H_n = \sum_{i=1}^{n} A_i \), we get the following

\[
M_{H_n}(t) = \prod_{i=1}^{n} M_{A_i}(t) = \prod_{i=1}^{n} [1 + p_{11}[(1-k)e^{-kt} + ke^{(1-k)t} - 1]] = \prod_{i=1}^{n} [1 + b p_{11}]
\]

where \( b = (1-k)e^{-kt} + ke^{(1-k)t} - 1 \). Expanding the expression for the mgf allows us to write that

\[
M_{H_n}(t) = 1 + \sum_{i=1}^{n} \prod_{j=1}^{n} p_{11} + b^n \sum_{i=1}^{n} \prod_{j=2}^{n} p_{11} + \ldots + b^n \sum_{i=1}^{n} \prod_{j=n}^{n} p_{11}.
\]

According to the multinomial theorem

\[
(x_1 + x_2 + x_3)^j = \sum_{k_1 \geq 0} \sum_{k_2 \geq 0} \sum_{k_3 \geq 0} (\begin{pmatrix} j \\ k_1, k_2, k_3 \end{pmatrix}) x_1^{k_1} x_2^{k_2} x_3^{k_3}
\]

where \( \sum_{i=1}^{3} k_i = j \). Thus, returning to the expression for the mgf for \( A_i \), we can write that

\[
[1 + p_{11}[(1-k)e^{-kt} + ke^{(1-k)t} - 1]]^n = \sum_{a_1 \geq 0} \sum_{a_2 \geq 0} \sum_{a_3 \geq 0} (\begin{pmatrix} n \\ a_1, a_2, a_3 \end{pmatrix}) [(1-k)e^{-kt}]^{a_1} [ke^{(1-k)t}]^{a_2} (-1)^{a_3}
\]

where \( \sum_{i=1}^{3} a_i = n \).

Next we put the above equation in a form that will be easier to follow by putting together all
of the exponents. We need to do this since we need to eventually get a mgf that is a polynomial in $e^t$ in order to be able to derive the probabilities in the same manner we did for the sum of independent non identically distributed Bernoulli random variables. Thus,

$$[1 + p_n[((1 - k)e^{-kt} + ke^{(1-k)t} - 1)]^n$$

$$= \sum_{a_1} \sum_{a_2} \sum_{a_3} (a_1, a_2, a_3) \left(1 - k\right)^{a_1} e^{-ka_1} k^{a_2} e^{(1-k)a_2} (-1)^{a_3}$$

$$= \sum_{a_1} \sum_{a_2} \sum_{a_3} (a_1, a_2, a_3) \left(1 - k\right)^{a_1} k^{a_2} e^{((1-k)a_2 - ka_1)} (-1)^{a_3}$$

and similarly

$$b^{n-1} = \sum_{a_1} \sum_{a_2} \sum_{a_3} (a_1, a_2, a_3) \left(1 - k\right)^{a_1} k^{a_2} e^{(a_2 - k(a_1 + a_2))} (-1)^{a_3}$$

such that $\sum_{i=1}^{3} a_i = n - 1$.

Now if $\sum_{i=1}^{3} a_i = n$ then $a_1 + a_2 = n - a_3$ and we can rewrite $b^n$ and express it as:

$$b^n = \sum_{a_1} \sum_{a_2} \sum_{a_3} (a_1, a_2, a_3) \left(1 - k\right)^{a_1} k^{a_2} e^{(a_2 - k(a_1 + a_2))} (-1)^{a_3}$$

$$= \sum_{a_3=0}^{n} \sum_{a_1+a_2=0}^{n-a_3} (a_1, a_2, a_3) \left(1 - k\right)^{a_1} k^{a_2} e^{((a_2 - k(a_1 + a_2)))} (-1)^{a_3}$$

$$= \sum_{a_1+a_2=n} (a_1, a_2, 0) \left(1 - k\right)^{a_1} k^{a_2} e^{(a_2 - k(n))} (-1)^{0}$$

$$+ \sum_{a_1+a_2=n-1} (a_1, a_2, 1) \left(1 - k\right)^{a_1} k^{a_2} e^{((a_2 - k(n-1)))} (-1)^{1}$$

$$+ \sum_{a_1+a_2=n-2} (a_1, a_2, 2) \left(1 - k\right)^{a_1} k^{a_2} e^{((a_2 - k(n-2)))} (-1)^{2}$$

$$+ \ldots$$

$$+ \sum_{a_1+a_2=n-1} (a_1, a_2, n-1) \left(1 - k\right)^{a_1} k^{a_2} e^{((a_2 - k(n-n)))} (-1)^{n-1}$$
\[ \sum_{a_1+a_2=0} \binom{n}{a_1 a_2} (1 - k)^{a_1} k^{a_2} e^{\theta(a_2 - k(m-\delta))} (-1)^n \]

where the last term
\[ \sum_{a_1+a_2=0} \binom{n}{a_1 a_2} (1 - k)^{a_1} k^{a_2} e^{\theta(a_2 - k(m-\delta))} (-1)^n = (-1)^n. \]

We need to break the equation above into term-by-term expressions in order to be able to regroup the terms later into a polynomial in \( e^\theta \) with a distinct form and coefficients that will enable us to specify the probabilities that \( H_n \) takes on a given value. This point will become clearer as the development of the equation proceeds.

First,
\[ \sum_{a_1+a_2=0} \binom{n}{a_1 a_2} (1 - k)^{a_1} k^{a_2} e^{\theta(a_2 - k(n-1))} (-1)^0 = \]
\[ \sum_{a_2=0}^n \binom{n}{a_2} (1 - k)^{n-a_2} k^{a_2} e^{\theta(a_2 - k(n-1))} \]

and similarly
\[ \sum_{a_1+a_2=0} \binom{n}{a_1 a_2} (1 - k)^{a_1} k^{a_2} e^{\theta(a_2 - k(n-1))} (-1)^1 = \]
\[ - \sum_{a_2=0}^{n-1} \binom{n}{a_2} (1 - k)^{n-a_2} k^{a_2} e^{\theta(a_2 - k(n-1))} \]

while in general
\[ \sum_{a_1+a_2=0} \binom{n}{a_1 a_2} (1 - k)^{a_1} k^{a_2} e^{\theta(a_2 - k(n-j))} (-1)^j = \]
\[ \sum_{a_2=0}^j \binom{n}{a_2} (1 - k)^{n-a_2} k^{a_2} e^{\theta(a_2 - k(n-j))} (-1)^j. \]

Finally, dropping the subscript for \( a_2 \) we get
\[ b^n = \sum_{j=0}^n \sum_{a=0}^j \binom{n}{a} (1 - k)^{n-a} k^a e^{\theta(a-k(n-j))} (-1)^j \]

which after substitution in the mgf gives us the following
\[ M_{H_n}(t) = [\sum_{j=0}^n \sum_{a=0}^j \binom{n}{a} (1 - k)^{n-a} k^a e^{\theta(a-k(n-j))} (-1)^j] \sum_{i=1}^n \prod_{p=1}^i p \]
\[
+ \left( \sum_{j=0}^{n-1} \sum_{a=0}^{n-1-j} \binom{n-1-j}{a+j} (1 - k)^{n-1-a-j} k^a \gamma^j \right) \sum_{j=1}^{n} \prod_{i=1}^{n-1} P_i \\
+ \left( \sum_{j=0}^{n-2} \sum_{a=0}^{n-2-j} \binom{n-2-j}{a+j} (1 - k)^{n-2-a-j} k^a \gamma^j \right) \sum_{j=1}^{n} \prod_{i=1}^{n-2} P_i \\
\vdots \\
+ \left( \sum_{j=0}^{n-(n-2)} \sum_{a=0}^{n-(n-2)-j} \binom{n-(n-2)-j}{a+j} (1 - k)^{(2-a-j)} k^a \gamma^j \right) \sum_{j=1}^{n} \prod_{i=1}^{n-(n-2)} P_i \\
+ \left( (1 - k)e^{-it} + ke^{(1-k)t} - 1 \right) \sum_{j=1}^{n} P_i \\
+ 1.
\]

Therefore,

\[
M_{H_0}(t) = \left[ \sum_{a=0}^{n} \binom{n}{a} (1 - k)^a \gamma^a \right] \sum_{j=1}^{n} \prod_{i=1}^{n-1} P_i \\
+ \left( \sum_{a=0}^{n-1} \binom{n}{a} (1 - k)^a \gamma^a \right) \sum_{j=1}^{n} \prod_{i=1}^{n-1} P_i \\
+ \left( \sum_{a=0}^{n-2} \binom{n}{a} (1 - k)^a \gamma^a \right) \sum_{j=1}^{n} \prod_{i=1}^{n-2} P_i \\
+ \left( \sum_{a=0}^{n-(n-1)} \binom{n-(n-1)}{a} \gamma^{n-(n-1)} \right) \sum_{j=1}^{n} \prod_{i=1}^{n-(n-1)} P_i \\
+ \left( \sum_{a=0}^{n-(n-2)} \binom{n-(n-2)}{a} \gamma^{n-(n-2)} \right) \sum_{j=1}^{n} \prod_{i=1}^{n-(n-2)} P_i \\
+ \left( (1 - k)e^{-it} + ke^{(1-k)t} - 1 \right) \sum_{j=1}^{n} P_i \\
+ 1.
\]
Thus, if we gather the common terms we get:

\[
M_{H_n}(t) = \left[ \sum_{a=0}^{n} \binom{n}{a} (1 - k)^{a} e^{(a-k)n} (-1)^{a} \sum_{j=1}^{n} \prod_{j} P_{n1} \right] + \sum_{a=0}^{n-1} \sum_{j=0}^{a} \binom{n}{a} \binom{a}{j} \binom{n}{a-j} \binom{n-a}{a-j} \binom{n-a-j}{a-j} \binom{n-a-j}{a-j} (-1)^{a-j} \sum_{j=1}^{n} \prod_{j} P_{n1} \\
+ \sum_{a=0}^{n-2} \sum_{j=0}^{a} \binom{n}{a} \binom{a}{j} \binom{n-a}{a-j} \binom{n-a-j}{a-j} \binom{n-a-j}{a-j} (-1)^{a-j} \sum_{j=1}^{n} \prod_{j} P_{n1} \\
+ \sum_{a=0}^{n-3} \sum_{j=0}^{a} \binom{n}{a} \binom{a}{j} \binom{n-a}{a-j} \binom{n-a-j}{a-j} \binom{n-a-j}{a-j} (-1)^{a-j} \sum_{j=1}^{n} \prod_{j} P_{n1} \\
+ \cdots \\
+ \sum_{a=0}^{n-(n-2)} \sum_{j=0}^{a} \binom{n}{a} \binom{a}{j} \binom{n-a}{a-j} \binom{n-a-j}{a-j} \binom{n-a-j}{a-j} (-1)^{a-j} \sum_{j=1}^{n} \prod_{j} P_{n1} \\
+ \left[ (1 - k)e^{-k} + ke^{(1-k)} - 1 \right] \sum_{j=1}^{n} P_{n1} \\
+ 1
\]

which gives us the form

\[
M_{H_n}(t) = \sum_{a=0}^{n} \binom{n}{a} (1 - k)^{a} e^{(a-k)n} (-1)^{a} \sum_{j=1}^{n} \prod_{j} P_{n1} \\
\cdot \binom{a}{j} \binom{n-a}{a-j} \binom{n-a-j}{a-j} \binom{n-a-j}{a-j} (-1)^{a-j} \sum_{j=1}^{n} \prod_{j} P_{n1} \\
\cdot \binom{n-a-j}{a-j} (-1)^{a-j} \sum_{j=1}^{n} \prod_{j} P_{n1} \\
\cdot \left[ (1 - k)e^{-k} + ke^{(1-k)} - 1 \right] \sum_{j=1}^{n} P_{n1} \\
+ 1
\]

Eventually, then, we can specify the probabilities for values of \( H_n \) in the same fashion that we did for the distribution of a sum of independent non-identically distributed Bernoulli random variables.

In particular,
Note that in enumerating the values for $H_n$ it is conceivable that some may repeat depending on the value of $k$, which will depend of the value of the parameter $\gamma$. We will illustrate this below.

Once we have $P(H_n = a - k(n - j))$, we can look at the exact power of a test of the form $H_0 : \gamma = \gamma_0$ versus $H_1 : \gamma = \gamma_1$. For example, the distribution of $H_n$ and subsequently the test statistic for $H_0 : \gamma = \gamma_0 = 0$ (or $k = \frac{1}{2}$) with $T = n - 2$ can be written as

<table>
<thead>
<tr>
<th>$H_n$</th>
<th>$a$</th>
<th>$j$</th>
<th>$S$</th>
<th>$P(H_n = h_n) = P(S = s)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-1$</td>
<td>$0$</td>
<td>$0$</td>
<td>$-\frac{1}{2}$</td>
<td>$\sum_{m=1}^{n} p_n$</td>
</tr>
<tr>
<td>$-\frac{1}{2}$</td>
<td>$0$</td>
<td>$1$</td>
<td>$-\frac{1}{2}$</td>
<td>$\sum_{m=1}^{n} p_n$</td>
</tr>
<tr>
<td>$0$</td>
<td>$2$</td>
<td>$0$</td>
<td>$0$</td>
<td>$\sum_{m=1}^{n} p_n$</td>
</tr>
<tr>
<td>$\frac{1}{2}$</td>
<td>$1$</td>
<td>$1$</td>
<td>$\frac{1}{2}$</td>
<td>$\sum_{m=1}^{n} p_n$</td>
</tr>
<tr>
<td>$0$</td>
<td>$1$</td>
<td>$0$</td>
<td>$0$</td>
<td>$\sum_{m=1}^{n} p_n$</td>
</tr>
<tr>
<td>$1$</td>
<td>$2$</td>
<td>$0$</td>
<td>$1$</td>
<td>$\sum_{m=1}^{n} p_n$</td>
</tr>
</tbody>
</table>

The probability values were obtained in the following manner. (Note that all that is required is the substitution of the $a$ and $j$ values.)

$$P(H_n = 0 - \frac{1}{2}(2 - 0)) = \sum_{m=0}^{j} (\frac{2}{-j - j} a) l(1 - \frac{1}{2})^{(2 - \sigma j)} \frac{1}{2}^{a} (-1)^{l} \sum_{m=1}^{n} \prod_{2-j+1} p_{21}$$
\[
\begin{align*}
    P(H_n = 0 - \frac{1}{2} (2 - 1)) & = \sum_{j=0}^{1} \binom{2-1}{j} \left( \frac{1}{2} \right)^{(2-1)} (-1)^{j} \sum_{i=1}^{2} \prod_{i=1}^{2} p_{i1} \\
    &= \sum_{j=0}^{1} \binom{2-1}{j} \left( \frac{1}{2} \right)^{(2-1)} (-1)^{j} \sum_{i=1}^{2} \prod_{i=1}^{2} p_{i1} \\
    &= \binom{2-1}{0} \left( \frac{1}{2} \right)^{(2-1)} (-1)^{0} \sum_{i=1}^{2} \prod_{i=1}^{2} p_{i1} \\
    &= \binom{2-1}{1} \left( \frac{1}{2} \right)^{(2-1)} (-1)^{1} \sum_{i=1}^{2} \prod_{i=1}^{2} p_{i1} \\
    &= \left( \frac{1}{2} \right)^{1} \sum_{i=1}^{2} \prod_{i=1}^{2} p_{i1} + 2 \left( \frac{1}{2} \right)^{(1)} (-1)^{1} \sum_{i=1}^{2} \prod_{i=1}^{2} p_{i1} \\
    &= -p_{11} p_{21} + \frac{1}{2} \sum_{i=1}^{2} p_{i1} \\
    P(H_n = 0 - \frac{1}{2} (2 - 2)) & = \sum_{j=0}^{2} \binom{2-2}{j} \left( \frac{1}{2} \right)^{(2-2)} (-1)^{j} \sum_{i=1}^{2} \prod_{i=1}^{2} p_{i1} \\
    &= \binom{2-2}{0} \left( \frac{1}{2} \right)^{(2-2)} (-1)^{0} \sum_{i=1}^{2} \prod_{i=1}^{2} p_{i1} \\
    &= \binom{2-2}{1} \left( \frac{1}{2} \right)^{(2-2)} (-1)^{1} \sum_{i=1}^{2} \prod_{i=1}^{2} p_{i1} \\
    &= \binom{2-2}{2} \left( \frac{1}{2} \right)^{(2-2)} (-1)^{2} \sum_{i=1}^{2} \prod_{i=1}^{2} p_{i1} \\
    &= p_{11} p_{21} - \sum_{i=1}^{2} p_{i1} + 1 \\
    P(H_n = 1 - \frac{1}{2} (2 - 1)) & = + \sum_{j=0}^{1} \binom{2-1}{j} \left( \frac{1}{2} \right)^{(2-1)} (-1)^{j} \sum_{i=1}^{2} \prod_{i=1}^{2} p_{i1} \\
    &= \binom{2-1}{0} \left( \frac{1}{2} \right)^{(2-1)} (-1)^{0} \sum_{i=1}^{2} \prod_{i=1}^{2} p_{i1} \\
    &= \binom{2-1}{1} \left( \frac{1}{2} \right)^{(2-1)} (-1)^{1} \sum_{i=1}^{2} \prod_{i=1}^{2} p_{i1} \\
    &= \left( \frac{1}{2} \right)^{1} \sum_{i=1}^{2} \prod_{i=1}^{2} p_{i1} + 2 \left( \frac{1}{2} \right)^{(1)} (-1)^{1} \sum_{i=1}^{2} \prod_{i=1}^{2} p_{i1} \\
    &= -p_{11} p_{21} + \frac{1}{2} \sum_{i=1}^{2} p_{i1} \\
    \end{align*}
\]
\[
P(H_n = 1 - \frac{1}{2}(2 - 0)) = \\
\sum_{i=0}^{\infty} \binom{2-o-1}{i} \left(\frac{1}{2}\right)^i \sum_{j=1}^{2-o} \prod_{2-o_1} p_{11} = 2! \left(\frac{1}{2}\right)^2 p_{11} p_{21} = \left(\frac{1}{2}\right) p_{11} p_{21}
\]

\[
P(H_n = 2 - \frac{1}{2}(2 - 0)) = \\
\sum_{i=0}^{\infty} \binom{2-o-2}{i} \left(\frac{1}{2}\right)^i \sum_{j=1}^{2-o} \prod_{2-o_1} p_{11} = \left(\frac{1}{2}\right)^2 p_{11} p_{21}
\]

The above table can thus be simplified to

<table>
<thead>
<tr>
<th>$H_n$</th>
<th>$S$</th>
<th>$P(H_n = h_n) = P(S = s)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>-2</td>
<td>( \left(\frac{1}{2}\right)^2 p_{11} p_{21} )</td>
</tr>
<tr>
<td>(\frac{1}{2})</td>
<td>-1</td>
<td>( \frac{1}{2} \sum_{j=1}^{2-o} p_{11} - p_{11} p_{21} )</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>( 1 + \frac{3}{2} p_{11} p_{21} - \sum_{j=1}^{2-o} p_{11} )</td>
</tr>
<tr>
<td>(\frac{1}{2})</td>
<td>1</td>
<td>( \frac{1}{2} \sum_{j=1}^{2-o} p_{11} - p_{11} p_{21} )</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>( \left(\frac{1}{2}\right)^2 p_{11} p_{21} )</td>
</tr>
</tbody>
</table>

A decision (rejection rule) using this distribution that allows for controlling the size of the test can initially be decided upon. Power against the alternative $H_a : \gamma = \gamma_1$ can then be determined by deriving the distribution of $H_n$ and $S$ under $\gamma = \gamma_1$ and applying the rejection rule to compute the appropriate rejection probability. For example if we wish to test $H_0 : \gamma = 0$ against $H_a : \gamma = \gamma_1$
such that \( k_1 = \frac{1}{3} \) then the distribution of \( H_n \) and \( S \) under \( H_a \) would be

\[
\begin{array}{ccc}
H_n & a & j \\
\frac{2}{3} & 0 & 0 \\
\frac{1}{3} & 0 & 1 \\
0 & 0 & 2 \\
\frac{1}{3} & 1 & 0 \\
\frac{2}{3} & 1 & 1 \\
\frac{4}{3} & 2 & 0 \\
\end{array}
\]

\[
P(H_n = h_n) = P(S = s) = \frac{2}{\sqrt{3}} \sum_{i=1}^{n} p_{ni}
\]

The probability values again were obtained in the following manner. (Note that all that is
required is the substitution of the \( a \) and \( j \) values.)

\[
P(H_n = 0 - \frac{1}{3} (2 - 0)) = \sum_{l=0}^{n} \binom{n-j}{l} (1 - \frac{1}{3})^{(n-a-j)} \frac{1}{3} \sum_{i=1}^{n} \prod_{m=1}^{n-j+l} p_{mi}
\]

\[
= \sum_{l=0}^{n} \binom{2-0+l}{l} (\frac{2}{3})^{(2-a-0)} \frac{1}{3} (-1)^l \sum_{i=1}^{2} \prod_{m=1}^{2-0+l} p_{mi}
\]

\[
= (\frac{2-0-0}{3}) (\frac{2}{3})^{(2-0-0)} \frac{1}{3} \sum_{i=1}^{2} \prod_{m=1}^{2-0} p_{mi}
\]

\[
P(H_n = 0 - \frac{1}{3} (2 - 1)) = \sum_{l=0}^{n} \binom{n-j}{l} (1 - \frac{1}{3})^{(n-a-j)} \frac{1}{3} \sum_{i=1}^{n} \prod_{m=1}^{n-j+l} p_{mi}
\]

\[
= \sum_{l=0}^{n} \binom{2-0+l}{l} (\frac{2}{3})^{(2-a-0)} \frac{1}{3} (-1)^l \sum_{i=1}^{2} \prod_{m=1}^{2-0+l} p_{mi}
\]

\[
= (\frac{2-0-0}{3}) (\frac{2}{3})^{(2-0-0)} \frac{1}{3} \sum_{i=1}^{2} \prod_{m=1}^{2-0} p_{mi}
\]

\[
P(H_n = 0 - \frac{1}{3} (2 - 2)) = \sum_{l=0}^{n} \binom{n-j}{l} (1 - \frac{1}{3})^{(n-a-j)} \frac{1}{3} \sum_{i=1}^{n} \prod_{m=1}^{n-j+l} p_{mi}
\]

\[
= \sum_{l=0}^{n} \binom{2-0+l}{l} (\frac{2}{3})^{(2-a-0)} \frac{1}{3} (-1)^l \sum_{i=1}^{2} \prod_{m=1}^{2-0+l} p_{mi}
\]

\[
P(H_n = 0 - \frac{1}{3} (2 - 1)) = \sum_{l=0}^{n} \binom{n-j}{l} (1 - \frac{1}{3})^{(n-a-j)} \frac{1}{3} \sum_{i=1}^{n} \prod_{m=1}^{n-j+l} p_{mi}
\]
\[
\sum_{i=0}^{1} \binom{2-1+i}{2-1-i \cdot i} (\frac{2}{3})^{(2-1-i) \cdot i} \cdot (-1)^{i} \sum_{p=1}^{2} P_{2-1+i} \cdot P_{p_1} \\
= \sum_{i=0}^{1} \binom{2-1+i}{2-0-1 \cdot 0} (\frac{2}{3})^{(2-0-1) \cdot 0} \cdot (-1)^{0} \sum_{p=1}^{2} P_{2-1+i} \cdot P_{p_1} \\
= (\frac{2-1+i}{2-0-1 \cdot 0}) (\frac{2}{3})^{(2-0-1) \cdot 0} \cdot (-1)^{0} \sum_{p=1}^{2} P_{2-1+i} \cdot P_{p_1} \\
+ (\frac{2-1+i}{2-0-1 \cdot 0}) (\frac{2}{3})^{(2-0-1) \cdot 0} \cdot (-1)^{1} \sum_{p=1}^{2} P_{2-1+i} \cdot P_{p_1} \\
= (\frac{2}{3}) \sum_{p=1}^{2} P_{p_1} - 2!(\frac{2}{3}) \cdot P_{11} \cdot P_{21} = (\frac{2}{3}) \sum_{p=1}^{2} P_{p_1} - (\frac{4}{3}) \cdot P_{11} \cdot P_{21} \\
P(H_p = 0 - \frac{1}{3}(2 - 2)) =
\sum_{i=0}^{2} \binom{2-2+i}{2-0-2 \cdot 0} (\frac{2}{3})^{(2-0-2) \cdot 0} \cdot (-1)^{0} \sum_{p=1}^{2} P_{2-2+i} \cdot P_{p_1} \\
= (\frac{2-2+i}{2-0-2 \cdot 0}) (\frac{2}{3})^{(2-0-2) \cdot 0} \cdot (-1)^{0} \sum_{p=1}^{2} P_{2-2+i} \cdot P_{p_1} \\
+ (\frac{2-2+i}{2-0-2 \cdot 0}) (\frac{2}{3})^{(2-0-2) \cdot 0} \cdot (-1)^{1} \sum_{p=1}^{2} P_{2-2+i} \cdot P_{p_1} \\
+ (\frac{2-2+i}{2-0-2 \cdot 0}) (\frac{2}{3})^{(2-0-2) \cdot 0} \cdot (-1)^{2} \sum_{p=1}^{2} P_{2-2+i} \cdot P_{p_1} \\
= 1 - \sum_{p=1}^{2} P_{p_1} + P_{11} \cdot P_{21} \\
P(H_p = 1 - \frac{1}{3}(2 - 1)) =
\sum_{i=0}^{1} \binom{2-1+i}{2-1-i \cdot i} (\frac{2}{3})^{(2-1-i) \cdot i} \cdot (-1)^{i} \sum_{p=1}^{2} P_{2-1+i} \cdot P_{p_1} \\
= (\frac{2-1+i}{2-1-i \cdot i}) (\frac{2}{3})^{(2-1-i) \cdot i} \cdot (-1)^{i} \sum_{p=1}^{2} P_{2-1+i} \cdot P_{p_1} \\
+ (\frac{2-1+i}{2-1-i \cdot i}) (\frac{2}{3})^{(2-1-i) \cdot i} \cdot (-1)^{i} \sum_{p=1}^{2} P_{2-1+i} \cdot P_{p_1} \\
= \frac{1}{3} \sum_{p=1}^{2} P_{p_1} - 2! \frac{1}{3} \cdot P_{11} \cdot P_{21} = \frac{1}{3} \sum_{p=1}^{2} P_{p_1} - \frac{2}{3} \cdot P_{11} \cdot P_{21} \\
P(H_p = 1 - \frac{1}{3}(2 - 0)) =
\sum_{i=0}^{0} \binom{2-0+i}{2-0-i \cdot i} (\frac{2}{3})^{(2-0-i) \cdot i} \cdot (-1)^{i} \sum_{p=1}^{2} P_{2-0+i} \cdot P_{p_1}
\[
\begin{align*}
&= \left(\frac{2^0}{2-1-0} \right) \left(\frac{2}{3}\right)^{\frac{1}{3}} (-1)^0 \sum_{j=0}^{2} \prod_{2-0} p_{11} \\
&= 2 \left(\frac{2}{3}\right)^{\frac{1}{3}} p_{11} p_{21} = 4 \left(\frac{1}{3}\right)^2 p_{11} p_{21} = \left(\frac{2}{3}\right)^2 p_{11} p_{21} \\
\end{align*}
\]

\[
P(H_n = 2 - \frac{1}{3} (2 - 0)) = \\
\sum_{i=0}^{2-0} c_{1,0,0} \left(\frac{2}{3}\right)^{\frac{1}{3}} (-1)^i \sum_{j=0}^{2} \prod_{2-0} p_{11} \\
= \left(\frac{2^0}{2-2-0} \right) \left(\frac{2}{3}\right)^{\frac{1}{3}} \left(\frac{1}{3}\right) (-1)^0 \sum_{i=0}^{2-0} \prod_{2-0} p_{11} \\
= \left(\frac{1}{3}\right)^2 p_{11} p_{21}
\]
7.3 Methodology for the Derivation of the Distribution of the Rao Score Statistic: Three Time Periods

As an extension to the results derived in section 7.1 and 7.2, we now consider an analogous situation where \( T = 3 \), and attempt to derive the exact distribution for the \( \sum_{i=1}^{n} A_i \) and the test statistic given by \( S \) in Chapter 4. We show that problems arise. Specifically, for the case of \( T = 3 \), we now have that

\[
A_i = Y_{i2}(Y_i - \lambda_{i3}) + Y_{i3}(Y_i - \lambda_{i2}).
\]

Letting

\[
P(Y_{it} = 1 | y_{t-1}) = e^{\gamma_{t}y_{t-1}}/(1 + e^{\gamma_{t}y_{t-1}}) = \lambda_{it},\text{ and bearing in mind that}
\]

\[
P(y_{i1}, y_{i2}, y_{i3}) = P(y_{i3} | y_{i2})P(y_{i2} | y_{i1})P(y_{i1})
\]

\[
= \lambda_{i3}^{y_{i3}} (1 - \lambda_{i3})^{1-y_{i3}} \lambda_{i2}^{y_{i2}} (1 - \lambda_{i2})^{1-y_{i2}} \lambda_{i1}^{y_{i1}} (1 - \lambda_{i1})^{1-y_{i1}}
\]

\[
= \lambda_{i3}^{y_{i3}} (1 - \lambda_{i3})^{1-y_{i3}} \lambda_{i2}^{y_{i2}} (1 - \lambda_{i2})^{1-y_{i2}} \lambda_{i1}^{y_{i1}} (1 - \lambda_{i1})^{1-y_{i1}}
\]
we can get the distribution of a single $A_i$:

<table>
<thead>
<tr>
<th>$Y_{i1}$</th>
<th>$Y_{i2}$</th>
<th>$Y_{i3}$</th>
<th>$P(y_{i1}, y_{i2}, y_{i3})$</th>
<th>$A_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>$\frac{1}{4}(1-p_n)$</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>$\frac{1}{4}(1-p_n)$</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>$(1-k)\frac{1}{2}(1-p_n)$</td>
<td>$-k$</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>$k\frac{1}{2}(1-p_n)$</td>
<td>$1-k$</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>$(1-k)\frac{1}{2}p_n$</td>
<td>$-k$</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>$(1-k)\frac{1}{2}p_n$</td>
<td>$-k$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>$(1-k)kp_n$</td>
<td>$1-2k$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>$k^2p_n$</td>
<td>$2-2k$</td>
</tr>
</tbody>
</table>

which simplifies to

<table>
<thead>
<tr>
<th>$A_i$</th>
<th>$P(a_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-k$</td>
<td>$\frac{1}{2}(1-k)(1+p_n)$</td>
</tr>
<tr>
<td>0</td>
<td>$\frac{1}{2}(1-p_n)$</td>
</tr>
<tr>
<td>$1-k$</td>
<td>$\frac{1}{2}k(1-p_n)$</td>
</tr>
<tr>
<td>$1-2k$</td>
<td>$(1-k)kp_n$</td>
</tr>
<tr>
<td>$2-2k$</td>
<td>$k^2p_n$</td>
</tr>
</tbody>
</table>
Then from here we can derive the mgf for a given $A_i$

\[
M_{A_i}(t) = E[e^{tA_i}] = e^{-tk} \left(1 - k\right)(1 + p_n) + e^{\alpha_1} \frac{1}{2} (1 - p_n)
+ e^{\alpha(t-k) \frac{1}{2} k(1 - p_n)} + e^{\alpha(1-2k)} (1 - k)kp_n + e^{\alpha(2-2k)k^2 p_n}
\]

\[
= \frac{1}{2} e^{-tk}(1 - k)(1 + p_n) + \frac{1}{2} (1 - p_n) + \frac{1}{2} e^{\alpha(t-k)k(1 - p_n)} + e^{\alpha(1-2k)} (1 - k)kp_n
+ e^{\alpha(2-2k)k^2 p_n} = p_n \left[ \frac{1}{2} e^{-tk}(1 - k) - \frac{1}{2} - \frac{1}{2} e^{\alpha(t-k)k} + e^{\alpha(1-2k)} (1 - k)k + e^{\alpha(2-2k)k^2} \right]
\]

\[
+ \frac{1}{2} e^{-tk}(1 - k) + \frac{1}{2} + \frac{1}{2} e^{\alpha(t-k)k} = \frac{1}{2} A + Bp_n.
\]

where

\[
A = 1 + e^{-tk}(1 - k) + e^{\alpha(t-k)k}
\]

and

\[
B = \frac{1}{2} e^{-tk}(1 - k) - \frac{1}{2} - \frac{1}{2} e^{\alpha(t-k)k} + e^{\alpha(1-2k)} (1 - k)k + e^{\alpha(2-2k)k^2}
\]

and so

\[
M_{\sum_{i=1}^{n} A_i}(t) = \prod_{i=1}^{n} M_{A_i}(t) = \prod_{i=1}^{n} (\frac{1}{2} A + Bp_n) = \prod_{i=1}^{n} \frac{1}{2} A[1 + \frac{2B}{A} p_n]
\]

\[
= \left( \frac{1}{2} A \right)^n \prod_{i=1}^{n} \left[ 1 + \frac{2B}{A} p_n \right] = \left( \frac{1}{2} A \right)^n \left[ 1 + \frac{2B}{A} \sum_{i=1}^{n} p_i + \left( \frac{2B}{A} \right)^2 \sum_{i=1}^{n} p_i \right]
+ \left( \frac{2B}{A} \right)^3 \sum_{i=1}^{n} p_i + ... + \left( \frac{2B}{A} \right)^{n-1} \sum_{i=1}^{n} p_i + \left( \frac{2B}{A} \right)^n \sum_{i=1}^{n} p_i
\]

\[
= \left( \frac{1}{2} A \right)^n + \left( \frac{1}{2} A \right)^{n-1} B \sum_{i=1}^{n} p_i + \left( \frac{1}{2} A \right)^{n-2} B^2 \sum_{i=1}^{n} p_i
+ \left( \frac{1}{2} A \right)^{n-3} B^3 \sum_{i=1}^{n} p_i + ... + \left( \frac{1}{2} A \right)^n B^{n-1} \sum_{i=1}^{n} p_i
+ \left( \frac{1}{2} A \right)^n B^n \sum_{i=1}^{n} p_i
\]

We must now use the multinomial expansion twice, firstly for $A^m$ where

\[
A^m = \left[ 1 + e^{-tk}(1 - k) + e^{\alpha(t-k)k} \right]^m
\]
which requires the use of a trinomial expansion

$$A^m = \sum_{j_1} \sum_{j_2} \sum_{j_3} \binom{m}{j_1 j_2 j_3} \left( e^{-k} (1 - k) \right)^{j_1} \left( e^{(1-k)k} \right)^{j_2} (1)^{j_3}$$

$$= \sum_{j_1} \sum_{j_2} \sum_{j_3} \binom{m}{j_1 j_2 j_3} \left( 1 - k \right)^{j_1} k^{j_2} e^{(1-k)k(j_1+j_2)} (1)^{j_3}$$

where $\sum_{i=1}^3 j_i = m$. Therefore, we can expand $A^m$ in the same manner as we did in the expansion when finding the distribution of $\sum_{i=1}^T A_i$ when $T = 2$, i.e. write it as a polynomial in $e^t$.

Next we need to apply the multinomial expansion to $B^m$:

$$B^m = \left[ \frac{1}{2} e^{-k} (1 - k) - \frac{1}{2} - \frac{1}{2} e^{(1-k)k} + e^{(1-2k)k} (1 - k)k + e^{(2-2k)k} \right]^w$$

$$= \sum_{j_1} \sum_{j_2} \sum_{j_3} \sum_{j_4} \binom{w}{j_1 j_2 j_3 j_4} \left( \frac{1}{2} e^{-k} (1 - k) \right)^{j_1} \left( e^{(1-2k)k} \right)^{j_2} \left( \frac{1}{2} e^{(1-k)k} \right)^{j_3} \left( - \frac{1}{2} e^{(1-k)k} \right)^{j_4}$$

Unfortunately, this expansion is much more complicated than that for $A^m$ or the multinomial expansion involved in the distribution of $\sum_{i=1}^n A_i$ where $T = 2$. Thus we are strongly motivated towards the use of the asymptotic distribution of $\sum_{i=1}^n A_i$ and $S$ when dealing with larger values of $T$.

However, the expansion of the above expression to a polynomial in $e^t$ will also eventually provide us with a pdf for $\sum_{i=1}^n A_i$ in the same manner in which it was done for the case where $T = 2$. 
8 A Rao Score Test for the Model Containing a Covariate

To date, the results developed have been restricted to the situation where the conditional probability distribution for \( Y_{it} \) is modelled solely as a function of \( Y_{i,t-1} \). Specifically, the model did not include any of the binary responses observed prior to the preceding time period, nor did it account for the possible effect of any covariates. In this section, we consider a model for the conditional probability distribution of the response at time \( t \) that is a function of the previous response at time \( t-1 \) along with a covariate, \( x_{it} \). We write the model as:

\[
P(Y_{it} = 1|y_{i,t-1}) = \log \frac{e^{\beta x_{it} + \gamma y_{i,t-1}}}{1 + e^{\beta x_{it} + \gamma y_{i,t-1}}} = \lambda_{it} \text{ for all } t > 1.
\]

Now since at time period \( t = 1 \) we have no information about the previous time period but do have information about \( x_{i1} \) we will use the following as a probability statement for \( P(Y_{i1} = 1) \):

\[
\lambda_{i1} = P(Y_{i1} = 1) = e^{\beta x_{i1}}/(1 + e^{\beta x_{i1}}) = \mu_{i1},
\]

which is similar to the case for the model that did not include a covariate.

From this point on we can derive a likelihood function in the same manner that we did for the model without covariates. Thus writing \( P(y_{i1}|y_{i,t-1}) \) in the form:

\[
P(Y_{it}|y_{i,t-1}) = \lambda_{it}^{y_{it}}(1 - \lambda_{it})^{1-y_{it}}
\]

leads to the joint density of \( Y_{i1}, \ldots, Y_{it} \):

\[
P(y_{i1}, \ldots, y_{it}) = P(y_{i1}|y_{i,t-1})P(y_{i,t-1}|y_{i,t-2})\ldots P(y_{i3}|y_{i2})P(y_{i2}|y_{i1})P(y_{i1})
\]
which can also be written in the form:

\[ P(y_{iT}, \ldots, y_{i1}) = \prod_{j=1}^{T} \lambda_{it}^{y_{ij}} (1 - \lambda_{it})^{1-y_{ij}} \]

Since we have \( n \) independent random vectors \( \tilde{Y}_1, \ldots, \tilde{Y}_n \) we can easily get the joint likelihood function by taking the product of all \( P(y_{iT}, \ldots, y_{i1}) \) over \( n \) as follows:

\[ L = \prod_{i=1}^{n} P(y_{iT}, \ldots, y_{i1}) = \prod_{i=1}^{n} \left( \prod_{j=1}^{T} \lambda_{it}^{y_{ij}} (1 - \lambda_{it})^{1-y_{ij}} \right) \mu_{it}^{y_{ij}} (1 - \mu_{it})^{1-y_{ij}} \] .

Next, taking the natural logarithm of the likelihood of the data

\[
\ln(L) = \ln(\prod_{i=1}^{n} \left( \prod_{j=1}^{T} \lambda_{it}^{y_{ij}} (1 - \lambda_{it})^{1-y_{ij}} \right) \mu_{it}^{y_{ij}} (1 - \mu_{it})^{1-y_{ij}}) \\
= \ln(\prod_{i=1}^{n} \prod_{j=1}^{T} \lambda_{it}^{y_{ij}} (1 - \lambda_{it})^{1-y_{ij}} + \ln(\prod_{i=1}^{n} \mu_{it}^{y_{ij}} (1 - \mu_{it})^{1-y_{ij}}) \\
= \sum_{i=1}^{n} \ln(\mu_{it}^{y_{ij}} (1 - \mu_{it})^{1-y_{ij}}) + \sum_{i=1}^{n} \sum_{j=1}^{T} \ln(\lambda_{it}^{y_{ij}} (1 - \lambda_{it})^{1-y_{ij}}) \\
= \sum_{i=1}^{n} \ln(\mu_{it}^{y_{ij}} (1 - \mu_{it})^{1-y_{ij}}) + \sum_{i=1}^{n} \sum_{j=1}^{T} y_{ij} \ln(\beta x_{it} + \gamma y_{ij-1}) + (1 - y_{ij}) \ln(1 - \log \beta x_{it} + \gamma y_{ij-1}) .
\]

Now note that:

\[
\ln(\log \beta x_{it} + \gamma y_{ij-1}) = \ln(e^{\beta x_{it} + \gamma y_{ij-1}}) - \ln(1 + e^{\beta x_{it} + \gamma y_{ij-1}}) - \beta x_{it} + \gamma y_{ij-1} - \ln(1 + e^{\beta x_{it} + \gamma y_{ij-1}}) \\
\]

and in addition

\[
1 - \log \beta x_{it} + \gamma y_{ij-1} = \frac{e^{\beta x_{it} + \gamma y_{ij-1}} - e^{-\beta x_{it} - \gamma y_{ij-1}}}{1 + e^{\beta x_{it} + \gamma y_{ij-1}}} = \frac{1}{1 + e^{\beta x_{it} + \gamma y_{ij-1}}} = (1 + e^{\beta x_{it} + \gamma y_{ij-1}})^{-1}
\]
so that
\[
\ln(1 - \log i((\beta_{ix} + \gamma_{iy}) \cdot (1 - \mu_{i})) = \ln((1 + e^{\beta_{ix} + \gamma_{iy} \cdot (1 - \mu_{i})})^{-1}) = -\ln(1 + e^{\beta_{ix} + \gamma_{iy} \cdot (1 - \mu_{i})}).
\]

Therefore
\[
\ln(L) = \sum_{i=1}^{n} \ln[\mu_{i}^{\gamma_{i}}(1 - \mu_{i})^{1 - \gamma_{i}}]
+ \sum_{i=1}^{n} \sum_{t=2}^{T} y_{it}[\beta x_{it} + \gamma y_{it} - \ln(1 + e^{\beta_{ix} + \gamma_{iy} \cdot (1 - \mu_{i})})] + (1 - y_{it})[-\ln(1 + e^{\beta_{ix} + \gamma_{iy} \cdot (1 - \mu_{i})})]
= \sum_{i=1}^{n} \ln[\mu_{i}^{\gamma_{i}}(1 - \mu_{i})^{1 - \gamma_{i}}]
+ \sum_{i=1}^{n} \sum_{t=2}^{T} y_{it}[\beta x_{it} + \gamma y_{it} - \ln(1 + e^{\beta_{ix} + \gamma_{iy} \cdot (1 - \mu_{i})})].
\]

Since the model in this case has two parameters of interest, \( \gamma \) and \( \beta \), we need to differentiate \( \ln(L) \) once with respect to parameter \( \gamma \), and once with respect to \( \beta \) in order to obtain a score vector:
\[
\frac{\partial \ln(L)}{\partial \gamma} = \frac{\partial}{\partial \gamma} \left[ \sum_{i=1}^{n} \ln[\mu_{i}^{\gamma_{i}}(1 - \mu_{i})^{1 - \gamma_{i}}] \right]
+ \frac{\partial}{\partial \gamma} \left[ \sum_{i=1}^{n} \sum_{t=2}^{T} y_{it}[\beta x_{it} + \gamma y_{it} - \ln(1 + e^{\beta_{ix} + \gamma_{iy} \cdot (1 - \mu_{i})})] \right]
= \left[ \sum_{i=1}^{n} \sum_{t=2}^{T} \frac{\partial}{\partial \gamma} y_{it}[\beta x_{it} + \gamma y_{it} - \ln(1 + e^{\beta_{ix} + \gamma_{iy} \cdot (1 - \mu_{i})})] \right]
\]
where
\[
\frac{\partial}{\partial \gamma} \ln(1 + e^{\beta_{ix} + \gamma_{iy} \cdot (1 - \mu_{i})}) = y_{i, t-1} \log i((\beta_{ix} + \gamma_{iy} \cdot (1 - \mu_{i})) = y_{i, t-1} \lambda_{it}
\]
and so
\[
\alpha(\gamma) = \frac{\partial \ln(L)}{\partial \gamma} = \sum_{i=1}^{n} \sum_{t=2}^{T} y_{it} y_{i, t-1} - Y_{i, t-1} \lambda_{it} = \sum_{i=1}^{n} \sum_{t=2}^{T} y_{i, t-1}(Y_{it} - \lambda_{it})
\]
Similarly

\[ a(\beta) = \frac{\partial \ln(L)}{\partial \beta} = \sum_{t=1}^{n} \sum_{t'=2}^{T} \left[ \frac{\partial}{\partial \beta}^{T} Y_{it \beta} + \gamma Y_{t+t-1} \right] - \frac{\partial}{\partial \beta} \ln(1 + e^{\beta \epsilon(t \beta \gamma)}) \]

\[ + \frac{\partial}{\partial \beta} \left( \sum_{t=1}^{n} \ln[\mu_{it} (1 - \mu_{ii})] Y_{it} \right) \]

\[ = \sum_{t=1}^{n} \sum_{t'=2}^{T} (Y_{it} x_{it} - x_{it} \lambda_{it}) + \sum_{t=1}^{n} (Y_{it} x_{it} - x_{it} \mu_{ii}) \]

\[ = \sum_{t=1}^{n} \sum_{t'=1}^{T} x_{it} (Y_{it} - \lambda_{it}) \cdot \]

Therefore, we get a score vector \( \check{a} \):

\[ \check{a} = \left( \frac{\partial \ln(L)}{\partial \gamma}, \frac{\partial \ln(L)}{\partial \beta} \right) = (a(\gamma), a(\beta)) \]

\[ = (\sum_{t=1}^{n} \sum_{t'=1}^{T} Y_{it-1} (Y_{it} - \lambda_{it}), \sum_{t=1}^{n} \sum_{t'=2}^{T} x_{it} (Y_{it} - \lambda_{it})) \cdot \]

To determine the information matrix

\[ I = \begin{bmatrix} E\left( \frac{\partial^{2} \ln(L)}{\partial \beta^{2}} \right) & E\left( \frac{\partial^{2} \ln(L)}{\partial \beta \partial \gamma} \right) \\ E\left( \frac{\partial^{2} \ln(L)}{\partial \gamma \partial \beta} \right) & E\left( \frac{\partial^{2} \ln(L)}{\partial \gamma^{2}} \right) \end{bmatrix} \]

we require second derivatives. Specifically

\[ \frac{\partial^{2} \ln(L)}{\partial \beta^{2}} = \frac{\partial}{\partial \beta} \left( \frac{\partial \ln(L)}{\partial \beta} \right) = \frac{\partial}{\partial \beta} \left( \frac{\partial \ln(L)}{\partial \beta} \right) = \frac{\partial}{\partial \beta} \left( \sum_{t=1}^{n} \sum_{t'=2}^{T} (Y_{it} x_{it} - x_{it} \lambda_{it}) \right) \]

\[ = \sum_{t=1}^{n} \sum_{t'=1}^{T} \left( \frac{\partial}{\partial \beta} Y_{it} x_{it} - \frac{\partial}{\partial \beta} x_{it} \lambda_{it} \right) = - \sum_{t=1}^{n} \sum_{t'=2}^{T} x_{it} \frac{\partial}{\partial \beta} \lambda_{it} \]
in which
\[
\frac{\partial}{\partial \beta} \lambda_{it} = \frac{\partial}{\partial \beta} \left( \frac{e^{\beta_1 x_{it} + \beta_2 y_{it}^2}}{1 + e^{\beta_1 x_{it} + \beta_2 y_{it}^2}} \right) = \frac{\frac{\partial}{\partial \beta} \left(e^{\beta_1 x_{it} + \beta_2 y_{it}^2}\right) (1 + e^{\beta_1 x_{it} + \beta_2 y_{it}^2}) - \frac{\partial}{\partial \beta} (1 + e^{\beta_1 x_{it} + \beta_2 y_{it}^2}) \left(e^{\beta_1 x_{it} + \beta_2 y_{it}^2}\right)}{(1 + e^{\beta_1 x_{it} + \beta_2 y_{it}^2})^2}
\]

\[
= x_{it} e^{\beta_1 x_{it} + \beta_2 y_{it}^2} (1 - e^{\beta_1 x_{it} + \beta_2 y_{it}^2})/(1 + e^{\beta_1 x_{it} + \beta_2 y_{it}^2})^2 = x_{it} e^{\beta_1 x_{it} + \beta_2 y_{it}^2} (1 - e^{\beta_1 x_{it} + \beta_2 y_{it}^2})/(1 + e^{\beta_1 x_{it} + \beta_2 y_{it}^2})
\]

and so we get
\[
\frac{\partial^2 \ln(L)}{\partial \beta^2} = -\sum_{j=1}^n \sum_{i=1}^T x_{ij}^2 \lambda_{it}(1 - \lambda_{it}) = -\sum_{j=1}^n \sum_{i=1}^T x_{ij}^2 \lambda_{it}(1 - \lambda_{it})
\]

Similarly, since
\[
\frac{\partial}{\partial \gamma} \lambda_{it} = y_{it-1} \lambda_{it}(1 - \lambda_{it})
\]

we have
\[
\frac{\partial^2 \ln(L)}{\partial \gamma^2} = -\sum_{i=1}^n \sum_{t=2}^T y_{it-1}^2 \lambda_{it}(1 - \lambda_{it})
\]

and
\[
\frac{\partial^2 \ln(L)}{\partial \gamma \partial \beta} = \frac{\partial^2 \ln(L)}{\partial \beta \partial \gamma} = \frac{\partial}{\partial \beta} \left( \frac{\partial \ln(L)}{\partial \gamma} \right) = \frac{\partial}{\partial \beta} \left( \sum_{i=1}^n \sum_{t=2}^T y_{it} y_{it-1} - y_{it-1} \lambda_{it} \right)
\]

\[
= \sum_{i=1}^n \sum_{t=2}^T \frac{\partial}{\partial \beta} \left( y_{it} y_{it-1} - y_{it-1} \lambda_{it} \right) = -\sum_{i=1}^n \sum_{t=2}^T x_{it} y_{it-1} \lambda_{it}(1 - \lambda_{it}).
\]

Next, we derive the equations for the expectations of these second derivatives. For example,
\[
-\mathbb{E} \left( \frac{\partial^2 \ln(L)}{\partial \beta^2} \right) = \mathbb{E} \left[ \sum_{i=1}^n \sum_{t=1}^T x_{it}^2 \lambda_{it}(1 - \lambda_{it}) \right] = \sum_{i=1}^n \sum_{t=1}^T x_{it}^2 \mathbb{E} \left[ \lambda_{it}(1 - \lambda_{it}) \right]
\]
where
\[ E[\lambda_{it}(1 - \lambda_{it})] = \sum_{j=0}^{1} P(Y_{it} = 1|y_{i,t-1})[1 - P(Y_{it} = 1|y_{i,t-1})]P(Y_{i,t-1} = y_{i,t-1}) \]
\[ = P(Y_{it} = 1|0)[1 - P(Y_{it} = 1|0)]P(Y_{i,t-1} = 0) \]
\[ + P(Y_{it} = 1|1)[1 - P(Y_{it} = 1|1)]P(Y_{i,t-1} = 1). \]

Now let
\[ k_{ij} = P(Y_{it} = 1|y_{i,t-1} = j) \]
so that
\[ E[\lambda_{it}(1 - \lambda_{it})] = \sum_{j=0}^{1} k_{ij}(1 - k_{ij})P(Y_{i,t-1} = j). \]

Note that \( \lambda_{i1} = \mu_{i1} \) is not a function of the binary response and can therefore be treated as constant.

Therefore, substituting gives the expectation
\[ -E\left(\frac{\partial^2 \ln L}{\partial \mu^2}\right) = \sum_{j=1}^{n} x_i^2 \mu_{i1} (1 - \mu_{i1}) + \sum_{j=1}^{n} \sum_{t=2}^{T} x_i^2 \sum_{j=0}^{1} k_{ij}(1 - k_{ij})P(Y_{i,t-1} = j) \]

Similarly, we next consider
\[ -E\left(\frac{\partial^2 \ln L}{\partial \tau^2}\right) = E(\sum_{j=1}^{n} \sum_{t=2}^{T} Y_{i,t-1}^2 \lambda_{it}(1 - \lambda_{it})) = \sum_{j=1}^{n} \sum_{t=2}^{T} E[Y_{i,t-1}^2 \lambda_{it}(1 - \lambda_{it})]. \]

Note that since \( Y_{i,t-1} \) is binary we get
\[ E[Y_{i,t-1}^2 \lambda_{it}(1 - \lambda_{it})] = E[Y_{i,t-1} \lambda_{it}(1 - \lambda_{it})] \]
\[ = \sum_{y_{i,t-1}=0}^{1} Y_{i,t-1} \lambda_{it}(1 - \lambda_{it})(1 - \lambda_{it})P(Y_{i,t-1} = y_{i,t-1}) = \lambda_{it}(1 - \lambda_{it}) \mu_{i,t-1} \]

and so
\[ -E\left(\frac{\partial^2 \ln L}{\partial \tau^2}\right) = \sum_{j=1}^{n} \sum_{t=2}^{T} k_{it}(1 - k_{it})\mu_{i,t-1}. \]

Finally
Similarly to the case when we considered an analogous model without any covariates, suppose interest still lies in assessing the extent and significance of the association between binary responses.

One form of test that can be proposed is

\[ H_0 : \gamma = \gamma_0 , \beta = \beta_0 \] versus \[ H_a : \gamma = \gamma_1 , \beta = \beta_1. \]

The score vector \( \mathbf{d} \) and the information matrix \( I \) as defined above, when combined as \( \mathbf{d}' I^{-1} \mathbf{d} \), has the following asymptotic property under \( H_0 \)

\[
\mathbf{d}' I^{-1} \mathbf{d} \rightarrow^d \chi^2_2
\]

That is

\[
(a(\gamma), a(\beta)) \frac{1}{\text{det}(I)} \begin{bmatrix}
-E\left( \frac{\partial^2 \ln(L)}{\partial \gamma^2} \right) & E\left( \frac{\partial^2 \ln(L)}{\partial \gamma \partial \beta} \right) \\
E\left( \frac{\partial^2 \ln(L)}{\partial \gamma \partial \beta} \right) & -E\left( \frac{\partial^2 \ln(L)}{\partial \beta^2} \right)
\end{bmatrix}
(a(\gamma), a(\beta)) \rightarrow^d \chi^2_2
\]

since

\[
I^{-1} = \frac{1}{\text{det}(I)} \begin{bmatrix}
-E\left( \frac{\partial^2 \ln(L)}{\partial \gamma^2} \right) & E\left( \frac{\partial^2 \ln(L)}{\partial \gamma \partial \beta} \right) \\
E\left( \frac{\partial^2 \ln(L)}{\partial \gamma \partial \beta} \right) & -E\left( \frac{\partial^2 \ln(L)}{\partial \beta^2} \right)
\end{bmatrix}
\]

and so
Therefore, we get a rejection rule of the form

\[
\chi^2 \sim \text{chisquare}(2). 
\]

\text{Reject \(H_0\) if } \frac{1}{\det(I)} [2a(\gamma)E(\frac{\partial^2 \ln(L)}{\partial \beta^2}) - a(\gamma)^2 E(\frac{\partial^2 \ln(L)}{\partial \beta^2}) - a(\beta)^2 E(\frac{\partial^2 \ln(L)}{\partial \gamma^2})] \rightarrow \chi^2_2. 

\text{(under } H_0)
The derivation of the above power function however is a bit more cumbersome in comparison to the model containing only the parameter $\gamma$ since the Fisher information is now 2x2. Alternatively, one might consider a test solely on $\gamma$ of the form:

$$H_0 : \gamma = \gamma_0 \quad \text{versus} \quad H_a : \gamma = \gamma_1.$$ 

If $\beta$ is considered to be a known constant then the derivation of a test statistic will follow in a similar manner to the way in which it was done for the model that did not contain any covariates. Specifically, we can get the fisher information corresponding to the parameter $\gamma$ when $\beta$ is considered to be a known constant directly from the 2x2 Fisher information matrix obtained above:

$$I(\gamma) = -E\left(\frac{\partial^2 \ln(L)}{\partial \gamma^2}\right) = \sum_{i=1}^n \sum_{t=2}^T k_{it} (1 - k_{it}) \mu_{i,t-1}.$$ 

Similarly the score corresponding to $\gamma$ can be obtained directly from the score vector. Specifically

$$a(\gamma) = \frac{\partial \ln(L)}{\partial \gamma} = \sum_{i=1}^n \sum_{t=2}^T Y_{i,t-1} (Y_{it} - \lambda_{it}).$$

Thus the asymptotic distribution of $\sqrt{a I^{-1} a}$ is given by

$$a(\gamma) I(\gamma)^{-1} a(\gamma) \rightarrow^d \chi_1^2$$

and a rejection rule of the form

Reject $H_0$ if $a(\gamma) I(\gamma)^{-1} a(\gamma) > c$

can be applied. In the case where $\beta$ is unknown we conduct the test approximately by first using an estimation method such as maximum likelihood to get an estimate $\hat{\beta}$ and then substituting $\hat{\beta}$ wherever $\beta$ appears in the test using the 2x2 matrix.
9 Conclusion and Discussion

In this paper an investigation of several different tests of hypothesis for the association parameter in a model proposed by Farrell and Sutradhar (2006) for longitudinally cluster correlated binary data where \( P(Y_{it} = 1|y_{i,t-1}, \ldots, y_{it}) = \logit(x_{it}^T \beta + \gamma_1 y_{i,t-1} + \ldots + \gamma_{t-1} y_{it}) \) have been suggested.

Specifically, focussing initially on a "lag1" model with no covariates given by

\[ P(Y_{it} = 1|y_{i,t-1}) = e^{\gamma y_{i,t-1}}/(1 + e^{\gamma y_{i,t-1}}), \]

a Rao-Score test for \( \gamma \) was derived for the case where \( T \) is permitted to take any positive integer value larger than two. Power functions were derived for the same test for the cases where \( T = 2 \) and \( T = 3 \) under the condition that the asymptotic distribution of the test statistic is considered. A small example that considered the case \( T = 2 \) demonstrated good asymptotic power when the absolute difference between the hypothesized values for the association parameter under the null and alternative hypothesis was at least 0.75.

Bear in mind, however, that the use of asymptotics necessitates the selection of a large sample, which could lead to a sharp increase in the cost of a study. In addition, in some cases the actual selection of such a sample size may not be possible due to technical reasons. As a result, we chose to derive the real distribution of the test statistic. For simplicity and comprehension, in Chapter 4 we initially considered the case where \( T = 2 \) and demonstrated how the distribution of such a test statistic could be used to construct a test of hypothesis with respect to \( \gamma \). The methodology for the derivation of the distribution of the test statistic for the case where \( T = 3 \) was demonstrated as well,
and a discussion as to a methodology for a derivation of the real distribution of the test statistic for a general value of \( T \) was also given.

We also considered the model \( P(Y_{it} = 1|y_{i,t-1}) = \log \text{it}(\beta x_{it} + \gamma y_{i,t-1}) \) containing a single covariate, developing a Rao-Score test for \( \gamma \). Although a power function for the test under this model was not derived, we recommend considering the asymptotic multivariate normal distribution of the score vector when setting up the probability statement for the power function before considering the \( \chi^2 \) distribution of the statistic which was derived in this paper.

So far as future research in concerned, the possibility of testing statistical hypothesis could be considered for models of the form

\[
P(Y_{it} = 1|y_{i,t-1}, y_{i,t-2}, \ldots, y_{i,2}, y_{i,1}) = \log \text{it}(\mathbf{x}_{it}^T \hat{\beta} + \gamma_1 y_{i,t-1} + \gamma_2 y_{i,t-2} \ldots + \gamma_2 y_{i,2} + \gamma_{t-1} y_{i,1})
\]

where the conditional probability of a binary response at time \( t \) is modeled given all previous responses. Another possibility for further research would be considering the vectors \( \mathbf{Y}_{1,n}, \mathbf{Y}_{2,n} \) where the binary responses within each vector are time dependent (as has been done in this paper), however the vectors will not be independent but will have a certain dependency structure which will need to be modeled and eventually tested for. For example, certain clusters may be correlated along spatial lines, possibly due to some factor such as geographic proximity. Thus, the models considered here for independent clusters could be extended to account for correlation such as this.
10 References

Agresti Alan, 2002 Categorical data analysis pp 409 and pp 418 (New York: John Wiley & sons)


11 Appendix

11.1 Minitab Code to Calculate Power of the Test Based on $\sum_{i=1}^n Y_{i2} Y_{i2}$ for the Case of $T = 2$, $n = 25$, $\gamma_0 = 0$

MTB > let k1=25 (k1 is the sample size n)

MTB > let k2=0.5 (k2 is $P_{Y_2}(Y_{i2} = 1|Y_{i1} = 1)$ under $H_0$)

MTB > let k3=exp(0.25)/(1+exp(0.25)) (k3 is $P_{Y_1}(Y_{i2} = 1|Y_{i1} = 1)$ under $H_a$)

MTB > let k4=k2*k2

MTB > let k5=k3*k3

MTB > let k6=sqrt(1-k4)

MTB > let k7=sqrt(1-k5)

MTB > let k8=sqrt(k1*k5*(1-k5))

MTB > let k9=(k1*(k4-k5)/k8)-1.96*(k2/k3)*(k6/k7)

MTB > let k10=(k1*(k4-k5)/k8)+1.96*(k2/k3)*(k6/k7)

MTB > CDF k9 k11;

SUBC> Normal 0.0 1.0.

MTB > CDF k10 k12;

SUBC> Normal 0.0 1.0.

MTB > let k14=1-k12

MTB > let k15=k14+k14 (computes power)

MTB > prin k15

98
11.2 Macro for the Power of the Rao Score Test when T=2

let k1=25  (sample size)

let k2=0.5  ($k_0 = P(Y_2 = 1|Y_1 = 1)$ under $H_0 : \gamma = 0$)

let k3=exp(0.25)/(1+exp(0.25))  ($k_1 = P(Y_2 = 1|Y_1 = 1)$ under $H_1 : \gamma = \gamma_1$)

let k4 = 1.96*sqrt[(k2*(1-k2)*k3]-sqrt(k1)*(k3-k2)*k3

let k5 = -1.96*sqrt[(k2*(1-k2)*k3]-sqrt(k1)*(k3-k2)*k3

let k6 =sqrt[(k3*(k3-2*k3*k2+k2*k2)-k3*k3*(k3-k2)*(k3-k2)]

let k7 = k4/k6

let k8 = k5/k6

CDF k7 k9;

Normal 0.0 1.0.

CDF k8 k10;

Normal 0.0 1.0.

let k11 = 1-k9

let k12 = k11+k10 (The power of the test)

print k12
11.3 Macro for the Correlation for the Lag 1 Model where $T=2$

let $k100=0$  ( Gamma under $H_a$)

MTB > let $k1=\exp(k100)/(1+\exp(k100))$  ($k$ under $H_a$)

MTB > let $k2=k1*[(3/2)*k1-k1*{(1/2)}]$ (covariance under $H_a$)

MTB > let $k3=\sqrt[k1*(1-k1)]$ (Standard deviation for $Y_1$)

MTB > let $k4=\sqrt[k1*(1/2)*k1+(1/2)]$

MTB > let $k5=\sqrt[1-k1*(1/2)*k1-(1/2)]$

MTB > let $k6=k2/(k3*k4*k5)$  (Correlation)

MTB > print $k6$ (prints correlation)