ISSN 1927-7032 (Print)
ISSN 1927-7040 (Online)

International Journal of Statistics and Probability

Vol. 9, No. 2 March 2020

CANADIAN CENTER OF SCIENCE AND EDUCATION

INTERNATIONAL JOURNAL OF STATISTICS AND PROBABILITY

An International Peer-reviewed and Open Access Journal for Statistics and Probability

International Journal of Statistics and Probability (ISSN: 1927-7032; E-ISSN: 1927-7040) is an open-access, international, double-blind peer-reviewed journal published by the Canadian Center of Science and Education. This journal, published **bimonthly** (January, March, May, July, September and November) in both **print and online versions**, keeps readers up-to-date with the latest developments in all areas of statistics and probability.

The scopes of the journal:

- Computational statistics
- Design of experiments
- Sample survey
- Statistical modelling
- Statistical theory
- Probability theory

The journal is included in:

- BASE
- Google Scholar
- JournalTOCs
- LOCKSS
- SHERPA/RoMEO
- Ulrich's

Submission Policy

Submission of an article implies that the work described has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the authorities responsible where the work was carried out. However, we accept submissions that have previously appeared on preprint servers (for example: arXiv, bioRxiv, Nature Precedings, Philica, Social Science Research Network, and Vixra); have previously been presented at conferences; or have previously appeared in other "non-journal" venues (for example: blogs or posters). Authors are responsible for updating the archived preprint with the journal reference (including DOI) and a link to the published articles on the appropriate journal website upon publication.

iThenticate Professional Plagiarism Prevention

The publisher and journals have a zero-tolerance plagiarism policy. We check the issue using two methods: a plagiarism prevention tool (iThenticate) and a reviewer check. All submissions will be checked by iThenticate before being sent to reviewers.

creative commons

We insist a rigorous viewpoint on the self-plagiarism. The self-plagiarism is plagiarism, as it fails to contribute to the research and science.

IJSP accepts both Online and Email submission. The online system makes readers to submit and track the status of their manuscripts conveniently. For any questions, please contact ijsp@ccsenet.org.

International Journal of Statistics and Probability; Vol. 9, No. 2; March 2020
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Editorial Team

International Journal of Statistics and Probability; Vol. 9, No. 2; March 2020
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Contents

International Journal of Statistics and Probability; Vol. 9, No. 2; March 2020
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Efficient Estimation of Interval-Valued Symbolic Data Regression Model

Chuanhua Wei[1], Nana Zheng[1] & Ke Tian[1]

[1] School of Science, Minzu University of China, Beijing 100081, China

Correspondence: Chuanhua Wei, School of Science, Minzu University of China, Beijing 100081, P.R.China.
E-mail: chweisd@163.com

Received: January 24, 2020 Accepted: February 10, 2020 Online Published: February 20, 2020

doi:10.5539/ijsp.v9n2p1 URL: https://doi.org/10.5539/ijsp.v9n2p1

Abstract

In the last two decades, regression analysis with interval-valued type data has received more and more attention. For the interval-valued symbolic data regression, the Minmax method and the center and range (CR) method are two widely used popular estimating approaches. In this paper, to improve the estimating efficiency of these two estimating methods, seemingly unrelated regression approach has been applied to take account of the dependence information of two regression models of the Minmax method or the CR method. Finally, real data sets are analysed to examine the performance of our proposed procedure.

Keywords: interval-valued data, minmax method, center and range method, seemingly unrelated regression, generalized least squares estimation

1. Introduction

Interval-valued data as a more general class of data type called symbolic data are observed as ranges instead of single values and frequently appear in some fields, such as finance, engineering, and medicine. In the last two decades, regression analysis with interval-valued type data has received more and more attention. Several approaches have been proposed to estimate the regression models with interval-valued data. The center method of Billard and Diday (2000) uses interval midpoints of both response and the associated explanatory variables to build the regression, and apply the fitted model to the lower and upper bounds of the independent variables to generate predictions respectively. In order to use more interval information than that of the center method, Billard and Diday (2002) proposed a MinMax method, which suggests modelling the lower and upper bounds of the intervals of both response and the associated explanatory variables by two linear regression models independently. Using two different models to predict lower and upper bounds can improve the linear fit and give an intuitive response interpretation. To include the information given by both the centre and the range of an interval on a linear regression model to improve the model prediction performance, Lima Neto and De Carvalho (2008) proposed a center and range (CR) method using two independent models: one for the interval midpoints and another for the semi-length of the interval. Other estimating approaches can be found in Xu (2010), Giordani (2015) and Souza *et al.* (2017). Furthermore, some generalized interval data models have been proposed, examples including additive models of Lim (2016), partially linear models of Wei *et al.* (2015).

It is noted that both the MinMax method and CR method analyze the interval data based on two independent linear regression models, and the assumption of independence between two regression models is not always true in practice. If these two regression models have some relationship, then, to solve this problem, how to combine the two regression models is an interesting topic. As we all know, the seemingly unrelated regression (SUR) introduced by Zellner (1962) is an important tool to analyze multiple equations with correlated disturbances. The SUR specification is expressed as a set of linear regressions where the disturbances in the different equations are correlated. Therefore, to take account of the dependence information of two regression models, we propose a seemingly unrelated regression approach based on the MinMax method or the CR method to modelling interval data.

The rest of this paper is organized as follows. We introduce the MinMax and CRM methods in Section 2. The proposed SUR approach is given in Section 3. Real interval-valued data sets are analyzed in Section 4 to illustrate the performance of the proposed approach. Conclusion is presented in Section 5.

2. Minmax and CR Methods for Linear Models With Interval-Valued Data

Let $\mathbf{E} = \{e_1, e_2, \cdots, e_n\}$ be a set of objects that are described by the p+1 symbolic interval-valued variables $Y, X_1, X_2, \cdots, X_p$. Each example $e_i \in E (i = 1, 2, \cdots, n)$ is represented as an interval quantitative feature vector $\mathbf{z}_i = (\mathbf{x}_i, \mathbf{y}_i)$. $\mathbf{x}_i =$

$(x_{i1}, x_{i2}, \cdots, x_{ip})^{\mathrm{T}}$, where $x_{ij} = [a_{ij}, b_{ij}] \in \mathfrak{I} = \{[a,b] : a, b \in \Re, a \le b\}(j = 1, 2, \cdots, p)$ and $y_i = [y_{Li}, y_{Ui}] \in \mathfrak{I}$ are the ith observed values of X_j and Y, respectively. Let us work with the matrix notation. Denote

$$\mathbf{Y}_L = \begin{bmatrix} y_{L1} \\ y_{L2} \\ \vdots \\ y_{Ln} \end{bmatrix}, \mathbf{Y}_U = \begin{bmatrix} y_{U1} \\ y_{U2} \\ \vdots \\ y_{Un} \end{bmatrix}, \mathbf{X}_L = \begin{bmatrix} 1 & a_{11} & \cdots & a_{1p} \\ 1 & a_{21} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & a_{n1} & \cdots & a_{np} \end{bmatrix}, \mathbf{X}_U = \begin{bmatrix} 1 & b_{11} & \cdots & b_{1p} \\ 1 & b_{21} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & b_{n1} & \cdots & b_{np} \end{bmatrix},$$

2.1 MinMax Method

By Billard and Diday (2002), we consider $X_1, X_2, \cdots, X_p$ related to Y according to the following linear regression relationship

$$y_{Li} = \beta_0^L + \beta_1^L a_{i1} + \beta_2^L a_{i2} + \cdots + \beta_p^L a_{ip} + \varepsilon_{Li}, \tag{1}$$

$$y_{Ui} = \beta_0^U + \beta_1^U b_{i1} + \beta_2^U b_{i2} + \cdots + \beta_p^U b_{ip} + \varepsilon_{Ui}. \tag{2}$$

Model (1) (2) can be written in the matrix form as

$$\mathbf{Y}_m = \mathbf{X}_m \boldsymbol{\beta}_m + \boldsymbol{\varepsilon}_m, \tag{3}$$

where

$$\mathbf{Y}_m = \begin{bmatrix} \mathbf{Y}_L \\ \mathbf{Y}_U \end{bmatrix}, \mathbf{X}_m = \begin{bmatrix} \mathbf{X}_L & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_U \end{bmatrix}, \boldsymbol{\beta}_m = \begin{bmatrix} \boldsymbol{\beta}^L \\ \boldsymbol{\beta}^U \end{bmatrix}, \boldsymbol{\varepsilon}_m = \begin{bmatrix} \boldsymbol{\varepsilon}_L \\ \boldsymbol{\varepsilon}_U \end{bmatrix},$$

and $\boldsymbol{\beta}^L = \left(\beta_0^L, \beta_1^L, \cdots, \beta_p^L\right)^{\mathrm{T}}, \boldsymbol{\beta}^U = \left(\beta_0^U, \beta_1^U, \cdots, \beta_p^U\right)^{\mathrm{T}}, \boldsymbol{\varepsilon}_L = (\varepsilon_{L1}, \cdots, \varepsilon_{Ln})^{\mathrm{T}}, \boldsymbol{\varepsilon}_U = (\varepsilon_{U1}, \cdots, \varepsilon_{Un})^{\mathrm{T}}$.

By applying the least squares approach to model (4), we can get the Minmax estimator of $\boldsymbol{\beta}_{\mathrm{m}}$ as

$$\hat{\boldsymbol{\beta}}^{mm} = \begin{bmatrix} \hat{\boldsymbol{\beta}}^L \\ \hat{\boldsymbol{\beta}}^U \end{bmatrix} = \left[\mathbf{X}_m^{\mathrm{T}} \mathbf{X}_m\right]^{-1} \mathbf{X}_m^{\mathrm{T}} \mathbf{Y}_m = \begin{bmatrix} \left[\mathbf{X}_L^{\mathrm{T}} \mathbf{X}_L\right]^{-1} \mathbf{X}_L^{\mathrm{T}} \mathbf{Y}_L \\ \left[\mathbf{X}_U^{\mathrm{T}} \mathbf{X}_U\right]^{-1} \mathbf{X}_U^{\mathrm{T}} \mathbf{Y}_U \end{bmatrix}. \tag{4}$$

Specifically, the MinMax estimators of $\boldsymbol{\beta}_L$ and $\hat{\boldsymbol{\beta}}_U$ are

$$\hat{\boldsymbol{\beta}}^L = \left[\mathbf{X}_L^{\mathrm{T}} \mathbf{X}_L\right]^{-1} \mathbf{X}_L^{\mathrm{T}} \mathbf{Y}_L, \quad \hat{\boldsymbol{\beta}}^U = \left[\mathbf{X}_U^{\mathrm{T}} \mathbf{X}_U\right]^{-1} \mathbf{X}_U^{\mathrm{T}} \mathbf{Y}_U.$$

2.2 The Centre and Range Method

Let $y_i^c = (y_{Li} + y_{Ui})/2, y_i^r = (y_{Ui} - y_{Li})/2$, $x_{ij}^c = (a_{ij} + b_{ij})/2, x_{ij}^r = (b_{ij} - a_{ij})/2$. Lima Neto and De Carvalho (2008) considered the following two regression models

$$y_i^c = \beta_0^c + \beta_1^c x_{i1}^c + \beta_2^c x_{i2}^c + \cdots + \beta_p^c x_{ip}^c + \varepsilon_i^c \tag{5}$$

$$y_i^r = \beta_0^r + \beta_1^r x_{i1}^r + \beta_2^r x_{i2}^r + \cdots + \beta_p^r x_{ip}^r + \varepsilon_i^r \tag{6}$$

Denote

$$\mathbf{Y}^c = \begin{bmatrix} y_1^c \\ y_2^c \\ \vdots \\ y_n^c \end{bmatrix}, \mathbf{Y}^r = \begin{bmatrix} y_1^r \\ y_2^r \\ \vdots \\ y_n^r \end{bmatrix}, \mathbf{X}^c = \begin{bmatrix} 1 & x_{11}^c & \cdots & x_{1p}^c \\ 1 & x_{21}^c & \cdots & x_{2p}^c \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1}^c & \cdots & x_{np}^c \end{bmatrix}, \mathbf{X}^r = \begin{bmatrix} 1 & x_{11}^r & \cdots & x_{1p}^r \\ 1 & x_{21}^r & \cdots & x_{2p}^r \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1}^r & \cdots & x_{np}^r \end{bmatrix},$$

Combing models (5) and (6), we have the following model

$$\mathbf{Y}_{cr} = \mathbf{X}_{cr} \boldsymbol{\beta}_{cr} + \boldsymbol{\varepsilon}_{cr}, \tag{7}$$

where

$$\mathbf{Y}_{cr} = \begin{bmatrix} \mathbf{Y}^c \\ \mathbf{Y}^r \end{bmatrix}, \mathbf{X}_{cr} = \begin{bmatrix} \mathbf{X}^c & \mathbf{0} \\ \mathbf{0} & \mathbf{X}^r \end{bmatrix}, \boldsymbol{\beta}_{cr} = \begin{bmatrix} \boldsymbol{\beta}^c \\ \boldsymbol{\beta}^r \end{bmatrix}, \boldsymbol{\varepsilon}_{cr} = \begin{bmatrix} \boldsymbol{\varepsilon}^c \\ \boldsymbol{\varepsilon}^r \end{bmatrix}.$$

and $\mathbf{Y}^{\mathrm{c}} = \frac{\mathbf{Y}_L + \mathbf{Y}_U}{2}, \mathbf{X}^{\mathrm{c}} = \frac{\mathbf{X}_L + \mathbf{X}_U}{2}, \mathbf{Y}^{\mathrm{r}} = \frac{\mathbf{Y}_L - \mathbf{Y}_U}{2}, \mathbf{X}^{\mathrm{r}} = \frac{\mathbf{X}_L - \mathbf{X}_U}{2}, \boldsymbol{\beta}^c = (\beta_0^c, \beta_1^c, \cdots, \beta_p^c)^{\mathrm{T}}, \boldsymbol{\beta}^r = (\beta_0^r, \beta_1^r, \cdots, \beta_p^r)^{\mathrm{T}}, \boldsymbol{\varepsilon}^c = (\varepsilon_1^c, \varepsilon_2^c, \cdots, \varepsilon_n^c)^{\mathrm{T}}$, and $\boldsymbol{\varepsilon}^r = (\varepsilon_1^r, \varepsilon_2^r, \cdots, \varepsilon_n^r)^{\mathrm{T}}$.

Then, by applying the least squares approach to model (7), we can obtain the CR estimator for $\boldsymbol{\beta}_{cr}$ as

$$\hat{\boldsymbol{\beta}}^{cr} = \begin{bmatrix} \hat{\boldsymbol{\beta}}^{c} \\ \hat{\boldsymbol{\beta}}^{r} \end{bmatrix} = \left[\mathbf{X}_{cr}^{\mathrm{T}}\mathbf{X}_{cr}\right]^{-1}\mathbf{X}_{cr}^{\mathrm{T}}\mathbf{Y}_{cr} = \begin{bmatrix} \left[(\mathbf{X}^{c})^{\mathrm{T}}\mathbf{X}^{c}\right]^{-1}(\mathbf{X}^{c})^{\mathrm{T}}\mathbf{Y}^{c} \\ \left[(\mathbf{X}^{r})^{\mathrm{T}}\mathbf{X}^{r}\right]^{-1}(\mathbf{X}^{r})^{\mathrm{T}}\mathbf{Y}^{r} \end{bmatrix}. \tag{8}$$

Specifically, the Centre-Range estimators of $\boldsymbol{\beta}^{c}$ and $\hat{\boldsymbol{\beta}}^{r}$ can be written as

$$\begin{aligned} \hat{\boldsymbol{\beta}}^{c} &= \left[(\mathbf{X}^{c})^{\mathrm{T}}\mathbf{X}^{c}\right]^{-1}(\mathbf{X}^{c})^{\mathrm{T}}\mathbf{y}^{c} = \left[(\mathbf{X}_{U}+\mathbf{X}_{L})^{\mathrm{T}}(\mathbf{X}_{U}+\mathbf{X}_{L})\right]^{-1}(\mathbf{X}_{U}+\mathbf{X}_{L})^{\mathrm{T}}(\mathbf{Y}_{U}+\mathbf{Y}_{L}) \\ &= \left(\mathbf{X}_{U}^{\mathrm{T}}\mathbf{X}_{U}+\mathbf{X}_{L}^{\mathrm{T}}\mathbf{X}_{L}+\mathbf{X}_{L}^{\mathrm{T}}\mathbf{X}_{U}+\mathbf{X}_{U}^{\mathrm{T}}\mathbf{X}_{L}\right)^{-1}\left(\mathbf{X}_{U}^{\mathrm{T}}\mathbf{Y}_{U}+\mathbf{X}_{L}^{\mathrm{T}}\mathbf{Y}_{L}+\mathbf{X}_{L}^{\mathrm{T}}\mathbf{Y}_{U}+\mathbf{X}_{U}^{\mathrm{T}}\mathbf{Y}_{L}\right), \\ \hat{\boldsymbol{\beta}}^{r} &= \left[(\mathbf{X}^{r})^{\mathrm{T}}\mathbf{X}^{r}\right]^{-1}(\mathbf{X}^{r})^{\mathrm{T}}\mathbf{y}^{r} = \left[(\mathbf{X}_{U}-\mathbf{X}_{L})^{\mathrm{T}}(\mathbf{X}_{U}-\mathbf{X}_{L})\right]^{-1}(\mathbf{X}_{U}-\mathbf{X}_{L})^{\mathrm{T}}(\mathbf{Y}_{U}-\mathbf{Y}_{L}) \\ &= \left(\mathbf{X}_{U}^{\mathrm{T}}\mathbf{X}_{U}+\mathbf{X}_{L}^{\mathrm{T}}\mathbf{X}_{L}-\mathbf{X}_{L}^{\mathrm{T}}\mathbf{X}_{U}-\mathbf{X}_{U}^{\mathrm{T}}\mathbf{X}_{L}\right)^{-1}\left(\mathbf{X}_{U}^{\mathrm{T}}\mathbf{Y}_{U}+\mathbf{X}_{L}^{\mathrm{T}}\mathbf{Y}_{L}-\mathbf{X}_{L}^{\mathrm{T}}\mathbf{Y}_{U}-\mathbf{X}_{U}^{\mathrm{T}}\mathbf{Y}_{L}\right). \end{aligned}$$

3. The Efficient SUR-Based Estimation

For both the Minmax and CR method, we can improve the estimating efficiency by considering the dependence between their two models. We propose the SUR-based Minmax estimator by applying the SUR method to models (1) and (2) directly. Similarly, we propose the SUR-based CR estimator by applying the SUR method to models (5) and (6) directly. It is noted that models (5) and (6) are derived by models (1) and (2), then we construct a two-step SUR-based CR estimator based on the relation between models (1)-(2) and models (5)-(6).

3.1 The Efficient SUR-Based Minmax Estimator

For models (1)-(2), we assume that

$$\mathrm{E}(\varepsilon_{Li}) = 0, \mathrm{D}(\varepsilon_{Li}) = \sigma_{LL}, \mathrm{E}(\varepsilon_{Ui}) = 0, \mathrm{D}(\varepsilon_{Ui}) = \sigma_{UU}, \mathrm{Cov}(\varepsilon_{Li}, \varepsilon_{Lj}) = 0, \mathrm{Cov}(\varepsilon_{Ui}, \varepsilon_{Uj}) = 0, i \neq j.$$

Then, we have

$$\mathrm{Cov}(\boldsymbol{\varepsilon}_{L}) = E\boldsymbol{\varepsilon}_{L}\boldsymbol{\varepsilon}_{L}^{\mathrm{T}} = \sigma_{LL}\mathbf{I}_{n}, \mathrm{Cov}(\boldsymbol{\varepsilon}_{U}) = E\boldsymbol{\varepsilon}_{U}\boldsymbol{\varepsilon}_{U}^{\mathrm{T}} = \sigma_{UU}\mathbf{I}_{n}, \tag{9}$$

Different to Billard and Diday (2002) and Lima Neto and De Carvalho (2008), for the models (1)(2), we consider the following assumptions of their disturbances

$$E(\varepsilon_{Li}\varepsilon_{Uj}) = \begin{cases} \sigma_{LU}, & i = j, \\ 0, & \text{otherwise}, \end{cases}$$

for $1 \leq i, j \leq n$. Then, we have

$$\mathrm{Cov}(\boldsymbol{\varepsilon}_{L}, \boldsymbol{\varepsilon}_{U}) = E\boldsymbol{\varepsilon}_{L}\boldsymbol{\varepsilon}_{U}^{\mathrm{T}} = \sigma_{LU}\mathbf{I}_{n}, \tag{10}$$

with $\mathbf{I}_{n}$ is the identity matrix of order n. Therefore, the $2n \times 1$ disturbance vector $\boldsymbol{\varepsilon}_{\mathrm{m}} = (\boldsymbol{\varepsilon}_{L}^{\mathrm{T}}, \boldsymbol{\varepsilon}_{U}^{\mathrm{T}})^{\mathrm{T}}$ has the following variance-covariance matrix

$$\boldsymbol{\Omega}_{\mathrm{m}} = E(\boldsymbol{\varepsilon}_{m}\boldsymbol{\varepsilon}_{m}^{\mathrm{T}}) = \boldsymbol{\Sigma}_{LU} \otimes \mathbf{I}_{n}, \tag{11}$$

with

$$\boldsymbol{\Sigma}_{LU} = \begin{bmatrix} \sigma_{LL} & \sigma_{LU} \\ \sigma_{LU} & \sigma_{UU} \end{bmatrix}.$$

By applying the generalized least squares estimation method for linear regression model (3) with (9), we can obtain the seeming unrelated regression-based Minmax estimator of $\boldsymbol{\beta}_{\mathrm{m}}$ as

$$\hat{\boldsymbol{\beta}}_{SUR}^{mm} = \begin{bmatrix} \hat{\boldsymbol{\beta}}_{SUR}^{L} \\ \hat{\boldsymbol{\beta}}_{SUR}^{U} \end{bmatrix} = \left[\mathbf{X}_{m}^{\mathrm{T}}\boldsymbol{\Omega}_{m}^{-1}\mathbf{X}_{m}\right]^{-1}\mathbf{X}_{m}^{\mathrm{T}}\boldsymbol{\Omega}_{m}^{-1}\mathbf{Y}_{m}. \tag{12}$$

However, the covariance matrix $\boldsymbol{\Omega}_{m}$ is unknown, then $\hat{\boldsymbol{\beta}}_{SUR}^{mm}$ is infeasible. To solve this problem, we can replace the unknown elements $\boldsymbol{\Omega}_{\mathrm{m}}$ by their estimators respectively. Define

$$\hat{\sigma}_{LL} = \frac{(\mathbf{Y}_{L} - \mathbf{X}_{L}\hat{\boldsymbol{\beta}}^{L})^{\mathrm{T}}(\mathbf{Y}_{L} - \mathbf{X}_{L}\hat{\boldsymbol{\beta}}^{L})}{n-p}, \hat{\sigma}_{UU} = \frac{(\mathbf{Y}_{U} - \mathbf{X}_{U}\hat{\boldsymbol{\beta}}^{U})^{\mathrm{T}}(\mathbf{Y}_{U} - \mathbf{X}_{U}\hat{\boldsymbol{\beta}}^{U})}{n-p},$$

and

$$\hat{\sigma}_{LU} = \frac{(\mathbf{Y}_L - \mathbf{X}_L\hat{\boldsymbol{\beta}}^L)^{\mathrm{T}}(\mathbf{Y}_U - \mathbf{X}_U\hat{\boldsymbol{\beta}}^U)}{n},$$

where $\hat{\boldsymbol{\beta}}^L$ and $\hat{\boldsymbol{\beta}}^U$ are the Minmax estimators which were defined in equation (4). Then we can define the feasible SUR-Minmax estimator for $\boldsymbol{\beta}$ as

$$\hat{\boldsymbol{\beta}}^{mm}_{FSUR} = \begin{bmatrix} \hat{\boldsymbol{\beta}}^L_{FSUR} \\ \hat{\boldsymbol{\beta}}^U_{FSUR} \end{bmatrix} = \left[\mathbf{X}_m^{\mathrm{T}}\hat{\boldsymbol{\Omega}}_m^{-1}\mathbf{X}_m\right]^{-1}\mathbf{X}_m^{\mathrm{T}}\hat{\boldsymbol{\Omega}}_m^{-1}\mathbf{Y}_m. \tag{13}$$

with $\hat{\boldsymbol{\Omega}}_m = \hat{\boldsymbol{\Sigma}}_{LU} \otimes \mathbf{I}_n$,

$$\hat{\boldsymbol{\Sigma}}_{LU} = \begin{bmatrix} \hat{\sigma}_{LL} & \hat{\sigma}_{LU} \\ \hat{\sigma}_{LU} & \hat{\sigma}_{UU} \end{bmatrix}.$$

By the classic theory of linear regression model, we know that the generalized least-squares estimator $\hat{\boldsymbol{\beta}}^{mm}_{FSUR}$ is efficient than the ordinary least-squares estimator $\hat{\boldsymbol{\beta}}^{mm}$. If $\sigma_{LU} = 0$, they are equal. In practice, We can first to test whether $\sigma_{LU} = 0$ is or not. we can use the Lagrange multiplier test method of Breusch and Pagan (1980) to this testing problem.

3.2 The Efficient SUR-Based CR Estimator

Similarly, for the linear regression models (5) and (6), the feasible SUR-based CR estimator for $\boldsymbol{\beta}^{cr}$ can be defined as

$$\hat{\boldsymbol{\beta}}^{cr}_{FSUR} = \begin{bmatrix} \hat{\boldsymbol{\beta}}^c_{FSUR} \\ \hat{\boldsymbol{\beta}}^r_{FSUR} \end{bmatrix} = \left[\mathbf{X}_{cr}^{\mathrm{T}}\hat{\boldsymbol{\Omega}}_{cr}^{-1}\mathbf{X}_m\right]^{-1}\mathbf{X}_{cr}^{\mathrm{T}}\hat{\boldsymbol{\Omega}}_{cr}^{-1}\mathbf{Y}_{cr}. \tag{14}$$

with $\hat{\boldsymbol{\Omega}}_{cr} = \hat{\boldsymbol{\Sigma}}_{cr} \otimes \mathbf{I}_n$, $\hat{\boldsymbol{\Sigma}}_{cr} = \begin{bmatrix} \hat{\sigma}_{cc} & \hat{\sigma}_{cr} \\ \hat{\sigma}_{cr} & \hat{\sigma}_{rr} \end{bmatrix}$, $\hat{\sigma}_{cr} = \frac{(\mathbf{Y}^c - \mathbf{X}^c\hat{\boldsymbol{\beta}}^c)^{\mathrm{T}}(\mathbf{Y}^r - \mathbf{X}^r\hat{\boldsymbol{\beta}}^r)}{n}$, and

$$\hat{\sigma}_{cc} = \frac{(\mathbf{Y}^c - \mathbf{X}^c\hat{\boldsymbol{\beta}}^c)^{\mathrm{T}}(\mathbf{Y}^c - \mathbf{X}^c\hat{\boldsymbol{\beta}}^c)}{n-p}, \hat{\sigma}_{rr} = \frac{(\mathbf{Y}^r - \mathbf{X}^r\hat{\boldsymbol{\beta}}^r)^{\mathrm{T}}(\mathbf{Y}^r - \mathbf{X}_U\hat{\boldsymbol{\beta}}^r)}{n-p},$$

where $\hat{\boldsymbol{\beta}}^c$ and $\hat{\boldsymbol{\beta}}^r$ are the CR estimators which were defined in equation (8).

3.3 The Two-Step SUR-Based CR Estimation

The direct SUR-based CR estimatior (13) is obtained on the assumption that model (5)-(6) are data generating models. However, it is not ture. Model (5)-(6) are generated from model (1)-(2). Then, the variance and covariance of $\boldsymbol{\varepsilon}^c$ and $\boldsymbol{\varepsilon}^r$ can be computed by the assumptions (9) and (10). We can show that

$$\begin{aligned} \boldsymbol{\Phi}_c = \mathrm{Cov}(\boldsymbol{\varepsilon}_c) &= \mathrm{Cov}\left(\tfrac{\boldsymbol{\varepsilon}_U+\boldsymbol{\varepsilon}_L}{2}\right) = \tfrac{1}{4}\mathrm{E}\left[(\boldsymbol{\varepsilon}_U + \boldsymbol{\varepsilon}_L)(\boldsymbol{\varepsilon}_U + \boldsymbol{\varepsilon}_L)^{\mathrm{T}}\right] \\ &= \tfrac{1}{4}\mathrm{E}(\boldsymbol{\varepsilon}_U\boldsymbol{\varepsilon}_U^{\mathrm{T}} + \boldsymbol{\varepsilon}_L\boldsymbol{\varepsilon}_L^{\mathrm{T}} + 2\boldsymbol{\varepsilon}_U\boldsymbol{\varepsilon}_L^{\mathrm{T}}) \\ &= \tfrac{\sigma_{UU}+\sigma_{LL}+2\sigma_{UL}}{4}\mathbf{I}_n, \end{aligned}$$

$$\boldsymbol{\Phi}_r = \mathrm{Cov}(\boldsymbol{\varepsilon}_r) = \mathrm{Cov}\left(\frac{\boldsymbol{\varepsilon}_U - \boldsymbol{\varepsilon}_L}{2}\right) = \frac{\sigma_{UU} + \sigma_{LL} - 2\sigma_{UL}}{4}\mathbf{I}_n,$$

$$\begin{aligned} \boldsymbol{\Phi}_{cr} = \mathrm{Cov}(\boldsymbol{\varepsilon}_c, \boldsymbol{\varepsilon}_r) &= \tfrac{1}{4}\mathrm{E}\left[(\boldsymbol{\varepsilon}_U + \boldsymbol{\varepsilon}_L)(\boldsymbol{\varepsilon}_U - \boldsymbol{\varepsilon}_L)^{\mathrm{T}}\right] \\ &= \tfrac{1}{4}\mathrm{E}(\boldsymbol{\varepsilon}_U\boldsymbol{\varepsilon}_U^{\mathrm{T}} - \boldsymbol{\varepsilon}_L\boldsymbol{\varepsilon}_L^{\mathrm{T}}) \\ &= \tfrac{\sigma_{UU}-\sigma_{LL}}{4}\mathbf{I}_n. \end{aligned}$$

Therefore, the $2n \times 1$ disturbance vector $\boldsymbol{\varepsilon}_{cr} = (\boldsymbol{\varepsilon}_c^{\mathrm{T}}, \boldsymbol{\varepsilon}_r^{\mathrm{T}})^{\mathrm{T}}$ has the following variance-covariance matrix

$$\boldsymbol{\Lambda}_{cr} = \mathrm{Cov}(\boldsymbol{\varepsilon}_{cr}) = \begin{bmatrix} \boldsymbol{\Phi}_c & \boldsymbol{\Phi}_{cr} \\ \boldsymbol{\Phi}_{cr} & \boldsymbol{\Phi}_r \end{bmatrix}.$$

Then, the two-step efficient SUR-based CR estimators are

$$\hat{\boldsymbol{\beta}}^{\mathrm{cr}}_{TSUR} = \begin{bmatrix} \tilde{\boldsymbol{\beta}}^c_{TSUR} \\ \tilde{\boldsymbol{\beta}}^r_{TSUR} \end{bmatrix} = \left[\mathbf{X}_{cr}^{\mathrm{T}}\boldsymbol{\Lambda}_{cr}^{-1}\mathbf{X}_{cr}\right]^{-1}\mathbf{X}_{cr}^{\mathrm{T}}\boldsymbol{\Lambda}_{cr}^{-1}\mathbf{Y}_{cr}. \tag{15}$$

Table 1. Mushroom interval-valued data set

Species	Y	X	Z	Species	Y	X	Z
1	[3,8]	[4,9]	[0.5,2.5]	13	[3.5,8]	[4,10]	[1,2]
2	[6,21]	[4,14]	[1,3.5]	14	[7,14]	[8,14]	[1.5,2.5]
3	[4,8]	[5,11]	[1,2]	15	[8,20]	[9,19]	[3,5]
4	[6,7]	[4,7]	[3,4.5]	16	[2.5,4]	[2.5,4.5]	[0.4,0.7]
5	[5,12]	[2,5]	[1.5,2.5]	17	[7,19]	[8,15]	[2,3.5]
6	[5,15]	[4,10]	[2,4]	18	[5,15]	[6,15]	[2.5,3.5]
7	[4,11]	[3,7]	[0.4,1]	19	[8,12]	[6,12]	[1.5,2]
8	[5,10]	[3,6]	[1,2]	20	[2,6]	[3,7]	[0.4,0.8]
9	[2.5,4]	[3,5]	[0.4,0.7]	21	[6,12]	[6,12]	[1.5,2]
10	[2.5,6]	[1.5,3.5]	[1,1.5]	22	[6,12]	[6,16]	[1,2]
11	[1.5,2.5]	[3,6]	[0.25,0.35]	23	[5,17]	[4,14]	[1,3.5]
12	[4,15]	[4,15]	[1.5,2.5]				

Table 2. Performance of the methods

Method	$\mathrm{RMSE_L}$	$\mathrm{RMSE_U}$	$\mathrm{r_L^2}$	$\mathrm{r_U^2}$
Minmax	1.131198	3.241086	0.6241114	0.6146621
SUR-Minmax	1.132775	3.169804	0.6308301	0.6275454
CR	1.479281	3.003173	0.5105622	0.6631535
SUR-CR	1.351316	2.898771	0.5473502	0.6813018
Two-step SUR-CR	1.348811	2.852429	0.5540649	0.7014462

The corresponding feasible estimator is

$$\hat{\boldsymbol{\beta}}_{FTSUR}^{cr} = \begin{bmatrix} \tilde{\boldsymbol{\beta}}_{FTSUR}^{c} \\ \tilde{\boldsymbol{\beta}}_{FTSUR}^{r} \end{bmatrix} = \left[\mathbf{X}_{cr}^{\mathrm{T}} \hat{\boldsymbol{\Lambda}}_{cr}^{-1} \mathbf{X}_{cr}\right]^{-1} \mathbf{X}_{cr}^{\mathrm{T}} \hat{\boldsymbol{\Lambda}}_{cr}^{-1} \mathbf{Y}_{cr}, \tag{16}$$

where $\hat{\boldsymbol{\Lambda}}$ is defined as $\boldsymbol{\Lambda}$ by replacing unknown parameters $\sigma_{UU}, \sigma_{UU}, \sigma_{UL}$ by their estimators, respectively.

It is noted that if $\sigma_{UL} = 0$, model (1) and (2) are independence, we have $\hat{\boldsymbol{\beta}}_{SUR}^{mm} = \hat{\boldsymbol{\beta}}^{mm}$. If $\sigma_{UU} \neq \sigma_{LL}$, model (5) and (6) are not independence as $\boldsymbol{\Phi}_{cr} = \frac{\sigma_{UU}-\sigma_{LL}}{4}\mathrm{I}_n \neq \mathbf{0}_n$, then $\tilde{\boldsymbol{\beta}}_{TSUR}^{cr} \neq \hat{\boldsymbol{\beta}}^{cr}$. If $\sigma_{UU} = \sigma_{LL}$, then $\boldsymbol{\Phi}_{cr} = \frac{\sigma_{UU}-\sigma_{LL}}{4}\mathrm{I}_n = \mathbf{0}_n$, we have $\hat{\boldsymbol{\beta}}_{TSUR}^{cr} = \hat{\boldsymbol{\beta}}^{cr}$.

4. Real Data Analysis

In this section, we shall analyse a real data to illustrate the finite sample properties of the proposed procedures.

Mushroom data set consists of a set of 23 species described by 3 interval variables, where the response variable Y is the stipe thickness, and the covariates X is the stipe length, Z is the pileus cap width. These mushroom species are members of the genus Agaricies. The specific variables and their values are extracted from the Fungi of California Species. The data set given in Table 1 was obtained from Billard and Diday (2006).

The performance assessment of the above estimating approaches will be based on the following measures: the lower boundary root mean-square error ($\mathrm{RMSE_L}$) and the upper boundary root mean-square error ($\mathrm{RMSE_U}$), the square of the lower bound correlation coefficient ($\mathrm{r_L^2}$) and the square of the upper bound correlation coefficient ($\mathrm{r_U^2}$) These measures, calculated from the observed values $[\mathrm{y_{Li}}, \mathrm{y_{Ui}}]$ and their corresponding leave-one-out cross-validation predicted values based on Minmax, SUR-Minmax, CR, SUR-CR and Two step SUR-CR estimating methods. They are defined by

$$\mathrm{RMSE_L} = \sqrt{\frac{\sum_{i=1}^{n}(y_{Li} - \hat{y}_{Li})^2}{n}}, \mathrm{RMSE_U} = \sqrt{\frac{\sum_{i=1}^{n}(y_{Ui} - \hat{y}_{Ui})^2}{n}}, \mathrm{r_L^2} = \rho^2(Y_L, \hat{Y}_L), \mathrm{r_U^2} = \rho^2(Y_U, \hat{Y}_U).$$

The results can be found in Table 2. Thus, we conclude that the SUR-MinMax method outperforms the MinMax method, the two-step SUR-CR method outperforms the SUR-CR method while the SUR-CR method outperforms the CR method.

5. Conclusion

This paper applied the SUR method to improve the efficiency of Minmax method and CR method. Real data sets are analysed to examine the performance of our proposed methods and the results are satisfactory.

For the Minmax method and CR method, we have some comments. In the simulation studies of Lima Neto and De Carvalho (2008) and other literatures, the interval data sets are often constructed in two different ways, the first with values of the centre and range of the intervals simulated independently, the second with values of the interval mid-points and ranges related according to a linear relationship. However, these assumptions are usually not satisfied in real data analysis.

Conflicts of Interest

The author declares no conflict of interest.

References

Billard, L., & Diday, E. (2000). Regression analysis for interval-valued data. In Kiers, H. A. L., Rasson, J. P. (Eds.), *Data Analysis, Classification, and Related Methods, Proceedings IFCS2000, Namur* (pp.369-374). Heidelberg: Springer Verlag. https://doi.org/10.1007/978-3-642-59789-35.8

Billard, L., & Diday, E. (2002). *Symbolic regression analysis.* In Classification, Clustering and Data Analysis, Proceedings of the Eighenth Conference of the International Federation of Classification Societies (IFCS02), Springer, Poland, 281-288. https://doi.org/10.1007/978-3-642-56181-83.1

Billard, L., & Diday, E. (2006). *Symbolic Data Analysis: Conceptual Statistics and Data Mining*. John Wiley, New York. https://doi.org/10.1002/9780470090183

Breusch, T. S., & Pagan, A. R. (1980). The Lagrange Multiplier Test and Its Applications to Model Specification in Econometrics. *Review of Economic Studies, 47,* 239-253. https://doi.org/10.2307/2297111

Giordani, P. (2015). Lasso-constrained regression analysis for interval-valued data. *Adv Data Anal Classif, 9*, 5-19. https://doi.org/10.1007/s11634-014-0164-8

Lim, C. (2016). Interval-valued data regression using nonparametric additive models. *Journal of the Korean Statistical Society, 45,* 358-370. https://doi.org/10.1016/j.jkss.2015.12.003

Lima Neto, E. A., & De Carvalho, F. A. T. (2008). Centre and range method to fitting a linear regression model on symbolic interval data. *Computational Statistics and Data Analysis, 52*(3), 1500-1515. https://doi.org/10.1016/j.csda.2007.04.014

Souza, L. C., & Souza, R. M. C. R., & Amaral, G. J. A., & Filho, T. M. S. (2017). A Parametrized Approach for Linear Regression of Interval Data. *Knowledge-Based Systems, 131*, 149-159. https://doi.org/10.1016/j.knosys.2017.06.012

Wei, Y., & Wang, S. S., & Wang, H. W. (2015). Interval-valued data regression using partial linear model. *Journal of Statistical Computation and Simulation, 87,* 3175-3194. https://doi.org/10.1080/00949655.2017.1360298

Xu, W. (2010). Symbolic Data Analysis: Interval-Valued Data Regression (Ph.D. thesis). University of Georgia.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association, 57*, 348-368. https://doi.org/10.1080/01621459.1962.10480664

Copyrights

International Journal of Statistics and Probability; Vol. 9, No. 2; March 2020
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

D-optimal Design in Linear Model With Different Heteroscedasticity Structures

BODUNWA, O. K.[1], FASORANBAKU, O. A.[1]

[1]Department of Statistics, Federal University of Technology, Akure

Correspondence: BODUNWA, O. K, Department of Statistics, Federal University of Technology, Akure. E-mail: okbodunwa@futa.edu.ng

Received: January 6, 2020 Accepted: February 21, 2020 Online Published: February 26, 2020

doi:10.5539/ijsp.v9n2p7 URL: https://doi.org/10.5539/ijsp.v9n2p7

Abstract

In this paper, we developed D-optimal design in linear model with two explanatory variables in the presence of heteroscedasticity. A sequential method of getting D-optimal design was adopted. Two different structures were used based on the literatures; it was found that the optimal design takes the extreme values of the design region. The results of simulated data was justified with real life data from the kinematic viscosity of a lubricant, in stokes, as a function of temperature and pressure which was used as discussed in Linssen (1975). The relative efficiency of other designs with respect to D-optimal designs was determined. Three correction methods was adopted from weighted least square method for heteroscedasticity problem, it was found that the correction method tagged HCW1 performed better.

Keywords: D-optimal design, Heteroscedasticity, experimental design, sequential method, correction measure

1. Introduction

Experimentation is the process of planning a study to meet specified objectives which constitutes a foundation of the empirical sciences (Zhu, 2012). One major advantage of experiment is its ability to control the experimental conditions; as well as to determine the variables to include in a study (FackleFornius, 2008). Since the introduction of experimental design principle in the first half of the 1930, optimal experimental designs have been gaining attention and had become useful tools among researchers in various fields (Atkinson and Donev, 1992; Atkinson, 1996; Atkinson, Donev and Tobias, 2007; Berger and Wong, 2009). There are various design criteria, D-optimality has been the most frequently used; and often performs better than other criteria (Zocchi and Atkinson, 1999; Atkinson et al., 2007). Hence, the D-optimality has become one of the most popular criteria which involve designs that minimize the generalized variance of the parameter vector. The D-optimal designs seek to minimize $|(\mathrm{X}'\mathrm{X})^{-1}|$ (dispersion matrix) or equivalently maximise the determinant of the information matrix $(\mathrm{X}'\mathrm{X})$ of the design through some forms of statistical modeling such as regression model. One of the important assumptions of the standard regression model is that the variance of the error terms (disturbance term, u_i) must be equal across the observations which is refers to as homoscedastic with the model$y = x\beta + u_i$ where $[E(u_i^2) = \sigma^2 \quad i = 1,2,\cdots,n]$. However, in real life situations, this assumption is often violated and the variances of the error terms are not the same. The condition where error terms have different variances is termed heteroscedasticity $[E(u_i^2) = \sigma_i^2 \quad i = 1,2,\cdots,n]$ that is, unequal variance across the observations (Lambert, 2013; Knaub, 2017). Heteroscedasticity, which is often referred to as a "problem" that needs to be "solved" or "corrected" is the change in variance of predicted y, given different values of the independent variables (Knaub, 2011, 2017). The aim of this research work is to examine D- optimal Designs with different heteroscedastcity Structures and the objectives are to construct D-optimal design with different heteroscedasticity structures, to obtain the relative efficiencies of other designs with respect to D-optimal design, to determine the heteroscedasticity correction measure that will produce the most efficient D-optimal design in the different structures, determining the relative efficiencies of the parameters of the D-Optimal design model and to establish the best heteroscedasticity correction measure to achieve the most Efficient Parameter Estimation for D-Optimal Design.

Yan and Raymond (2001) presented D-optimal designs for two- variable logistic regression models where two-variable were fitted in the logistic regression models.Jafari (2013) found locally D-optimal design for a logit model in discrete choice experiment where there are many alternative set for people to make their choice using D-optimal design for the combination of the level of attributes to create alternatives. Jafari, *et.al,*(2014) worked on D-optimal design for logistic regression model with three independent variables; they obtained a locally D-optimal design for several specific states, presented certain designs with different points and calculated the subject optimality based on space of the parameters.

Jafari and Maram (2015) explored the notion of Bayesian D-optimal design for logistic regression model with exponential distribution for random intercept and obtained Bayesian D-optimal design; the method to maximize the Bayesian D-optimal criterion which is a function of the quasi- information matrix that depends on the unknown parameters of the model.

Jesús López-Fidalgo and Garcet-Rodríguez, (2004) considered the problem of constructing optimal designs for regression models when the design space is a product space and some of the variables are not under the control of the practitioner.Zhide and Douglas (2004) found locally D-optimal designs for multistage models and heteroscedastic polynomial regression model where they considered the construction of locally D-optimal designs for non-linear, multistage model in which one observes a binary response variable.Gaviriaa and López-Ríosb (2014) worked on locally D-optimal designs with Heteroscedasticity: a comparison between two methodologies, it was found that the optimal design point takes the extreme values for both methods. These prior studies were more particular about the construction of the optimal designs with different models under some assumptions of the explanatory variables. In this study, construction of D-optimal designs in linear model with two explanatory variables in which there is a problem of heteroscedasticity in the model were examined. Different structures were used and the effects were also found on the optimal design.

2. Material and Method

2.1 Simulation Study

Starting with a linear regression model of the form (2.1)

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + e_i \tag{2.1}$$

Where e_i is the error term which is a stochastic term assumed to be normally distributed with mean zero and variance σ_i^2 i.e. $e_i \sim N(0,\ \sigma_i^2)$.These $x_i s$ are fixed independently variables and y_i is the dependent variable and β_i are parameters that are known. The generations of the data used for independent variables are random variables that are normally distributed

$$x_1 = \left((1 - K\text{^}2)\text{^}0.5\right) * E_1 + K * E_2 \tag{2.2}$$

$$x_2 = \left((1 - K\text{^}2)\text{^}0.5\right) * E_2 + K * E_2 \tag{2.3}$$

Where K is the correlation between the explanatory variables, $E_1\ and\ E_2$ are the independent standard normal distribution with mean zero and the unit variance. The response variable was therefore obtained with equation

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i \tag{2.4}$$

Where $e_i = Z_i\sqrt{Var(e_i)}$, $Z_i \sim N(0,1)$ $\quad i = 1,\ 2.$

$e_1 \sim N(0, X_{2i}^2)$ (Park, 1966, White, 1980, Guajarati *et. al* 2012)

$e_i \sim N\left(0, Exp(x_2^2)\right)$ (Box and Hill, 1974, Harvey, 1976)

The $Var(e_i)$ took any of the structures in equations 2 and 3. The simulations were carried out in one thousand times (1000) at eight sample sizes of 10, 20, 30, 40, 50, 100, 250 and 500.

In order to correct the heteroscedasticity problem with the selected structures, the weighted least square methods was adopted and the $log\hat{e}_i^2\ was$ regress $on\ (x_1, x_2)$ to have Heteroscedasticity Correction Weighted 1 (HCW1), $log\hat{e}_i^2\ on\ (x_1,\ x_2, x_1^2, x_2^2)$ to have Heteroscedasticity Correction Weighted 2 (HCW2) and $log\hat{e}_i^2\ on\ (x_1,\ \ x_2, x_1^2,\ \ x_2^2,\ x_1 x_2)$ to have Heteroscedasticity Correction Weighted 3 (HCW3).

2.2 Construction of D-optimal Design

There are several methods at hand on the practices of determining the optimal design. These include algorithms, sequential, analytical, numerical and graphical methods, used separately or in combinations. There is no method that is generally favorable; it depends on the problem at hand. The method selected in this research work is sequential method of getting D-optimal design; we find the D-optimal design for model with different variance structure of the error term was essentially obtained. For the model (2.1) used in this study, the number of p is 3. Therefore the partial derivative for the model is

$$f'(x_i) = (1,\ x_1,\ x_2) \tag{2.5}$$

The information matrix is now

$$M(\xi) = \sum w_i f(x_i)\, f'(x_i) \tag{2.6}$$

Beginning with p-point design, we get 3×3 design matrix of the form

$$X_3 = \begin{bmatrix} 1.000000 & 1.000000 & 0.322035 \\ 1.000000 & -0.494439 & 0.413935 \\ 1.000000 & -0.235592 & -0.026634 \end{bmatrix} \quad (2.7)$$

It should be noted that the procedure requires a sufficient number of observations because we have to ensure that the inverse $|X_N'X_N|^{-1}$ exist. A simple condition that will guarantee the inverse exists is to have the number of different design points greater than or equal to the number of parameters, that is $N \geq p$

The design points are selected within the range of $-1 \leq x \leq 1$ for the variables. The largest $s(x_a, \xi)$ is found for $x_1 = 1.00000\ and x_2 = -1.000000$, so these design points were added to design matrix X_3 and the design matrix is now

$$X_4 = \begin{bmatrix} 1.000000 & 1.000000 & 0.322035 \\ 1.000000 & -0.494439 & 0.413935 \\ 1.000000 & -0.235592 & -0.026634 \\ 1.000000 & 1.000000 & -1.000000 \end{bmatrix} \quad (2.8)$$

The iteration continued until the condition for getting optimal design was reached. The maximum $s(x_i, \xi)$ value decreases as N increases, according to the general equivalence theorem (Kiefer and Wolfowitz, 1960), a D-optimal design satisfies the condition that $s(x_a, \xi) \leq p$.

2.3 Relative Efficiencies of D-optimal to Other Designs

The Efficiency of D-optimal design ξ_D with respect to the other design is

$$D_{eff} = \left(\frac{|M(\xi)|}{|M(\xi_D|)}\right)^{1/p} \quad (2.9)$$

Where p is the number of parameters of the model and $M(\xi)$ denotes the information matrix of the design ξ which is another design different from D-optimal design. Relative efficiencies of the parameters of the D-optimal design and non optimal designs models were also done to establish the result of D-optimal designs point. The design points for all the structures were obtained with respect to the probability, number of iteration, the standardized variance.

2.4 Most Efficient Correction Method

The best correction method among the one named HCW1, HCW2 and HCW3 was determined. This was done by calculating the variances for the probabilities of the D-optimal designs taking the design points as x and the probabilities as $f(x)$. The minimum variances were selected for the structures for all the sample sizes and the method that has highest values was chosen to be the most efficient.

2.5 Real Life Application

Construction of D-optimal design in the presence of heteroscedasticity for the model (1) was applied to a real life data, a secondary data from the kinematic viscosity of a lubricant, in stokes, as a function of temperature(o_C), and pressure in atmospheres (atm), was used as discussed in Linssen (1975) where y is predicted In (viscosity), x_1 is temperature, and x_2 is pressure to justify the simulated data.

3. Result and Discussion

In this work, D-optimal designs with two different heteroscedasticity structures were constructed when there is no heteroscedasticity (No H) and when there is (HR). It was generally found that the D-optimal designs take the extreme values of the response variables which follow uniform distribution of the experimental units

Table 3.1. D-Optimal Designs for the Structures

Structures		(-1, -1)	(-1, 1)	(1, -1)	(1, 1)
$\sigma^2 X_{2i}^2$	No H	44(0.25143)	44(0.25143)	44(0.25143)	43(0.24571)
	HR	28(0.24138)	30(0.25862)	295(0.25000)	29(0.25000)
$\sigma^2 Exp(x_2^2)$	No H	44(0.25143)	44(0.25143)	43(0.24571)	44(0.25143)
	HR	22(0.23656)	24(0.25806)	24(0.25806)	23(0.24731)

Table 3.1 presents the construction of the D-optimal when there is no heteroscedasticity and when there is heteroscedasticity for the error structures. It can be seen that the D-optimal designs when there is no heteroscedasticity for the two structures were same reason being that the error term have equal variance. The optimal designs even though the model has three parameters the design consists four points which are the extreme points of the regression range.From the table, it can been seen that

$$\xi^* = \begin{Bmatrix} (-1,-1) & (-1,1) & (1,-1) & (1,1) \\ 0.24138 & 0.258620.25000 & & 0.250000 \end{Bmatrix} \quad (3.1)$$

if there are 116 experimental units, 28 should be allocated to when $x_1 = -1$ and$x_2 = -1$, 30 should be for when $x_1 = -1$ and$x_2 = 1$. In the same vein, 29 should be allocated to when $x_1 = 1$ and $x_2 = -1$ and when$x_1 = 1$ and$x_2 = 1$.

Considering D-optimal design for the second structure,

$$\xi^* = \begin{Bmatrix} (-1,-1) & (-1,1) & (1,-1) & (1,1) \\ 0.23656 & 0.258060.25806 & & 0.24731 \end{Bmatrix} \quad (3.2)$$

Equation shows that if there are 93 experimental units, 22 should be allocated to when $x_1 = -1$ and $x_2 = -1$, 24 should be for when $x_1 = -1$ and $x_2 = 1$and when$x_1 = 1$ and $x_2 = -1$, 23 for when $x_1 = 1$ and $x_2 = 1$.

Table 3.2. D-optimal Designs for the real life data

	(-1,-1)	**(-1,1)**	**(1,-1)**	**(1,1)**
HR	16(0.30000)	11(0.20000)	16(0.30000)	11(0.20000)

The results still revealed that the D-optimal design for the real life data presented above affirmed the result from simulated data in the sense that the design point takes the extreme values of the design region.

The relative efficiencies of D-optimal design with respect to other designs that are not optimal using the same method of construction of D-optimal design from the starting design matrix of point 4 is given below for the structures.

Table 3.3. Relative Efficiency Table

$\sigma^2 X_{2i}^2$		$\sigma^2 Exp(x_2^2)$	
No of Iteration	D-efficiency	No of Iteration	D-efficiency
4	0.0019	4	0.0043
5	0.0225	5	0.0329
6	0.0331	6	0.0453
⋮	⋮	⋮	⋮
114	0.9829	91	0.9788
115	0.9914	92	0.9894

Table 3.3 shows that the D-optimal design has close efficiency to other design especially the one closed to the design point meaning that the closer the D-efficient to one, the better. The no of iteration for D-optimal design for the first structure is 116 and for the second structure 93. Next table present the D-efficiency of the real life data.

Table 3.4. Relative Efficiencies of other Designs for real life data

I	$\boldsymbol{D_{eff}}$
4	0.002128
5	0.003511
6	0.004728
⋮	⋮
901	0.9869
902	0.9931

To determine the best correction method, the variances of the probability in the design point of the D-optimal design were calculated using different sample sizes. The best method was chosen on the basis of the one with minimum variance. Table 3.5 presented the variances of design points.

Table 3.5. Determination of the best Correction Method

Forms	Correction Methods	Sample size 10	20	30	40	50	100	250	500
$\sigma^2 X_{2i}^2$	HCW1	1.24996	1.26050	1.24407	**1.23217**	1.24290	**1.24434**	**1.24407**	1.25584
	HCW2	**1.24978**	**1.25912**	1.25000	1.24386	**1.24113**	1.25762	1.25584	**1.24172**
	HCW3	1.24995	1.25969	**1.24362**	1.24386	1.25718	1.25000	1.25598	1.24223
$\sigma^2 Exp(x_2^2)$	HCW1	**1.25774**	1.26146	1.24362	**1.23030**	1.24362	**1.25534**	**1.24481**	**1.25628**
	HCW2	1.25786	**1.25868**	1.23777	1.24401	**1.24144**	1.25786	1.25546	1.25739
	HCW3	1.27398	1.25899	**1.21102**	1.2386	1.25899	1.25718	1.25523	1.25762

From the table, number of appearance of minimum variance values in HCW1 is more than the other two. Therefore HCW1 is assumed to be performing better.

4. Conclusion

In the study, constructions of D-optimal designs in the presence of Heteroscedasticity for two different structures were considered with when there is no Heteroscedasticity in the data.

It was generally found that the D-optimal designs take the extreme values of the response variables which follow uniform distribution of the experimental units which can be interpreted as taking the least and the highest values of the explanatory variables in order to get best output through the response variable. To verify the above findings, a set of real life data (secondary data) was used and the design points for D-optimal designs were same with simulated data.

The relative efficiencies of other designs under different Heteroscedasticity structures were found to prove the strength of the design. Determination of the best correction method was also found. This was achieved by comparing the variances of the selected correction methods with respect to sample sizes for all the structures used in the study. It was found that the correction method with minimum variance that showed the efficiency of the method represented by (HCW1) which was done by regressing $log\hat{e}_i^2$ on the linear combinations of x_1 and $x_2$performed better than the remaining two.

References

Atkinson, A. C., Donev, A. N., & Tobias, R. D. (2007). *Optimum Experimental Designs, with SAS.* Oxford University Press.

Atkinson, A. C. (1996). The usefulness of optimum experimental designs.*Journal of the Royal Statistical Society. Series B*, *58*, 59-76.https://doi.org/10.1111/j.2517-6161.1996.tb02067.x

Atkinson, A. C., & Donev, A. N. (1992). *Optimum Experimental Designs*. Oxford University Press, Oxford.

Berger, P. F., & Wong, K. W. (2009). *An Introduction to optimal designs for social and Biomedical research.*A John Wiley & Sons, Ltd. Publication.https://doi.org/10.1002/9780470746912

Box, G. E. P., & Hill, W. J. (1974). Correction Inhomogenity of Variance with Power Transformation Weighting. *Technometrics, 16*(3), 385-389. https://doi.org/10.1080/00401706.1974.10489207

Fackle, F. E. (2008). *Optimal Design of Experiments for the Quadratic Logistic Model.* A Thesis submitted to the Department of Statistics, Stockholm University, Stockholm, in partial fulfillment of Doctor of Philosophy in Statistics.

Kiefer, J., & Wolfowitz, J. (1960). The Equivalence of Two Extremum Problems. *Canad. J. Math., 12*, 363-366. https://doi.org/10.4153/CJM-1960-030-4

Knaub, J. R. J. (2011). Ken Brewer and the Coefficient of Heteroscedasticity as Used in Sample Survey Inference.*Pakistan Journal of Statistics*, *27*(4), 397-406.

Knaub, J. J. R. (2017). Essential Heteroscedasticity. Retrieved from https://www.researchgate.net/publication/320853387_Essential_Heteroscedasticity

Gaviriaa, J. A., &López-Ríosb, V. I. (2014). Locally D-Optimal Designs with Heteroscedasticity: A Comparison between Two Methodologies. *RevistaColombiana de Estadística, 37*(1), 95-110. https://doi.org/10.15446/rce.v37n1.44360

Gujarati, N. D., Porter, C. D., & Gunasekar, S. (2012). "Basic Ecocnometric" (Fifth Edition) New Delhi: Tata

McGraw-Hill.

Jafari, H., & Maram. (2015). Bayesian D-optimal design for Logistic Regression model with Exponential distribution for random intercept. *Journal of Statistical Computation and Simulation.*

Jafari, H., Khazai, S., & Khaki, Y. (2014). D-optimal design for logistic regression model withthree independent variables.*Journals of Asian Scientific Research, 4*(3), 120-124.

Lambert, B. (2013). Heteroscedasticity Summary, June 3, 2013, YouTube. Retrieved fromhttps://youtu.be/zRklTsY9w9c

Linssen, H. N. (1975). Nonlinearity measures: a case study, Statist. *Neerland, 29*, 93-99. https://doi.org/10.1111/j.1467-9574.1975.tb00253.x

López-Fidalgo, J., & Garcet-Rodríguez, S. A. (2004). Optimal experimental designs when some independent variables are not subject to control. *Journal of the American Statistical Association*, *99*(468), 1190-1199. https://doi.org/10.1198/016214504000001736

Park, R. E. (1966). Estimation with Heteroscedastic Error terms. *Econometrica, 34*, 888-892. https://doi.org/10.2307/1910108

White, H. (1980). A Heteroscedastic-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity. *Econometrica, 48*, 817-838. https://doi.org/10.2307/1912934

Zhu, C. (2012). *Construction of Optimal Designs in Polynomial Regression Models*. A Thesis submitted to the Faculty of Graduate Studies of The University of Manitoba in Partial Fulfillment of the Requirements for the Degree of Master of Science Department of Statistics University of Manitoba Winnipeg, Manitoba, Canada.

Zocchi, S. S., & Atkinson, A. C. (1999). Optimum experimental designs for multinomial logistic models. *Biometrics*, *55*, 437-444. https://doi.org/10.1111/j.0006-341X.1999.00437.x

International Journal of Statistics and Probability; Vol. 9, No. 2; March 2020
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

An Algorithmic Approach to Modelling the Co-Evolution of Parasites and Their Hosts

Charles J. Mode

Correspondence: Charles J. Mode, Department of Mathematics, Drexel University, Philadelphia PA 1, USA

Received: January 17, 2020 Accepted: February 19, 2020 Online Published: February 26, 2020

doi:10.5539/ijsp.v9n2p13 URL: https://doi.org/10.5539/ijsp.v9n2p13

Abstract

This paper is a reformulation of the paper, Mode 1958 Evolution 12:158 - 165, which was written in terms of a deterministic paradigm, using differential equations In this paper, however, the working paradigm will be stochastic, and from the mathematical point of view, it will be a stochastic process that may be viewed as a branching process within a branching process. In particular, it will be assumed that the population of host plants will evolve as a multitype branching process, and the pathogen, which grows on the leaves of the host in every generation of the host, will also be assumed to evolve as a multitype branching processes during each generation of the host. The contents of this paper, were motivated by problems in Agriculture in which Plant Pathologists and Plant Breeders work together to control the damage inflicted by a pathogen on a growing crop of a cultivar such as flax, wheat. and many other cultivars. The focus of attention in this paper is the development of algorithms that will guide the development of software to run Monte Carol simulation experiments taking into account mutations in the host and pathogen. The writing of software to implement the algorithms developed in this paper would require a major effort, and is, therefore, beyond the scope of this paper.

Keywords: host, pathogen, genes for resistance in the host, genes to pathogenicity in the pathogen, mutations pathogenicity in the pathogen, mutations for resistance to the pathogen in the host, embedded deterministic model

1. A Stochastic Model Describing the Coevolution of an Obligate Parasite and Its Host

The model formulated in Mode (1958) was based on a system of differential equations that were used as a framework to describe the coevolution of an obligate parasite and its host. These equations were also used to suggest that the host and the parasite populations would coevolve to a state of equilibrium in which both the host and the pathogen coexist. During the 1950s, computing technology was very primitive when compared to technology that is available in the present era. But, nevertheless, some simple calculations were included in the(1958) paper to suggest that a host - pathogen system would indeed converge to a state of a mutual equilibrium in which the host and pathogen populations would coexist indefinitely. Two other papers on the dynamics of Host - Pathogen systems were also published. In Mode 1961, a deterministic formulation, which generalized the model in Mode 1958 was published. A stochastic model of a Host - Pathogen system was published in 1964. In this paper, the working paradigm belonged to a class of stochastic processes known as birth and death processes.

In this paper, however, the model that will be formulated to describe the coevolution of a host - parasite system will belong to a class of stochastic processes called branching processes. To illustrate the ideas, in spring of some year in the evolution of a host - pathogen system, let the symbols H_i for $i = 1, 2, \cdots, n$, denote the number $n > 0$ of host plants in the initial generation. For the sake of simplicity, it will be assumed that all host plants will be infected with the parasite or some other pathogen. It will also be assumed that the infection of a plant by the pathogen, starts with one spore which leads to a pustule that produces a spore, this spore in turn produces one or more spores and that this process continues until a leaf or leaves of a plant is fully covered by the pustules of the parasite. In the literature on stochastic processes, the type of multiplicative process of spores just described in known as a branching process.

Mathematically, a simple branching process may be described as follows. Let the random variable $X_0 = 1$ denote the initial pustule on a leaf , and let the random variable ξ, which takes values in the set $\mathbb{I} = \{0, 1, 2, \cdots\}$ of nonnegative integers, denote the number of pustules produce by any initial pustule. It will be assumed that the random variable $\boldsymbol{\xi}$ has a Poison distribution with positive parameter λ and probability density function

$$P[\boldsymbol{\xi} = \boldsymbol{x}] = f(x) = \exp[-\lambda] \frac{\lambda^x}{x!} \tag{1.1}$$

for $x = 0, 1, 2, 3. \cdots$. It is well known that the expectation and variance of a random variable with a Poisson distribution are $E[\xi] = \lambda$ and $var[\xi] = \lambda$. In the context of the model under consideration, the parameter λ may be interpreted as a measure of the virulence of the pathogen.

The infection process for any of the n plants will be formulated in terms of class of stochastic processes that are known

as a branching process. For the sake of simplicity, it will be assumed that an infection of any plant begins with one spore that which produces one pustule, Then, by assumption, the number of pustules produced by the first pustule is a random variable and is given by the equation

$$X_1 = \xi \,, \tag{1.2}$$

where $\boldsymbol{\xi}$ is a realization of the Poisson random variable. In general, let the random variable X_n denote the number of pustules on a plant in the *n-th* generation of the pathogen, and let $(\xi_k \mid k = 1, 2, \cdots)$ denote a sequence of independent random variables whose common distribution is that of the random variable ξ**.** Then, it follows that the number of pustules on a plant in generation 2 is

$$X_2 = \sum_{k=1}^{X_1} \xi_k \,. \tag{1.3}$$

And in general, this line of reasoning may be extended to conclude that

$$X_\nu = \sum_{k=1}^{X_{\nu-1}} \xi_k \,. \tag{1.4}$$

is the number of pustules on a plant for all generations $\nu = 1, 2, \cdots$ of the pathogen. For the case of flax rust, the generation time of the pathogen is about 10 days. In reality and in any computer simulation experiment, only some finite number $m > 0$ of generations of the pathogen will be considered.

From this equation, it can be seen that the conditional expectation of X_ν , given $X_{\nu-1}$, is

$$E\left[X_\nu \mid X_{\nu-1}\right] = X_{\nu-1}\lambda \tag{1.5}$$

for $\nu \geq 1$. Therefore, the unconditional expectation of the number of pustules in generation $\nu \geq 1$ on any plant satisfies the recursive equation

$$E\left[X_\nu\right] = E\left[X_{\nu-1}\right]\lambda \tag{1.6}$$

for $\nu \geq 1$. The algorithms just described may be used to program a computer to simulate that number of pustules on any of the n plants under consideration at the end of a season. In any computer experiment, it would be necessary to repeat the infection process just described for each of the n plants under consideration.

For any host plant H_i, let the random variable Y_i denote the total number of pustules on the plant at the end of the growing season. Then, the total load of pustules on the $n > 0$ plants under consideration is given by the equation

$$T = \sum_{k=1}^{n} Y_i \,. \tag{1.7}$$

The value of the random variable will influence that quantity and quality of the seeds produced by the plants. In extreme cases, the host plants may not be able to produce a sufficient number of viable seeds to produce the next generation of plants. Under such circumstances, the host-pathogen system would go extinct. But, even though the plants are infected with a pathogen, it may also happen that the plants produce a sufficient number of viable seeds to produce the next generation of plants. In such cases, the host - pathogen system would survive for another generation.

Given the a value of the random variable T, let $S\left(\nu\right)$ denote the conditional probability that in any generation a plant in generation $\nu \geq 1$ survives to produce plants in generation $\nu + 1$. If a plant is not infected by a pathogen, then it will be assumed that this conditional survival probability has the form

$$S\left(\nu\right) = \exp\left[-(\beta\nu)^2\right] \,. \tag{1.8}$$

If, however, a plant is infected with a pathogen, then it will be assumed that this survival probability has the form

$$S\left(\nu \mid T\right) = \exp\left[-(T\beta\nu)^2\right] \,. \tag{1.9}$$

Let n denote the number of plants at the end of a season. Then, it will be assumed that the number of plants that will survive and produce seeds in the next generation is a random variable W with a binomial distribution with sample size n and probability p. It is well known that the probability density function of the binomial distribution is

$$f_{BN}\left(x \mid n, p\right) = \binom{n}{x} p^x \left(1 - p\right)^{n-x} \tag{1.10}$$

for $x = 0, 1, 2, \cdots, n$.

In what follows, the notation $X \sim BN(n, p)$ will be used to denote that a random variable X has a binomial distribution sample size n and probability p. In any computer simulation experiment, a realization of the random variable W would be computed using the formula

$$W \sim BN(n, S(v \mid T)) \ , \tag{1.11}$$

and the realization of the random variable W would be the initial number of plants in the next generation. At this point in a simulation, the algorithms outlined above would be used to continue the Monte Carlo simulation experiment for another season.

As indicated in the title of this section, some obligate parasite was considered as the pathogen in the formulation above, but the model would could also be used for cases in which the pathogen may also grow in the soil or on media prepared in a laboratory.

2. Genetics of Host and Pathogen

The genetics of pathogenicity and host resistance to flax rust, Melampsora lini (Pers) Lev., has been reported in a pioneering works by Flor (1955 and 1956). Through a series of genetic studies, Flor has shown that the host and parasite possess complementary genetic systems. That is to say, any gene in the host for resistance acts if, and only if, there is a corresponding gene in the pathogen for avirulence. The genes for host resistance exist as a series of multiple alleles at five loci, designated as the K, L, M, N, and P in the genome of the host. There is one gene at the K locus, 11 at the L, six at the M, three at the N, and four at the P, for a total of twenty genes. The K, L, and M loci are inherited independently, but the N and P loci are linked with about 26 percent recombination. Unlike the genes for host resistance, the genes for pathogenicity in the pathogen exist at twenty-five separate loci.

Resistance of a host genotype to a particular pathogen genotype occurs whenever any allele in the host (at any one of the five multiple allelic loci) and its complementary gene for avirulence in the pathogen at any one of the twenty-five (diallelic loci) are present simultaneously. The alleles for host resistance are all dominant or semi-dominant so that the heterozygote as well as the homozygote are resistant. The genes for avirulence in the pathogen are also dominant, with the exception of one locus where the homozygous recessive in combination with one or two doses of a gene in the host is necessary for resistance. Further details on the genetics flax resistance to flax rust may be found in Flor (1955) and (1956).

These complementary genetic systems of the host and parasite are illustrated in the model presented below for the simple case of one locus in both the host and parasite with two alleles at each locus. Let R and r, respectively, denote alleles for resistance and susceptibility to the pathogen in the host, and let A and a denote alleles for avirulence and virulence, respectively, in the pathogen. In this model, there are three genotypes of the host, namely RR, Rr and rr. With respect to the pathogen there are also three genotypes: AA, Aa and aa. The interactions of the three genotypes of the host with three genotypes of the pathogen are illustrated in the table below.

Table 2.1. Interactions of the Genotypes of the Host and Pathogen

Genotypes	*RR*	*Rr*	*rr*
AA	*RIS*	*RIS*	*SUS*
Aa	*RIS*	*RIS*	*SUS*
aa	*SUS*	*SUS*	*SUS*

Observe that the columns of the table represent the genotypes of the host, and the rows of the table represent the genotypes of the pathogen. The symbols in the body of the table represent the interactions of the genotypes of the host and pathogen. For example in Table 2.1, the symbol RIS denotes resistance of the host to the pathogen, and the symbol SUS denotes susceptibility of the host to the pathogen. By inspecting the first row of this table, it can be seen that according to the model under consideration, the allele R in the host for resistance to the pathogen is assumed to be dominant to the recessive allele r for host susceptibility to the pathogen. The dominance effect of the allele A is also illustrated in row 2 of the table. In row 3 of the table, the susceptibility of each of the three genotypes of the host to the virulent genotype aa of the pathogen is indicated.

The simple illustrative case illustrating the gene to gene relationship in the host and pathogen considered in this section could in principle be extended to the case of one locus with multiple alleles. But, such an extension would require a more complex notation, and, therefore, will not be considered in this section. In the next section, however, cases in which the gene for gene relationship in the host and the pathogen will be considered with respect to two or more loci with multiple alleles in the host and pathogen.

Research on the genetics of host - pathogen systems is still an active field of research, but the focus of attention of current research differs from that of Flor. Currently, quite a number of researchers are studying the genetics of host - pathogen systems by sequencing the genomes of the host and parasite. Among those researchers, who are sequencing the gnomes of host and pathogens are Dr. Robert Brueggeman, Department of Plant Pathology, North Dakota State University. His current focus of attention is the sequencing the genomes of barley and one of its parasites, which is a fungus called net blotch. Two Ph.D. theses Boyle (2009) and Richards (2016) have also been devoted to research on net blotch.

Teams of researchers are working on sequencing the genomes of barley and net blotch. For example, the paper Wyatt et al. 2017 contains an account of the sequencing of the Pyrenophora teres f. teres Isolate 0-1, which the scientific name for net blotch. The paper Mascher et al. 2017 is devoted to an account of the sequencing of the barley genome and the paper by Tamang et al. 2019 is devoted to finding susceptibility/resistance to net blotch in the barley genome. In the paper Richards et al. 2016, there is an account of the fine mapping of the barley genome 6H net blotch susceptibility locus G3.. This brief review will provide the reader with some of the areas of genomic research that is currently in progress among scientists working on resistance of barley to net blotch, other plant species and pathogens.

3. Linkage of Genes in the Host for Resistance to Two Pathogens

It has been observed by plant geneticists during the past several decades that genes in the host for resistances to two or more pathogens are often linked, i.e. they are on the same chromosome in the host. An example of such linkage was reported in the paper Schaller and Briggs (1955) on genes for bunt resistance in wheat. Currently, one can find many reports on the internet on the location of genes for resistance in partially sequenced genomes of barley and wheat and their functions at the molecular level, but it is beyond the scope to this paper to review this more recent literature. The main thrust of this section is to focus on the formulation of a mathematical structure that accommodates linkage of genes for resistance in the host to two or more pathogens that may be used in computer simulation experiments.

In chapter 2 of the book Mode and Sleeman (2012), a mathematical structure for dealing with linkage of genes at two or more loci is described for diploid organisms. In that chapter a diploid genotype was represented by the symbol (x, y), where x is the allele contributed by the maternal parent and y is the allele contributed by the paternal parent. For the case of two alleles denoted by 0 and 1, four genotypes may be identified; namely, $(0, 0), (0, 1). (1, 0)$ and $(1, 1)$. Observe that in genotype $(0, 1)$ the maternal parent contributed the allele 0, but in genotype $(1, 0)$ the maternal parent contributed the allele 1.

It should be noted that not all cereal species that are important in agriculture are diploids. A plant species is said to be a polyploid, if during its evolution two or more genomes have been incorporated to make a single genome. For example, durum wheat and bread wheat are polyploids. Durum wheat (Triticum durum) for example, has 14 pairs of chromosomes. It is thought that during its evolution, two genomes that were originally 7 pairs of chromosomes, were joined to make a species of 14 pairs of chromosomes. Bread wheat (Triticum aestivum).has 21 pairs of chromosomes, and it is thought that the genome of this species is made of the genomes of three species with genomes of 7 pairs of chromosomes. Polyploidal species may not follow the same laws of inheritance than those of diploid species. Consequently, the linkage structure under consideration may not be applicable for polyploid species. Barley is an important spices in agriculture, and it has a genome made up of 7 pairs of chromosomes. Moreover, this species follows the laws of inheritance for diploid species, and therefore the linkage structure that will be developed will be applicable to experiments with barley.

In chapter 2, section 2 of Mode and Sleeman (2012), formulas are derived that accommodate linkage at two or more loci that can be applied in computer simulation experiments involving numerical calculations. In this section, a revision of some formulas will be derived for the case of two autosomal linked loci. In this case, an arbitrary genotype of diploid individual with respect to two linked loci may be represented in the form

$$\frac{00}{11}, \tag{3.1}$$

where 00 represents the alleles at two loci contributed by the females parent. Similarly, the symbol 11 denoted the alleles contributed by the male parent.

An individual of this genotype may in turn produce four types of gametes; namely $00, 01, 10$, and 11. Gametes 00 and 11 represent copies of the maternal and paternal gametes respectively; while 01 and 10 are recombination type gametes containing both maternal and paternal genes. In particular cases, the generic gametic symbols could be identified with particular haplotypes when the focus of attention is at the molecular level, or with phenotypes when the discussion is at the Mendelian level. It should also be noted that the Boolean notation under consideration is not phase dependent with respect to some specific genes at two loci to which attention is being directed, because in terms of Boolean indicators only parental origins of genes are under consideration.

The four types of gametes will be produced by a given genotype with certain probabilities depending on the probability

of recombination. Let $\gamma(00)$, $\gamma(01)$,$\gamma(10)$, and $\gamma(11)$ denote the probabilities an arbitrary genotype produces gametes 00, 01, 10, and 11, respectively. In the two loci case, the linkage (gametic) distribution is the set

$$(\gamma(00), \gamma(01), \gamma(10), \gamma(11)) \tag{3.2}$$

of non-negative numbers whose sum is one. Because of the complementary nature of the meiotic mechanism, i.e., gametes are almost always produced in pairs, the mechanism of meiosis will be called balanced if the equations

$$\begin{aligned} \gamma(00) &= \gamma(11) \\ \gamma(10) &= \gamma(01) \end{aligned} \tag{3.3}$$

are satisfied.

If we let ρ be the probability of recombination, then it follows that

$$\gamma(10) + \gamma(01) = \rho \tag{3.4}$$

and

$$\gamma(00) + \gamma(11) = 1 - \rho \ . \tag{3.5}$$

But, if the mechanism of meiosis is balanced, then

$$\gamma(01) = \gamma(10) = \frac{1}{2}\rho \tag{3.6}$$

and

$$\gamma(00) = \gamma(11) = \frac{1}{2}(1 - \rho) \ . \tag{3.7}$$

When a objective of an investigation is to map the locations of two loci on the same chromosome expressed in terms of centimorgens (cM), then the pertinent values of ρ will satisfy the condition $0 \leq \rho \leq \frac{1}{2}$. The case of random assortment occurs when $\rho = \frac{1}{2}$ so that the probability of each type of gamete is $1/4$. In principle, however, ρ may be any value such that $0 \leq \rho \leq 1$. Note if $0.5 < \rho \leq 1$, then $\gamma(01)$ and $\gamma(10)$ are greater than $\gamma(00)$ and $\gamma(11)$.

A problem of considerable theoretical importance in assessing the effects of linkage in populations is that of defining recombination probabilities and finding a relation between these recombination probabilities and the gametic distribution, when the number of loci under consideration is greater than two. An interesting step towards a solution of this problem was made by Schnell (1961), who observed that if one makes the transformation

$$\rho = \frac{1}{2}(1 - \lambda) \ , \tag{3.8}$$

or equivalently

$$\lambda = 1 - 2\rho \ , \tag{3.9}$$

then a certain orthogonality is introduced which leads to an extension to cases of an arbitrary number of linked loci. Observe that as ρ varies over the interval $[0, 1]$, the parameter λ varies over the interval $[-1, 1]$.

By using these equations, it can be seen that

$$\begin{aligned} \gamma(00) &= \frac{1}{4}(1 + \lambda) \\ \gamma(10) &= \frac{1}{4}(1 - \lambda) \ . \end{aligned} \tag{3.10}$$

These equations can perhaps be most easily comprehended in vector-matrix notation. Let

$$\begin{aligned} \gamma^T &= (\gamma(00), \gamma(10)) \\ \lambda^T &= (1, \lambda) \end{aligned} \tag{3.11}$$

be 1×2 vectors, where the superscript T stands for transpose of a vector or matrix, and let

$$\mathbf{A}_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \tag{3.12}$$

denote a 2×2 matrix. Observe that the rows and columns of this matrix is orthogonal. Then, in vector-matrix notation the above equation may be written in the form

$$\gamma = \frac{1}{4}\mathbf{A}_2\lambda \ . \tag{3.13}$$

This equation is of interest, because it expresses $\boldsymbol{\gamma}$ as a function of λ. In a computer simulation experiment, if the components of vector λ are assigned numerical values, then the vector $\boldsymbol{\gamma}$ is determined. From now on the symbol T will be used to denote the transpose of a matrix.

The following observations are very helpful in finding an extension to the case of an arbitrary number of linked loci. Firstly,

$$\mathbf{A}_2^T = \mathbf{A}_2 \tag{3.14}$$

so that $\mathbf{A}_2$ transpose is $\mathbf{A}_2$. When a square matrix remains invariant under the operation of transposition, it is said to be symmetric. Secondly, observe that

$$\mathbf{A}_2^T\mathbf{A}_2 = \mathbf{A}_2^2 = 2\mathbf{I}_2 \ , \tag{3.15}$$

where $\mathbf{I}_2$ is a 2×2 identity matrix., which has the form

$$\boldsymbol{I}_2 = \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) \tag{3.16}$$

Equation (3.15) expresses an orthogonality condition which, as we shall see, may be generalized to an arbitrary number of loci greater than two. Thirdly, if the symbols 00 and 10 are regarded as the vectors $\xi_1^T = (0, 0)$ and $\xi_2^T = (1, 0)$, then the matrix $\mathbf{A}_2$ may be represented in the form

$$\mathbf{A}_2 = \left(a_{ej}\right) = \left((-1)^{\xi_i^T\xi_j}\right) \ , \tag{3.17}$$

for $i, j = 1, 2$. Lastly, we see the column vector λ may be represented in the form

$$\lambda = 2\mathbf{A}_2\gamma \tag{3.18}$$

This equation could be taken as a definition of the column vector λ. In the next section, these results will be extended to the case of three linked loci.

4. Linkage of Genes for Resistance in the Host to Three Pathogens

In this section in order to make the paper self contained, a revised version of section 2.5 in Mode and Sleeman 2012 will be presented that will provide a structure to consider genes in the host for resistance to three pathogens. A question that naturally arises is whether the above scheme considered in section 3 can be generalized to any arbitrary number of loci greater than two. As we shall see such a generalization is possible, but the manner in which the scheme may be generalized will not become clear until the three loci case is considered. In the three loci case, an arbitrary diploid genotype may be represented in the form

$$\frac{000}{111} \ , \tag{4.1}$$

where as before the zeros and ones represent genes contributed by maternal and paternal parents, respectively. This genotype is capable of generating $2^3 = 8$ types of gametes containing various combinations of maternal and paternal genes.

The set of these eight types of gametes will be represented in the form

$$\mathbb{G} = (000, 100, 010, 110, 111, 011, 101, 001) \ , \tag{4.2}$$

and let the vector

$$(\gamma(\xi) \mid \xi \in \mathbb{G}) \tag{4.3}$$

denote the linkage distribution, *i.e.*, the set of non-negative numbers, whose sum is one, giving the probability that each type of gamete is produced by the meiotic process. Because the meiotic mechanism is assumed to be balanced, gametes are produced in pairs so that following symmetry or complementary conditions

$$\begin{aligned} \gamma(000) &= \gamma(111) \\ \gamma(100) &= \gamma(011) \\ \gamma(010) &= \gamma(101) \\ \gamma(110) &= \gamma(001) \end{aligned} \tag{4.4}$$

will be assumed. From these symmetry conditions, it follows that it will be sufficient to consider only four probabilities from the linkage distribution in setting up a correspondence with a set of recombination probabilities.

In the three loci case, a recombination probability may be associated with each pair of loci; namely, the pairs $(1,2),(1,3)$ and $(2,3)$. Let ρ_{12} , ρ_{13}, and ρ_{23}, denote the probability of recombination between the respective pairs of loci. With each recombination probability, we may associate a lambda parameter denoted by λ_{12}, λ_{13} and λ_{23}. From the definitions of ρ_{12} , ρ_{13}, and ρ_{23}, it follows that

$$\begin{aligned} \rho_{12} &= 2\left(\gamma(100)+\gamma(010)\right) \\ \rho_{13} &= 2\left(\gamma(100)+\gamma(110)\right) \\ \rho_{23} &= 2\left(\gamma(010)+\gamma(110)\right) \\ 1-\rho_{12} &= 2\left(\gamma(000)+\gamma(110)\right) \ . \end{aligned} \tag{4.5}$$

Observe that only four gametic probabilities appear on the right so it may be possible to solve four simultaneous linear equations. Also observe that the symmetry conditions were used in choosing the four gametic probabilities on the right.

By substituting the $\lambda' s$ for the $\rho' s$ in these equations, it can be shown that

$$\begin{aligned} \frac{1}{4}\left(1-\lambda_{12}\right) &= \gamma(100)+\gamma(010) \\ \frac{1}{4}\left(1-\lambda_{13}\right) &= \gamma(100)+\gamma(110) \\ \frac{1}{4}\left(1-\lambda_{23}\right) &= \gamma(010)+\gamma(110) \\ \frac{1}{4}\left(1+\lambda_{12}\right) &= \gamma(000)+\gamma(110) \ . \end{aligned} \tag{4.6}$$

By solving these four equations for $\gamma(000),\gamma(100),\gamma(010)$ and $\gamma(110)$, it can be seen that

$$\begin{aligned} \gamma(000) &= \frac{1}{8}(1+\lambda_{13}+\lambda_{23}+\lambda_{12}) \\ \gamma(100) &= \frac{1}{8}(1-\lambda_{13}+\lambda_{23}-\lambda_{12}) \\ \gamma(010) &= \frac{1}{8}(1+\lambda_{13}-\lambda_{23}-\lambda_{12}) \\ \gamma(110) &= \frac{1}{8}(1-\lambda_{13}-\lambda_{23}+\lambda_{12}) \ . \end{aligned} \tag{4.7}$$

Just as in the two loci case, these equations may be most easily comprehended if they are cast in vector-matrix notation. Let

$$\gamma^T = (\gamma(000),\gamma(100),\gamma(010),\gamma(110)) \tag{4.8}$$

and

$$\lambda^T = (1,\lambda_{13},\lambda_{23},\lambda_{12}) \tag{4.9}$$

denote row vectors, and let

$$\mathbf{A}_3 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} . \tag{410}$$

denote a 4×4 matrix. Then the equations may be written in the compact form

$$\gamma \ = \frac{1}{2^3}\mathbf{A}_3\lambda \ . \tag{4.11}$$

It should also be noted that the matrix $\mathbf{A}_2$ is related to the matrix $\mathbf{A}_3$ by the simple recursion formula

$$\mathbf{A}_3 = \begin{bmatrix} \mathbf{A}_2 & \mathbf{A}_2 \\ \mathbf{A}_2 & -\mathbf{A}_2 \end{bmatrix} . \tag{4.12}$$

From this recursive equation, it can be seen that

$$\mathbf{A}_3^T = \begin{bmatrix} \mathbf{A}_2^T & \mathbf{A}_2^T \\ \mathbf{A}_2^T & -\mathbf{A}_2^T \end{bmatrix} = \begin{bmatrix} \mathbf{A}_2 & \mathbf{A}_2 \\ \mathbf{A}_2 & -\mathbf{A}_2 \end{bmatrix} = \mathbf{A}_3 \tag{4.13}$$

because $\mathbf{A}_2$ is symmetric. Furthermore, from this equation, it follows that

$$\begin{aligned} \mathbf{A}_3^T\mathbf{A}_3 = \mathbf{A}_3^2 &= \begin{bmatrix} 2\mathbf{A}_2^2 & \mathbf{0} \\ \mathbf{0} & 2\mathbf{A}_2^2 \end{bmatrix} = 2\begin{bmatrix} 2\mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & 2\mathbf{I}_2 \end{bmatrix} \\ &= 2^2\mathbf{I}_4 \end{aligned} \tag{4.14}$$

where $\mathbf{0}$ is a 2×2 zero matrix and $\mathbf{I}_4$ is an identity matrix of order 4.

We thus see that orthogonality condition carries over to the three loci case. Moreover, let $\xi_1^T = (0,0,0), \xi_2^T = (1,0,0), \xi_3^T =$ $((0,1,0)$ and $\xi_4^T = (1,1,0)$ denote row vectors. Then, by inspection, it can be seen that

$$\mathbf{A}_3 = \left(a_{ij} = (-1)^{\xi_i^T \xi_j}\right) \tag{4.15}$$

for all $i, j = 1,2,3,4$. Just as in the case of two loci, it also follows, by using the above results and solving for the vector λ, that the equation

$$\lambda \ = 2\mathbf{A}_3\gamma \ , \tag{4.16}$$

expresses the vector λ as a linear function of the vector γ, given the matrix $\mathbf{A}_3$.

At this juncture it is important to note that the ordering of the gametic symbols

$$(000, 100, 010, 110) \tag{4.17}$$

played a basic role in extending the observations made in the case two loci to the case of three loci. Furthermore, the linear relation (4.16) , connecting the vectors λ and γ , suggests that to extend the results for the case of three loci to cases of four or more loci, it would be prudent to simplify the notation by abandoning subscripts on the λ and ρ parameters and replacing them by a function notation of the form $\lambda(\xi)$ and $\rho(\xi)$, where ξ is an arbitrary gametic symbol containing $0's$ and $1's$.

Given this function notation, the matrix $\mathbf{A}_3$ and the ordering of the elements of the gametic vector γ, equation (4.2) leads to an automatic ordering of the elements of the vector λ. This observation suggests that extensions of equation (4.1) to cases of four or more loci could be used to define the vector λ for an arbitrary number of loci. Then, given any $\xi \in \mathbb{G}$, a corresponding recombination probability may be determined by the equation

$$\lambda(\xi) = 1 - 2\rho(\xi) \ . \tag{4.18}$$

By way of an illustrative example, suppose $\xi = 000$, indicating that in gametes of this type there was no genetic recombination. For the case of three loci under consideration, it was observed that $\lambda(\xi) = 1$, which implies $\rho(\xi) = 0$, indicating with gametes of type $\xi = 000$ recombination occurs with probability 0, which is consistent with our intuition. In this paper, however, there will be no attempt to extend the results presented in this section to the cases of four or more loci, because it seems likely that demand for such cases will be very limited. If, however, a reader is interested in considering four or more linked loci, chapter 2 of the book Mode and Sleeman 2012 may be consulted for an account of working with four or more linked loci.

It would be of interest to present an example of data such that the probabilities of recombination for the case of three linked loci could be estimated, but no attempt to construct such an example will be undertaken. For readers who are interested in three point test for linkage, the book by Liu 1998 may be consulted.

5. Mutations in the Host and Pathogen

It is well known that mutations do occur in plant pathogens. For example, a variety of durum or bread wheat may have been grown extensively in an area for many years, because it's resistant to stem rust, a plant pathogen. However, it may happen that there is a mutation in the pathogen, stem rust, so that a variety of wheat that was resistant to the pathogen is no longer resistant, due to a mutation in the pathogen. As the science of genomics continues to develop, it may become possible to identity the location in the genome of the pathogen where the mutation occurred and characterize it at the molecular level. In this section, however, mutations in the host and pathogen will be described and analyzed mathematically with a view towards providing a framework for studying the interactions of hosts and pathogens as well as their coevolution.

Let A denote a gene in the pathogen for avirulence, and let a denote an allele for virulence. Similarly, let R let denote a gene in the host for resistance to the pathogen. and r denote an allele for susceptibility in the host . It will be assumed that the interactions of the host and pathogen follow those described in Table 2.1 in section 2. Let μ_{12} denote the probability per generation that gene A in the pathogen mutates to gene a for virulence. Similarly, let ν_{12} denote the probability per generation that gene R for resistance in the host mutates to gene r, which confers susceptibility of the host to the pathogen. If gene a in the pathogen mutates to gene A for avirulence, then, by definition, a back mutation has occurred.

The set of possible mutations in the pathogen may be represented in a two by two matrix

$$\mathbb{P} = \begin{bmatrix} \nu_{11} & \nu_{12} \\ \nu_{21} & \nu_{22} \end{bmatrix}. \tag{5.1}$$

In the first row of this matrix $\nu_{11} = 1 - \nu_{12}$ so that if $\nu_{12} = 10^{-9}$, then ν_{11}, which is the probability a mutation does not occur, is close to 1. Row 2 of this matrix, takes into account back mutations, which may occur with probability $\nu_{21} > 0$. In computer simulation experiments, the row of the matrix $\mathbb{P}$ are assigned numbers such that the rows of this matrix sum to 1. Similarly, mutations in the host may be represented by the matrix

$$\mathbb{H} = \begin{bmatrix} \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \end{bmatrix}, \tag{5.2}$$

row sum to 1. Observe that ν_{12} is the probability per generation that gene R in the host mutates to gene r. The probabilities in row 2 of the matrix $\mathbb{H}$ take into account that back mutations may occur in the host.

It seems likely that before wheats, barley, flax and other plants that were domesticated, they were outbreeders so that in an evolving population random mating was prevalent, because winds and perhaps insects would distribute pollen randomly among plants. According to this view, as these species of plants were domesticated they evolved to a state in which they reproduced by self fertilization. In this situation, the anatomy of a flowering plant evolved in such a way that the pollen fertilizes only its eggs in same flower. Consequently, when formulating models describing the evolution of these species of plants, it will be necessary to take into account not only random mating but also reproduction by self fertilization. in models dealing with the joint evolution of the host and pathogen. It will be assumed that the pathogen reproduces asexually. That is individuals of type A produce only individuals of type A if a mutation does not occur. Similar remarks hold for individuals of type a of the pathogen.

A computer experiment that would mimic that rise of virulent mutations in the pathogens and their impact on populations of domestic wheats, barley and other crops would be of practical interest, because they are actually observed in such populations. Bread and durum wheats are polyploids, but with respect to one locus in which genes for resistance or susceptibility to a pathogen are located, the genetic laws governing these traits are similar to those of diploids. The laws governing the resistance to pathogens in barley follow those of a diploid population. Consider a host pathogen system in which pathogen is haploid and reproduces asexually. It will be assumed that the pathogen has two alleles A for avirulence and a for virulence. As is section 2, let R denote a allele in the host for resistance to the pathogen and let r denote at the same locus for susceptibility to the pathogen. In this model, the host population would consist of three genotypes RR. Rr and rr. In the absence of mutation, when a population reproduces either by random mating or selfing, individuals of genotype RR would produce offspring only of genotype RR. But, is such population individuals of genotype Rr would produce offspring with genotypes RR, Rr and rr with corresponding probabilities 0.25, 0.50, and 0.25. Similarly, in the absence of mutation, individuals of genotype rr would produce offspring only of genotype rr.

Mutations are rare events and thus need to be taken into account in a model according the laws of evolution of a stochastic processes. For example, with regard to the pathogen population, let M denote the number of individuals of genotype A in the population in some generation , and let m denote the number of mutations from genotype A to the virulent genotype a. If these mutations occur independently, then it follows that the random variable m has a binomial distribution with the probability density function

$$P[m = x] = f(x) = \binom{M}{x} \mu_{12}^{x} (1 - \mu_{12})^{M-x} \tag{5.3}$$

for $x = 0, 1, 2, \cdots, M$. When M is large, then the probability density function in 5.3 may be approximated by a Poisson distribution with parameter $M\mu_{12}$.

As a first step in formulating a model in a host population with respect to one locus, consider a population of plants that are homozygous for a gene R for resistance to some pathogen. In this case, all plants in a population would be of genotype RR. It will be assumed that the population reproduces by selfing and that mutations may occur. Specifically, it will be assumed that allele R may mutate to allele r with probability ν_{12} per generation, and that allele r may back mutate to allele

R with probability ν_{21} per generation, see 5.2. Given a population of individuals of genotype RR, it will be necessary to describe the set of events that may occur when mutation occurs in the evolution of a population. In particular, under the assumption that mutations may occur, in what follows formulas will be derived for the distribution of the three genotypes RR, Rr and rr among the offspring of parents with genotypes RR, Rr and rr.

Observe that under the assumption that copies of allele R mutate independently, it follows that

$$(1-\mu_{12})\times(1-\mu_{12}) = (1-\mu_{12})^2 \tag{5.4}$$

is the probability per generation that neither of the copies of the allele R in the genotype RR mutate to allele r. Suppose that among the offspring of a plant of genotype RR an offspring of genotype Rr is found, indicating that a mutation of the form $R \longrightarrow r$ has occurred, This mutation can occur in two ways. Suppose the left allele in the genotype RR does not mutate with probability $(1-\mu_{12})$, but the right allele does mutate with probability μ_{12}. In this case, the probability of the occurrence of the mutant genotype Rr is $(1-\mu_{12})\mu_{12}$. It can be seen by the same line of reasoning, that the probability that that left allele in the genotype RR mutates but the right allele dose not mutate is $\mu_{12}(1-\mu_{12})$. The two mutational events just described are disjoint. Therefore, the probability of finding an offspring of genotype Rr among the offspring of an individual of genotype RR is

$$(1-\mu_{12})\mu_{12} + \mu_{12}(1-\mu_{12}) = 2\mu_{12}(1-\mu_{12}). \tag{5.5}$$

The genotype rr may also be occur among the offspring of individuals of genotype RR. In this case, both the left and right alleles in the genotype RR have mutated to the allele r. The probability of this rare event is

$$\mu_{12}^2 . \tag{5.6}$$

The next step in formulating the model of mutation under consideration is to derive formulas for the probabilities of finding mutant genotypes among the offspring of individuals of genotype Rr. The probability that neither of the alleles in the genotype mutate is

$$(1-\mu_{12})(1-\mu_{21}) . \tag{5.7}$$

The probability that the right allele r in the genotype Rr back mutates to allele R is ν_{21}. Therefore, the probability of finding an individual of genotype RR among the offspring of an individual of genotype Rr is

$$(1-\mu_{12})\mu_{21}. \tag{5.8}$$

Similarly, the probability of finding an individual of genotype rr among the offspring of an individuals of genotype Rr is

$$\mu_{12}(1-\mu_{21}) . \tag{5.9}$$

The last step in the process of deriving formulas for finding mutant genotypes among the offspring of individuals of genotype rr. The probability that there are no mutant offspring among the offspring of an individual of genotype rr is

$$(1-\mu_{21})^2 . \tag{5.10}$$

The probability of finding an individual of genotype Rr among the offspring on an individual of genotype rr is

$$2\mu_{21}(1-\mu_{21}) . \tag{5.11}$$

Note that the argument used to derive this formula is the same as that used in deriving formula 5.5. Finally, the probability of finding an offspring of genotype RR among the offspring of an individual of genotype rr is

$$\mu_{21}^2 . \tag{5.12}$$

The next step towards reaching the goal stated above is to derive a formula the probabilities of finding genotypes RR, Rr and rr among the offspring an individual of the genotype RR. Let TOR_{RR} denote the number of offspring on an individual of genotype RR. From 5.4, 5.5 and 5.6, it can be seen that the total probability of finding non-mutant as well as mutant genotypes among the offspring of an individual of genotype RR is

$$P[TOT_{RR}] = (1-\mu_{12})^2 + 2\mu_{21}(1-\mu_{21}) + \mu_{12}^2 \tag{5.13}$$

As an aid in deriving formulas for the probabilities of finding mutant offspring *RR*, *Rr* and *rr* among the offspring of genotype *RR* under the assumption mutation, a random variable X_{RR} will be introduced to indicate the genotype of an offspring of genotype *RR*. Therefore, the conditional probability of finding an offspring of genotype *RR* among those of genotype *RR* is

$$P\left[X_{RR} = RR \mid TOT_{RR}\right] = \frac{(1-\mu_{12})^2}{(1-\mu_{12})^2 + 2\mu_{21}(1-\mu_{21}) + \mu_{12}^2} \tag{5.14}$$

Similarly,

$$P\left[X_{RR} = Rr \mid TOT_{RR}\right] = \frac{2\mu_{21}((1-\mu_{21}))}{(1-\mu_{12})^2 + 2\mu_{21}(1-\mu_{21}) + \mu_{12}^2} \tag{5.15}$$

and

$$P\left[X_{RR} = rr \mid TOT_{RR}\right] = \frac{\mu_{12}^2}{(1-\mu_{12})^2 + 2\mu_{21}(1-\mu_{21}) + \mu_{12}^2}\ . \tag{5.16}$$

Let X_{Rr} denote a random variable indicating the genotype of an offspring of an parental individual of genotype *Rr*. Observe that

$$TOT_{Rr} = (1-\nu_{12})(1-\nu_{21}) \ + (1-\nu_{12})\nu_{21} + (1-\nu_{21})^2\ . \tag{5.17}$$

Therefore, from 5.8 it follows that

$$P\left[X_{Rr} = RR \mid TOT_{Rr}\right] = \frac{(1-\mu_{12})\mu_{21}.}{(1-\mu_{12})(1-\mu_{21}) \ + (1-\mu_{12})\mu_{21} + (1-\mu_{21})^2}\ . \tag{5.18}$$

Similarly, from 5.9 it can be seen that

$$P\left[X_{Rr} = Rr \mid TOT_{Rr}\right] = \frac{2\mu_{21}(1-\mu_{21})}{(1-\mu_{12})(1-\mu_{21}) \ + (1-\mu_{12})\mu_{21} + (1-\mu_{21})^2} \tag{5.18}$$

$\mu$and from 5.10 it can be see that

$$P\left[X_{Rr} = rr \mid TOT_{Rr}\right] = \frac{\mu_{12}(1-\mu_{21})}{(1-\mu_{12})(1-\mu_{21}) \ + (1-\mu_{12})_{21} + (1-\mu_{21})^2} \tag{5.19}$$

For the case of finding genotypes *RR*, *Rr* and *rr* among the offspring of an individual of genotype *rr*, the total probability of a mutation is

$$TOT_{rr} = \mu_{21}^2 + 2\mu_{21}(1-\mu_{21}) + (1-\mu_{21})^2 \tag{5.20}$$

Therefore,

$$P\left[X_{rr} = RR \mid TOT_{rr}\right] = \frac{\mu_{21}^2}{\mu_{21}^2 + 2\mu_{21}(1-\mu_{21}) + (1-\mu_{21})^2}\ . \tag{5.21}$$

By using the same line of reasoning, it follows that

$$P\left[X_{rr} = Rr \mid TOT_{rr}\right] = \frac{2\mu_{21}(1-\mu_{21})}{\mu_{21}^2 + 2\mu_{21}(1-\mu_{21}) + (1-\mu_{21})^2} \tag{5.22}$$

and

$$P\left[X_{rr} = rr \mid TOT_{rr}\right] = \frac{(1-\mu_{21})^2}{\mu_{21}^2 + 2\mu_{21}(1-\mu_{21}) + (1-\mu_{21})^2}\ . \tag{5.23}$$

The expressions in the formulas derived above will be interpreted as elements in three dimensional 1×3 vectors that are defined below.

$$\begin{aligned} p_{RR} &= \left(P\left[X_{RR} = RR \mid TOT_{RR}\right], P\left[X_{RR} = Rr \mid TOT_{RR}\right], P\left[X_{RR} = rr \mid TOT_{RR}\right]\right) \\ p_{Rr} &= \left(P\left[X_{Rr} = RR \mid TOT_{Rr}\right], P\left[X_{Rr} = Rr \mid TOT_{Rr}\right], P\left[X_{Rr} = rr \mid TOT_{Rr}\right]\right) \\ p_{rr} &= \left(P\left[X_{rr} = RR \mid TOT_{rr}\right], P\left[X_{rr} = Rr \mid TOT_{rr}\right], P\left[X_{rr} = rr \mid TOT_{Rr}\right]\right) \end{aligned} \tag{5.24}$$

In the next section., it will be shown that these three vectors play an essential role in Monte Carlo simulation experiments designed to study the occurrence of virulent genotypes in the pathogen and genes for susceptibilities in the host that arise the host and pathogen populations by the process of mutation.

6. Evolution of the Host Population as a Multitype Branching Processes

In this section, the formulation of a mulitype branching process evolving on a time scale of discrete generations will be given along with a description of algorithms to compute a sample of Monte Carlo realizations of such a multitype branching processes. The class of multitype branching processes considered in this section, is an extension of the one type branching process described in section 1. To illustrate the algorithms underlying a Monte Carlo simulation procedures to simulate a sample of realizations of this stochastic process, for the sake of simplicity, attention will be focused on the case of $m = 3$ types, which will be referred as genotypes. Let $\mathbb{G} = \{1, 2, 3\}$ denote the set of three genotypes. For the case of the host with one locus with two alleles R and r, these three genotypes would be RR, Rr and rr as illustrated in section 2.

The number of offspring will be characterized in terms of random variables N_ν for $\nu \in \mathbb{G}$, taking values in the set $\boldsymbol{I} = \{n \mid n = 0, 1, 2, 3, \cdots\}$ of non-negative integers. The probability density functions of these random variables will be denoted by

$$P[N_\nu] = g_\nu(n) \text{ for } n \in \boldsymbol{I} \tag{6.1}$$

and $\nu \in \mathbb{G}$. The expected value of a random variable N_ν is

$$E[N_\nu] = \lambda_\nu \geq 0 \tag{6.2}$$

and may be interpreted as the average number of offspring contributed to the next generation by each genotype $\nu \in \mathbb{G}$.

As in the foregoing sections of this paper, the probability density functions for the random variables N_ν, $\nu \in \mathbb{G}$, will be chosen as the simple Poisson densities

$$g_\nu(n) = \exp[-\lambda_\nu] \frac{\lambda_\nu^n}{n!} \tag{6.3}$$

for $n \in \mathbb{N}$ and $\nu \in \mathbb{G}$. It is easy to show that for the density in (6.3), $E[N_\nu] = \lambda_\nu > 0$ so that the measure of reproductive success λ_ν is the parameter for a Poisson density for each genotype. In the experiments reported in this paper, a decision was made to consider only the special case of Poisson distributions for the random variables N_ν for $\nu = 1, 2, 3$. It is well known for this simple distribution that the expectation and variance both equal the parameter λ.

A multitype branching process in discrete time may be defined as follows. In generation t, where $t = 0, 1, 2, 3, \cdots$, let the random function $X_i(t)$, taking values in the set $\mathbb{N}$, denote the number of individuals of genotype $i \in \mathbb{G}$ in generation t, and let

$$\boldsymbol{X}(t) = (X_1(t)), X_2(t), X_3(t) \tag{6.4}$$

denote a vector of these random functions. For $t = 0$, an experimenter needs to assign initial values in the set $\mathbb{N}$ to each of the elements in the vector (6.4). Let the 1×3 vector $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, Y_{i3})$ denote the total number of offspring each genotype produced by genotype $i \in \mathbb{G}$ in any generation. For each genotype i, let

$$Y_{i1}, Y_{i2}, \cdots \tag{6.5}$$

denote a sequence of conditionally independent random variables, given $X_i(t)$ the number of individuals of genotype i in generation t.Each of the random variables in this sequence has a Poisson distribution with parameter λ_i. Then, the number of individuals in the population at time $t + 1$ of genotype i is

$$X_i(t + 1) = \sum_{\nu=1}^{X_i(t)} Y_{i\nu} \tag{6.6}$$

for $i = 1, 2, 3$.

The next step in the formulation of the model is to take into account mutation. To include mutations in the model, it will be necessary to introduce the multinomial distribution. The probability density function of the multinomial distribution for the case of three dimensions is

$$f(z_1, z_2, z_3) = \frac{n!}{z_1! z_2! z_3!} p_1^{s_1} p_2^{z_2} p_3^{z_3} \tag{6.7}$$

where

$$z_1 + z_2 + z_3 = n \tag{6.8}$$

and

$$p_1 + p_2 + p_3 = 1 \tag{6.9}$$

and $p_i > 0$ for $i = 1, 2, 3$.

Let p_{RR} denote the 1×3 row vector defined in 5.24.Then for an individual of genotype RR,when mutations occur the number of offspring of each of the three genotypes, RR, Rr and rr,is given by the 1×3 row vector OFF_{RR}, which is a realization of a multinomial distribution with probability vector p_{RR} and sample size $n = X_i(t+1)$, with genotype $i = RR$. For the case of genotype Rr, the number of offspring of each genotype is given by the 1×3 vector OFF_{Rr}, which is a realization of a multinomial distribution with probability vector p_{Rr} and sample size $n = X_i(t+1)$ with $i = Rr$. Lastly, for the case of genotype rr the number of offspring of the three genotypes is given by the 1×3 vector OFF_{rr}, which is a realization of a multinomial distribution with a 1×3 probability vector p_{rr} and sample size $n = X_i(t+1)$ with $i = rr$. An algorithm for simulating a realization of a multinomial random vector may be found in the paper by Mode and Gallop 2008.

It will be helpful to arrange the realization of the three 1×3 vectors in the form of a 3×3 matrix of the form

$$\begin{bmatrix} off_{11} & off_{12} & off_{13} \\ off_{21} & off_{22} & off_{23} \\ off_{31} & off_{32} & off_{33} \end{bmatrix} \tag{6.10}$$

when it is necessary to compute the number of individuals of each genotype when mutations occur in a population. For example, when mutations are taken into account, the number of individuals of genotypes RR is at time $t+1$

$$X_{RR}(t+1) = \sum_{j=1}^{3} off_{j1} \, . \tag{6.11}$$

Similarly, when mutations are taken into account the numbers of individuals in the population of genotypes Rr and rr are given by the sums

$$X_{Rr}(t+1) = \sum_{j=1}^{3} off_{J2} \tag{6.12}$$

and

$$X_{rr}(t+1) = \sum_{j=1}^{3} off_{J3} \tag{6.13}$$

7. Embedding a Deterministic Model in a Stochastic Process

Let $Z_1, Z_2, \cdots$ denote a sequence of random variables taking values in the set of real numbers $R = (-\infty.\infty)$, and suppose the expectation $E\left[Z_i^2\right]$ is finite for every $i = 1, 2, \cdots$. It is of interest to find the best estimator of the random variable Z_{n+1}, given the random variable Z_n. Let $\widehat{Z}_{n+1}$ denote this estimator. Then it is well known that the conditional expectation

$$\widehat{Z}_{n+1} = E\left[Z_{n+1} \mid Z_n\right] \tag{7.1}$$

minimizes the expectation

$$E\left[\left(Z_{n+1} - \widehat{Z}_{n+1}\right)^2\right] \, . \tag{7.2}$$

Thus $\widehat{Z}_{n+1}$ in 7.1 is the best estimator of Z_{n+1} in the sense of least squares. In what follows, conditional expectations of the form 7.1 will be used extensively .

The objective of this section is to describe a procedure for embedding a deterministic model in a stochastic process. As a first step in describing the derivation of the embedded model, consider the Galton-Watson process defined by equation 1.4 in section 1 by the equation

$$X_\nu = \sum_{k=1}^{X_{\nu-1}} \xi_k \tag{7.3}$$

for $\nu = 1, 2, \cdots$. In this case, for a given ν

$$E[X_\nu \mid X_{\nu-1}] = E[X_{\nu-1}]\lambda \tag{7.4}$$

because for each k it is assumed that the random variable ξ_k has a Poison distribution with expectation λ. As indicated above, let

$$\widehat{X}_i = \widehat{X}_{i-1}\lambda \tag{7.5}$$

for $\nu = 1, 2, \cdots$.

For $\nu = 1$ equation 7.5 takes the form

$$\widehat{X_1} = E[X_0]\lambda\,, \tag{7.6}$$

where $E[X_0] = X_0$, the initial size of the population, which is an assigned number. Thus,

$$\widehat{X_1} = X_0\lambda\,. \tag{7.7}$$

is known. From equations 7.5 and 7.7, it follows that

$$\widehat{X_2} = E[X_2 \mid X_1] = \widehat{X_1}\lambda\,. \tag{7.8}$$

By using the process just described, it can be shown that

$$\widehat{X_k} = \widehat{X}_{k-1}\lambda \tag{7.9}$$

for $k \geq 1$. By definition, equation 7.9 is the deterministic model embedded in the Galton-Watson process.

For the case of a multitype branching process described in section 6, the basic equation of the process is

$$X_i(t+1) = \sum_{\nu=1}^{X_i(t)} Y_{i\nu} \tag{7 .10}$$

for genotypes $i = 1, 2$ and 3, where given $X_i(t)$, $Y_{i1}, Y_{i2}, \cdots$ is a sequence of conditionally independent Poisson random variables for expectation λ_i for each genotype $i = 1, 2, 3$. For each genotype i,it can be shown that

$$E[X_i(t+1) \mid X_i(1)] = X_i(t)\lambda_i\,. \tag{7.11}$$

Therefore, just as in the case of a one type branching process, the recursive equation

$$\widehat{X_i}(t+1) = \widehat{X_i}(t)\lambda_i \tag{7.12}$$

will be used to calculate estimates of the sample functions for each genotype $i = 1, 2, 3$ and $t = 0, 1, 2, \cdots$.

To take mutation into account in the embedded deterministic model, it will be necessary the use the modified matrix

$$\begin{bmatrix} \widehat{off_{11}} & \widehat{off_{12}} & \widehat{off_{13}} \\ \widehat{off_{21}} & \widehat{off_{22}} & \widehat{off_{23}} \\ \widehat{off_{31}} & \widehat{off_{32}} & \widehat{off_{33}} \end{bmatrix}, \tag{7.13}$$

which was defined in section 6 in equation 6.10. The symbol $\widehat{off}$ indicated that the elements of the matrix have been calculated using estimates of the sample functions of the process. In section 6, the elements of the matrix in 7.13 were calculated as realizations of multinomial random vectors. But, in the embedded deterministic model the rows of the desired matrix, will be calculated as the mean or expectation of a vector with a multinomial distribution.

Consequently, the modified version of the matrix in 7.13 has the form

$$\begin{bmatrix} \widehat{X_1}\widehat{off_{11}} & \widehat{X_1}\widehat{off_{12}} & \widehat{X_1}\widehat{off_{13}} \\ \widehat{X_2}\widehat{off_{21}} & X_2\widehat{off_{22}} & \widehat{X_2}\widehat{off_{23}} \\ \widehat{X_3}\widehat{off_{31}} & \widehat{X_3}\widehat{off_{32}} & X_3\widehat{off_{33}} \end{bmatrix} \tag{7.14}$$

In this equation, $\widehat{X_1}$ is actually $\widehat{X_1}(t+1)$, but for the sake of simplicity, the symbol $(t+1)$ was not shown in matrix 7.14. Given this matrix, the estimates of the number of each of the three genotypes at time $t+1$ are as follows:

$$\begin{aligned} \widehat{X_1}(t+1) &= \sum_{\nu=1}^{3} \widehat{X_\nu}\widehat{off_{\nu 1}} \\ \widehat{X_2}(t+1) &= \sum_{\nu=1}^{3} \widehat{X_\nu}\widehat{off_{\nu 2}} \\ \widehat{X_3}(t+1) &= \sum_{\nu=1}^{3} \widehat{X_\nu}\widehat{off_{\nu 3}} \end{aligned} \tag{7.15}$$

8. Evolution of the Pathogen Population as a Multitype Branching Process

For the sake of simplicity, it will be assumed that the pathogen is a haploid so that each individual of the pathogen population has only one copy of a gene at each locus. In what follows the genotype of an individual in the pathogen population will be denoted by the symbols A and a, which denote avirulence and virulence respectively. According the mutation matrix in 5.1, gene A of the pathogen mutates to gene a with probability ν_{12} per generation, and gene a mutates back to gene A with probability ν_{21} per generation.

At time t let the random variable $W_1(t)$ denote the number of individuals of genotype A in the pathogen population, and let the random variable $W_2(t)$ denote the number of individuals of genotype a in the pathogen population at time t. Let $\xi_i(j)$ for $j = 1, 2, \cdots$ denote a sequence of conditionally independent Poison random variables with parameter λ_i, where $i = 1$ or $i = 2$ denotes the genotype of an individual in the pathogen population. Then the evolution of the number of individuals of genotype $1 = A$ in the pathogen populate is governed by the recursive equation

$$W_1(t+1) = \sum_{j=1}^{W_1(t)} \xi_j(1) \ , \tag{8.1}$$

for $t = 0, 1, 2, \cdots$. If $W_1(t) = 0$, then $W(t+1) = 0$ for all t. Similarly, the evolution of the number of individuals in the pathogen population of genotype $2 = a$ has the form

$$W_2(t+1) = \sum_{j=1}^{W_2(t)} \xi_j(2) \ , \tag{8.2}$$

for $t = 0, 1, 2, \cdots$. To initialize the recursive procedures in equations 8.1 and 8.2, the number of individuals of each genotype in the initial population, $W_1(0)$ and $W_2(0)$, must be specified by an experimenter. For each j the expection of the Poison random variables are $E\left[\xi_j(1)\right] = \lambda_1$ and $E\left[\xi_j(2)\right] = \lambda_2$.Because genotype 2 is a, the gene for virulence, the values of λ_1 and λ_2 chosen by an experimenter must satisfy the inequality $\lambda_1 < \lambda_2$ to take into account that virulent genotypes of the pathogen will have many more offspring per generation that an avirulent genotype.

The next step in the formulation of the pathogen process is to take into account that mutations may occur in populations avirulent individuals that are of genotype A. Mutations in individuals of the virulent genotype a will also be included in the formulation. According to the 2×2 matrix of mutation probabilities in 5.1 for the pathogen, the mutation probabilities for genotype A are ν_{11} and ν_{12}, and those for genotype a are ν_{21} and ν_{22}. Let $OA(t) = (n_{A1}(t), n_{a1}(t))$ denote a 1×2 vector of the numbers of genotypes A and a that are offspring of genotype A at time $t+1$. Then, by assumption, the vector $OA(t+1)$ has a multinomial distribution with sample size $W_1(t+1)$ and probability vector $p_A = (\nu_{11}, \nu_{12})$ at time $t+1$ in the evolution of the pathogen process. Similarly, let the 2×1 vector $Oa(t+1) = (n_{A2}(t+1), n_{a2}(t+1))$ denote the number of offspring of genotype a with genotypes A and a. By assumption this vector also has a multinomial distribution with sample size $W_2(t)$ and probability vector $p_a = (\nu_{21}, \nu_{22})$.

To complete the derivation of the process for the pathogen in any generation, it will be helpful for arrange the two vectors just derived in the 2×2 matrix

$$\begin{bmatrix} n_{A1}(t+1) & n_{a1}(t+1) \\ n_{A2}(t+1) & n_{a2}(t+1) \end{bmatrix} . \tag{8.3}$$

From this matrix it can be seen that, when mutation is taken into account, the number of individuals of genotype A in the population at time $t+1$ is

$$W_1(t+1) = \sum_{j=1}^{2} n_{Aj}(t+1) \ , \tag{8.4}$$

and the number of individuals of genotype a in the population at time $t+1$ is

$$W_2(t+1) = \sum_{j=1}^{2} n_{aj}(t+1) \tag{8.5}$$

By using procedures similar to those in section 7, a deterministic model may be embedded in the pathogen process under consideration. For example, the recursive deterministic model corresponding to equation 8.1 is

$$\widehat{W}_1(t+1) = \widehat{W}_1(t)\lambda_1 \tag{8.6}$$

for $t = 0, 1, 2, \cdots$. Similarly, the recursive deterministic model corresponding to equation 8.2 is

$$\widehat{W}_2(t+1) = \widehat{W}_2(t)\,\lambda_2\,. \tag{8.7}$$

To include mutation in the embedded deterministic model, the probabilities in the mutation matrix

$$\begin{bmatrix} \nu_{11} & \nu_{12} \\ \nu_{21} & \nu_{22} \end{bmatrix} \tag{8.8}$$

will play a fundamental role. The estimated expectation matrix for the numbers of mutations for the embedded deterministic model is

$$\begin{bmatrix} \widehat{W}_1(t+1)\,\nu_{11} & \widehat{W}_1(t+1)\,\nu_{12} \\ \widehat{W}_2(t+1)\,\nu_{21} & \widehat{W}_2(t+1)\,\nu_{22} \end{bmatrix}. \tag{8.9}$$

Therefore, it follows that the estimated number of individuals of genotype 1 in the pathogen population at time $t+1$ is

$$\widehat{W}_1(t+1) = \sum_{k=1}^{2} \widehat{W}_k(t+1)\,\nu_{k1} \tag{8.10}$$

Similarly, the estimated number of individuals of genotype 2 in the pathogen population at time $t+1$ is

$$\widehat{W}_2(t+1) = \sum_{k=1}^{2} \widehat{W}_k(t+1)\,\nu_{k2}\,. \tag{8.11}$$

9. Generation Times of Small Grains, Other Cultivars, Their Pathogens and Balanced Polymorphisms

An essential ingredient of an evolutionary model of small grains, such as wheat and barley, is their generation times. By definition, a generation time, is the time from the germination of seeds to the time that the mature plants produce seeds. For those varieties of wheat and barley that are planted in the spring in northern temperate regions of the earth, the generations times of these species are in the range of 90 for 100 days. The generation times for varieties of flax in northern temperate regions also are in the range of 90 to 100 days. When these species are grown in the southern temperate regions of the earth, the generation times of the species under consideration are essentially the same.

The generation times of the pathogens, which vary among pathogens, are usually much shorter than those of the host. In the computer experiments reported in this paper, it will be assumed that the generation time of a pathogen is 10 days. Under this assumption, for every generation of the host, there are about 10 generations of the pathogen that will be simulated. Mutations in the host and pathogen often occur during the reproduction process in connection with the copying of *DNA* in both the host and pathogen when cells divide. Because for every generation of the host, there will be 10 generations of the pathogen, it is more likely that during every growing season, there many more mutations in the pathogen than the host.

After the seeds sprout in the hosts under consideration, there is a rapid increase in biomass, in the forms of leaves, stems and structures connected with reproduction, such as heads that contain the seeds in wheat and barley and structures that contain the seeds in flax. To model the process of plant growth after germination would technically be a very difficult. Consequently, no attempt will be made to model plant growth in this paper. But, it will be tacitly assumed that plant grow does occur in computer simulation experiments and provides a medium on which a pathogen grows and damages the plant. The procedures described in section 1 will be used in all computer simulation experiments to quantify the damage the pathogen does to the host in each generation of a computer experiment.

The term, balanced polymorphisms, is used in evolutionary biology. From the mathematical point of view, it refers to the structure that results form the convergence of a Host-Pathogen model to a limit. For the case of a deterministic model the so called balanced polymorphism will be the constant structure that is the limit of the solution of deterministic equations, describing the evolution of a host-pathogen system as $t \longrightarrow \infty$ when it exits. For the case of a stochastic model describing the evolution of a host-parasite population, in some formulations, such as a Markov processes, the model will converge in distribution to what is called a quasi-stationary distribution. If a reader is interested of examples in which a stochastic model converges to a quasi-stationary distribution, the book by Mode and Sleeman (2012) may be consulted. For a so called quasi - stationary distribution, the mean and variance of the distribution remains constant for all t after convergence. In the case of a stochastic model, the quasi-station distribution may be referred to as a balanced polymorphism. The models described in the foregoing sections are Markov processes.

References

Flor, H. H. (1955). Host-parasite interaction in flax rust-its genetics and other implications. *Phytopath., 45,* 680-685.

Flor, H. H. (1956). *The complementary genic systems in flax and flax rust.* In Advances in Genetics, New York: Academic Press, Inc. https://doi.org/10.1016/S0065-2660(08)60498-8

Liu, B. H. (1998). *Statistical Genomics, Linkage, Mapping and QTL Analysis.* CRC Press, Boca Raton and New York.

Liu, Z., Holmes, D., Faris, J. D., Chao, S., Brueggeman, R. S., Edwards, M. C., & Friesen, T. L. (2015). *QTL mapping reveals effector-triggered susceptibility underlying the barley-Pyrenophora teres f. teres interaction.* Molecular Plant Pathology.

Mascher, M., Gundlach, H., & Himmelbach, A. (2017). A chromosome conformation capture ordered sequence of the barley genome. Nature 544: 427-433.

Mode, C. J. (1958). A mathematical model for the co-evolution of obligate parasites and their hosts. *Evolution, 12*, 158-165. https://doi.org/10.2307/2406026

Mode, C. J. (1961). A Generalized Model of a Host-Pathogen System. *Biometrics, 17,* 386-403. https://doi.org/10.2307/2527833

Mode, C. J. (1964). A Stochastic Model of the Dynamics of Host-Pathogen Systems With Mutation. *The Bulletin of Mathematical Biophysics, 26,* 205-233. https://doi.org/10.1007/BF02479043

Mode, C. J., & Gallop, R. J. (2008). A review of Monte Carlo simulation methods as they apply to selection and mutation in Wright-Fisher models of evolutionary genetics. *Mathematical Biosciences, 211*, 205-225.

Mode, C. J., & Sleeman, C. K. (2012). *Stochastic Processes in Genetics and Evolution.* World Scientific, New Jersey, London, Singapore, Beijing, and Hong Kong. https://doi.org/10.1142/8159

O' BOYLE, P. D. (2009). *Genetic characterization and linkage mapping of barley and net blotch resistance genes.* Ph. D. thesis, Department of Crop and Soil Sciences, Virginia Polytechnic Institute and State University.

Richards, J. K. (2016). *Genetic and molecular characterization of host resistance and susceptibility to Pyrenophora F. teres in Hordeum vulgare.* Ph. D. thesis, Department of Plant Pathology, North Dakota State University.

Richards, J., Chao, S., Friesen, T., & Brueggeman, R. (2016). Fine mapping of the barley chromosome 6H net form net blotch susceptibility locus. *G3: Genes, Genomes, Genetics, 6*(7), 1809-1818. https://doi.org/10.1534/g3.116.028902

Copyrights

International Journal of Statistics and Probability; Vol. 9, No. 2; March 2020
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

D-Optimal Slope Design for Second Degree Kronecker Model Mixture Experiment With Three Ingredients

Ngigi Peter Kung'u[1], J. K. Arap Koske[2], Josphat K. Kinyanjui[1]

[1] Department of mathematics, statistics and actuarial science, Karatina University, Karatina, Kenya

[2] Department of mathematics & computer science, Moi University, Eldoret, Kenya

Correspondence: Thomas C. Keane, Director, ISES Labs, Department of Chemistry and Biochemistry, Russell Sage College, Troy, NY 12180, USA. E-mail: keanet@sage.edu

Received: October 24, 2019 Accepted: February 5, 2020 Online Published: February 28, 2020

doi:10.5539/ijsp.v9n2p30 URL: https://doi.org/10.5539/ijsp.v9n2p30

Abstract

This study presents an investigation of an optimal slope design in the second degree Kronecker model for mixture experiments in three dimensions. The study is restricted to weighted centroid designs, with the second degree Kronecker model. A well-defined coefficient matrix is used to select a maximal parameter subsystem for the model since its full parameter space is inestimable. The information matrix of the design is obtained using a linear function of the moment matrices for the centroids and directly linked to the slope matrix. The discussion is based on Kronecker product algebra which clearly reflects the symmetries of the simplex experimental region. Eventually the matrix means are used in determining optimal values of the efficient developed design.

Keywords: information matrix, moment matrix, optimal design, response surface methodology, weighted centroid design, Kiefer ordering

1. Introduction

This study deals with the exploration and optimization of response surface. This is a problem faced by experimenters in many technical fields, where in general the response of interest is affected by a set of independent factors. In this response surface methodology (RSM) problem we assume a response of interest is influenced by three factors with the intent of optimizing this response. The response in linked to the factors through a second degree polynomial model.

In this mixture experiment the response is a function of the proportions of each ingredient. Let x_i represent the proportion of the ith ingredient in the mixture. Then, we have two conditions, $x_i \geq 0, i = 1,2,3$ and $\sum_{i=1}^{3} x_i = 1$. Evidently the levels of the factors x_i are interdependent. The experimental region for the mixture problem is a two dimensional simplex.

2. Materials and Methods

Let $1_m=(1, \ldots, 1)' \in \Re^m$ be a unity vector. The experimental conditions $t=(t_1, t_2, \ldots, t_m)$ with $t_i \geq 0$ of a mixture experiment are points in the probability simplex,

$$T_m = \{t = (t_1, \quad t_2, \quad \ldots, \quad t_m)' \in [0, \quad 1]^m : 1_m' t = 1\}.$$

Under experimental conditions, $t \in T_m$, the response Y_t is taken to be a quantitative random variable. The responses are assumed to be uncorrelated with equal but unknown finite variance say $\sigma^2 \in (0, \infty)$. The design in point has finite number of support points.

This study adopts a second degree polynomial regression function with the expected response:

$$E(Y_t) = f(t)'\theta = \sum_{i=1}^{m} \theta_{ii} t_i^{\,2} + \sum_{\substack{i,j=1 \\ i<j}}^{m} (\theta_{ij} + \theta_{ji}) t_i t_j \tag{1}$$

where Y_t, is the response under experimental condition $t \in T_m$, and $\theta = (\theta_{11}, \theta_{12}, ..., \theta_{mm}) \in \Re^{m^2}$ an unknown parameter. (see (Draper & Pukelsheim, 1998)).

A general review of design environment is done by (Pukelsheim, 1993) while (Klein, 2004) showed that the class of weighted centroid designs with at least two ingredients is essentially complete for the Kiefer ordering, (Draper, Heilijers, & Pukelsheim, 2000). As a consequence, we restrict the study to weighted centroid design.

General Design Problem

The problem of finding a design with maximum information on the parameter subsystem $K'\theta$ can be formulated as;

$$\text{Maximize } \varphi_p(C_k(M(\tau))) \text{ with } \tau \in T \tag{2}$$

$$\text{Subject to } C_k(M(\tau)) \in PD(s)\ \tau \in T$$

where T denotes the set of all designs T_m. The side condition $C_k(M(\tau)) \in PD(s)$ is equal to the existence of an unbiased linear estimator for $K'\theta$ under τ, Pukelsheim (1993). In which case, the design τ is called feasible for $K'\theta$. Any design solving problem (2) above for a fixed p∈ (-∞, 1] is called $\phi_p - optimal$for $K'\theta$ in T. For all p∈ (-∞, 1], the existence of $\phi_p - optimal$design for $K'\theta$ is certain, (Pukelsheim, 1993).

Moment Matrix

An experimental design τ is a probability measure on the experimental domain with a finite number of support points. Each support point s ∈ supp(τ) directs the experimenter to take a proportion T({t}) of all observations under experimental condition T. The statistical properties of a design are reflected by its moment matrix:

$$M(\tau) = \int_{\tau} f(t) f(t)' d\tau \in NND(m^2) \tag{3}$$

where, NND(m²) denotes the cone of nonnegative definite m²×m² matrices. The entries of M(τ) are fourth moments of τ, since the regression function $f(t)$ is purely quadratic.

Information matrix

We use unit vectors e_1, e_2, e_3 and set $e_{ij}=e_i \otimes e_j$ for i<j i,j={1, 2,3} and define the coefficient matrix

$$K = (K_1; K_2) \in \Re^{m^2 \times \binom{m+1}{2}}$$

where

$$K_1 = \sum_{i=1}^{m} e_{ii} e_i'$$

$$\text{and } K_2 = \frac{1}{m} \sum_{\substack{i,j=1 \\ i<j}}^{m} (e_{ij} + e_{ji}) E_{ij}' \tag{4}$$

Obtainable as follows:

From $e_1 = (1\ 0\ 0)'$, $e_2 = (0\ 1\ 0)'$ and $e_3 = (0\ 0\ 1)'$ we have:

$$e_{11} = e_1 \otimes e_1 = (1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0)',\ e_{22} = e_1 \otimes e_1 = (0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0)',$$

$$e_{33} = e_3 \otimes e_3 = (0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1)',\ e_{12} = e_1 \otimes e_2 = (0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0)',$$

$$e_{21} = e_2 \otimes e_1 = (0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0)',\ e_{13} = e_1 \otimes e_3 = (0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0)',$$

$$e_{31} = e_3 \otimes e_1 = (0\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0)',\ e_{23} = e_2 \otimes e_3 = (0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0)',$$

$$e_{32} = e_3 \otimes e_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}', \quad E_{12} = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}', \quad E_{13} = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}' \text{ and}$$

$$E_{23} = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}'$$

Therefore, we obtain;

$$K_1 = e_{11}e_1' + e_{22}e_2' + e_{33}e_3' = \begin{pmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{pmatrix}$$

and

$$K_2 = (e_{12} + e_{21})E_{12}' + (e_{13} + e_{31})E_{13}' + (e_{23} + e_{32})E_{23}' = \begin{pmatrix} 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 \end{pmatrix}$$

Thus

$$K = \begin{pmatrix} K_1 & K_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{3} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

The full parameter vector $\theta \in \Re^{m^2}$ for model equation (1) is not estimable. We select a maximal sub parameter vector:

$$K'\theta = \begin{Bmatrix} (\theta_{ii})_{1 \le i \le m} \\ \frac{1}{m}(\theta_{ij} + \theta_{ji}), {}_{1 \le i < j \le m} \end{Bmatrix} \in \Re^{\binom{m+1}{2}}$$

for all

$$\theta \in \Re^{m^2} \tag{5}$$

To optimize the response, we focus on the movement of the design center along the direction of the directional derivatives of the response function, that is, $\frac{\partial Y_t}{\partial t}$. Since the designs that attain certain properties in Y (estimated

response) do not enjoy the same properties for the estimated derivatives (slopes), we consider experimental designs that are constructed with derivatives in mind, (Murty & Studden, 1972) and (Ott & Mendenhall, 1972).

In practice, it is often of interest to investigate the slope of the response surface at a point t, not only over the axial directions, but also over any specified direction. We develop the concept of robust slope over all directions. Define D, a matrix arising from the differentiation of $f(t)'\theta$ with respect to each of the m independent factors, (see (Sung, Hyang, & Rabindra, 2009)). That is;

$$D=\left(\frac{\partial f'(t)}{\partial t_1},\frac{\partial f'(t)}{\partial t_2},...,\frac{\partial f'(t)}{\partial t_m}\right)', \quad \text{where,} \quad f(t)=t\otimes t \tag{6}$$

An important matrix for the design with three ingredients is the adjusted 3×6 slope matrix $H_0=DK$.

The amount of information a design contains on K'θ is captured by the information matrix:

$$C_k(M(\tau))=\min\{LM(\tau)L' \quad L\in \Re^{\binom{m+1}{2}\times m^2};\ LK=I_{\binom{m+1}{2}}\} \tag{7}$$

where $I_{\binom{m+1}{2}}$ denotes the $\binom{m+1}{2}\times\binom{m+1}{2}$ identity matrix and L is the left inverse of K derived from the linear relation, $L=(K'K)^{-1}K'$. The information matrices for K'θ takes the form:

$$C_0=LM(\tau)L'\in NND\left(\binom{m+1}{2}\right) \tag{8}$$

Thus the information matrices for K'θ are linear transformations of the moment matrices.

We then consider optimizing the information matrices for K'θ of the form:

$$C=H_0C_0H_0'\in NNND(m) \tag{9}$$

Optimality Criteria

We will compute optimal design for the polynomial fit model using matrix mean ϕ_p which is an information function (Pukelsheim, 1993). For an information matrix $C_k(M(\tau))\in$ PD(m) the kiefers ϕ_p-criteria are defined by:

$$\phi_p(C)=\begin{cases}\lambda_{\min}(C) & if \quad p=-\infty\\ \det(C)^{\frac{1}{\binom{m+1}{2}}} & if \quad p=0\\ \left[\frac{1}{\binom{m+1}{2}}traceC^p\right]^p & if \quad p\in[-\infty;1]\setminus\{0\}\end{cases} \tag{10}$$

where $\lambda_{\min}(C)$ refers to the smallest eigenvalue of C. By definition $\phi_p(C)$ is a scalar measure which is a function of the eigenvalues of C for all $p\in[-\infty;1]$. (Pukelsheim, 1993) .

Consequently a design with maximum information on the parameter subsystem $K'\theta$ solves the problem;

$$\text{Maximize} \quad \phi_p(C_k(M(\tau))) \text{ with } \tau\in T$$

$$\text{Subject to } C_k(M(\tau))\in \text{ PD(m)} \tag{11}$$

Suppose $\eta(\alpha)$ satisfies the side condition $C_k(M(\tau))\in$ PD(m) and write $C_j=C_k(M(\eta_j))$ for j=(1, 2, 3). For all $p\in(-\infty;1]$, $\eta(\alpha)$ solves problem (11) if and only if;

$$traceH_0C_jC^{p-1}H_0'\begin{cases}=traceH_0C^pH_0' & for\ all\ j\in\partial(\alpha)\\ \le traceH_0C^pH_0' & otherwise\end{cases} \tag{12}$$

(T. K. , 2004).

3. Construction of the design

We consider the weighted centroid design $\eta(\alpha)=\sum_{j=1}^{3}\alpha_j\eta_j=\alpha_1\eta_1+\alpha_2\eta_2+\alpha_3\eta_3$ with three elementary centroids (captured from the support points):

$$\eta_1=\left\{\begin{pmatrix}1\\0\\0\end{pmatrix},\ \begin{pmatrix}0\\1\\0\end{pmatrix},\ \begin{pmatrix}0\\0\\1\end{pmatrix}\right\},\ \eta_2=\left\{\begin{pmatrix}1/2\\1/2\\0\end{pmatrix},\begin{pmatrix}1/2\\0\\1/2\end{pmatrix},\begin{pmatrix}0\\1/2\\1/2\end{pmatrix}\right\}\ \text{and}\ \eta_3=\left\{\begin{pmatrix}1/3\\1/3\\1/3\end{pmatrix}\right\}.$$

These designs discovered by ((Scheffe', 1958) and (H., 1963)), are exchangeable and invariant under permutations, (T. K. , 2002). Weighted centroid designs are exchangeable.

The moment matrices for η_1 and η_2 are:

$$M(\eta_1)=\begin{pmatrix}1/3&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0\\0&0&0&0&1/3&0&0&0&0\\0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&1/3\end{pmatrix}$$

and

$$M(\eta_2)=\begin{pmatrix}1/24&1/48&1/48&1/48&1/48&0&1/48&0&1/48\\1/48&1/48&0&1/48&1/48&0&0&0&0\\1/48&0&1/48&0&0&0&1/48&0&1/48\\1/48&1/48&0&1/48&1/48&0&0&0&0\\1/48&1/48&0&1/48&1/24&1/48&0&1/48&1/48\\0&0&0&0&1/48&1/48&0&1/48&1/48\\1/48&0&1/48&0&0&0&1/48&0&1/48\\0&0&0&0&1/48&1/48&0&1/48&1/48\\1/48&0&1/48&0&1/48&1/48&1/48&1/48&1/24\end{pmatrix}.$$

Defining matrix $\widetilde{L}=(K'K)^{-1}K'$ where K is the earlier defined (equation 4) coefficient matrix,

$$\widetilde{L}=(K'K)^{-1}K'=\begin{pmatrix}1&0&0&0&0&0&0&0&0\\0&0&0&0&1&0&0&0&0\\0&0&0&0&0&0&0&0&1\\0&3/2&0&3/2&0&0&0&0&0\\0&0&3/2&0&0&0&3/2&0&0\\0&0&0&0&0&3/2&0&3/2&0\end{pmatrix}$$

The information matrices for the designs η_1 and η_2 are obtained as follows:

$$C_1 = C_k(M(\eta_1)) = \widetilde{L}(M(\eta_1))\widetilde{L}' = \begin{pmatrix} 1/3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{13}$$

and

$$C_2 = C_k(M(\eta_2)) = \widetilde{L}(M(\eta_2))\widetilde{L}' = \begin{pmatrix} 1/24 & 1/48 & 1/48 & 1/16 & 1/16 & 0 \\ 1/48 & 1/24 & 1/48 & 1/16 & 0 & 1/16 \\ 1/48 & 1/48 & 1/24 & 0 & 1/16 & 1/16 \\ 1/16 & 1/16 & 0 & 3/16 & 0 & 0 \\ 1/16 & 0 & 1/16 & 0 & 3/16 & 0 \\ 0 & 1/16 & 1/16 & 0 & 0 & 3/16 \end{pmatrix} \tag{14}$$

Using equations (13) and (14), we obtain the information matrix for the design $\eta(\alpha)$ from; $C_k(M(\eta(\alpha))) = \alpha_1 C(M(\eta_1)) + \alpha_2 C(M(\eta_2))$, as

$$C_k = C_k(M(\eta(\alpha))) = \begin{pmatrix} \frac{8\alpha_1+\alpha_2}{24} & \frac{\alpha_2}{48} & \frac{\alpha_2}{48} & \frac{\alpha_2}{16} & \frac{\alpha_2}{16} & 0 \\ \frac{\alpha_2}{48} & \frac{8\alpha_1+\alpha_2}{24} & \frac{\alpha_2}{48} & \frac{\alpha_2}{16} & 0 & \frac{\alpha_2}{16} \\ \frac{\alpha_2}{48} & \frac{\alpha_2}{48} & \frac{8\alpha_1+\alpha_2}{24} & 0 & \frac{\alpha_2}{16} & \frac{\alpha_2}{16} \\ \frac{\alpha_2}{16} & \frac{\alpha_2}{16} & 0 & \frac{3\alpha_2}{16} & 0 & 0 \\ \frac{\alpha_2}{16} & 0 & \frac{\alpha_2}{16} & 0 & \frac{3\alpha_2}{16} & 0 \\ 0 & \frac{\alpha_2}{16} & \frac{\alpha_2}{16} & 0 & 0 & \frac{3\alpha_2}{16} \end{pmatrix}$$

This matrix has a regular inverse,

$$[C(M(\eta(\alpha)))]^{-1} = \begin{pmatrix} \frac{3}{\alpha_1} & 0 & 0 & \frac{-1}{\alpha_1} & \frac{-1}{\alpha_1} & 0 \\ 0 & \frac{3}{\alpha_1} & 0 & \frac{-1}{\alpha_1} & 0 & \frac{-1}{\alpha_1} \\ 0 & 0 & \frac{3}{\alpha_1} & 0 & \frac{-1}{\alpha_1} & \frac{-1}{\alpha_1} \\ \frac{-1}{\alpha_1} & \frac{-1}{\alpha_1} & 0 & \frac{2(8\alpha_1+\alpha_2)}{3\alpha_1\alpha_2} & \frac{1}{3\alpha_1} & \frac{1}{3\alpha_1} \\ \frac{-1}{\alpha_1} & 0 & \frac{-1}{\alpha_1} & \frac{1}{3\alpha_1} & \frac{2(8\alpha_1+\alpha_2)}{3\alpha_1\alpha_2} & \frac{1}{3\alpha_1} \\ 0 & \frac{-1}{\alpha_1} & \frac{-1}{\alpha_1} & \frac{1}{3\alpha_1} & \frac{1}{3\alpha_1} & \frac{2(8\alpha_1+\alpha_2)}{3\alpha_1\alpha_2} \end{pmatrix} \tag{15}$$

The slope matrix D as defined by equation (6) is obtained as

$$D = \begin{pmatrix} 2t_1 & t_2 & t_3 & t_2 & 0 & 0 & t_3 & 0 & 0 \\ 0 & t_1 & 0 & t_1 & 2t_2 & t_3 & 0 & t_3 & 0 \\ 0 & 0 & t_1 & 0 & 0 & t_2 & t_1 & t_2 & 2t_3 \end{pmatrix}$$

A corresponding adjusted slope matrix $H_0 = DK$ is thus given by;

$$H_0 = \begin{pmatrix} 2t_1 & 0 & 0 & \frac{2}{3}t_2 & \frac{2}{3}t_3 & 0 \\ 0 & 2t_2 & 0 & \frac{2}{3}t_1 & 0 & \frac{2}{3}t_3 \\ 0 & 0 & 2t_3 & 0 & \frac{2}{3}t_1 & \frac{2}{3}t_2 \end{pmatrix}$$

To get the D-optimal design we employ the relation, that $\eta(\alpha)$ is $\phi_p - optimal$ for $K'\theta$ in T if and only if;

$$trace\, H_0 C_j C^{p-1} H_0' \begin{cases} = traceH_0 C^p H_0' \quad for \;\; j = 1,2 \\ < traceC^p \quad otherwise \end{cases}$$

From which, the unique D-optimal design for $K'\theta$ is derived using the equation (putting p=0)

$$trace\, H_0 C_j C^{-1} H_0' = traceH_0 C^0 H_0' = traceH_0 H_0' \quad for \;\; j = 1,2 \tag{16}$$

The following results can be easily demonstrated using condition (16):

- For j=1

$$H_0 C_1 C^{-1} H_0' = \frac{1}{3\alpha_1} \begin{pmatrix} 12t_1^2 - \frac{4}{3}(t_1t_2 + t_1t_3) & -\frac{4}{3}t_1^2 & -\frac{4}{3}t_1^2 \\ -\frac{4}{3}t_2^2 & 12t_2^2 - \frac{4}{3}(t_1t_2 + t_2t_3) & -\frac{4}{3}t_2^2 \\ -\frac{4}{3}t_3^2 & -\frac{4}{3}t_3^2 & 12t_3^2 - \frac{4}{3}(t_1t_3 + t_2t_3) \end{pmatrix}, \text{ with}$$

$$trace\, H_0 C_1 C^{-1} H_0' = \frac{1}{3\alpha_1}\left[12(t_1^2 + t_2^2 + t_3^2) - \tfrac{8}{3}(t_1t_2 + t_1t_3 + t_2t_3)\right] = \frac{496}{27\alpha_1} \text{ and}$$

$$H_0 H_0' = \begin{pmatrix} 4t_1^2 + \frac{4}{9}(t_2^2 + t_3^2) & \frac{4}{9}t_1t_2 & \frac{4}{9}t_1t_3 \\ \frac{4}{9}t_1t_2 & 4t_2^2 + \frac{4}{9}(t_1^2 + t_3^2) & \frac{4}{9}t_2t_3 \\ \frac{4}{9}t_1t_3 & \frac{4}{9}t_2t_3 & 4t_3^2 + \frac{4}{9}(t_1^2 + t_2^2) \end{pmatrix}, \text{ with}$$

$$traceH_0 H_0' = \frac{44}{9}(t_1^2 + t_2^2 + t_3^2) = \frac{638}{27}$$

The condition, $traceH_0 C_1 C^{-1} H_0' = traceH_0 H_0'$ implies that $\frac{496}{27\alpha_1} = \frac{638}{27}$, giving $\alpha_1 = \frac{248}{319}$

- For j=2;

$$H_0 C_2 C^{-1} H_0' = \frac{4}{9\alpha_2} \begin{pmatrix} t_2^2 + t_3^2 + t_1t_2 + t_1t_3 & t_1^2 + t_1t_2 & t_1^2 + t_1t_3 \\ t_2^2 + t_1t_2 & t_1^2 + t_3^2 + t_1t_2 + t_2t_3 & t_2^2 + t_2t_3 \\ t_3^2 + t_1t_3 & t_3^2 + t_2t_3 & t_2^2 + t_2^2 + t_1t_3 + t_2t_3 \end{pmatrix}$$

with $traceH_0 C_2 C^{-1} H_0' = \frac{8}{9\alpha_2}\left(t_1^2 + t_2^2 + t_3^2 + t_1t_2 + t_1t_3 + t_2t_3\right) = \frac{142}{27\alpha_2}$

The equation, $traceH_0 C_2 C^{-1} H_0' = traceH_0 H_0'$ implies that $\frac{142}{27\alpha_2} = \frac{638}{27}$, giving $\alpha_2 = \frac{71}{319}$

Therefore the unique D-optimal design for $K'\theta$ is

$$\eta(\alpha^{(D)}) = \alpha_1\eta_1 + \alpha_2\eta_2 = \frac{248}{319}\eta_1 + \frac{71}{319}\eta_2 .$$

The information matrix:

$$H_0CH_0' = \begin{pmatrix} 4at_1^2 + 4b(t_2^2 + t_3^2 + 2t_1t_2 + 2t_1t_3) & 4b(t_1^2 + t_2^2 + 2t_1t_2) & 4b(t_1^2 + t_3^2 + 2t_1t_3) \\ 4b(t_1^2 + t_2^2 + 2t_1t_2) & 4at_2^2 + 4b(t_1^2 + t_3^2 + 2t_1t_2 + 2t_2t_3) & 4b(t_2^2 + t_3^2 + 2t_2t_3) \\ 4b(t_1^2 + t_3^2 + 2t_1t_3) & 4b(t_2^2 + t_3^2 + 2t_2t_3) & 4at_3^2 + 4b(t_1^2 + t_2^2 + 2t_1t_3 + 2t_2t_3) \end{pmatrix}$$

Where $a = \frac{8\alpha_1 + \alpha_2}{24}$, $b = \frac{\alpha_2}{48}$, $t_1^2 = \frac{29}{18}$ *and* $t_it_j = \frac{13}{36}$

The maximum of the D-criterion is $v(\phi_0) = \left(\frac{1}{5.964}\right)^{\frac{1}{3}} = 0.5514 \cdot$

4. Conclusion

The design presented is highly efficient and can be employed as a design for a finite sample size. Of importance is to relate the weights to the number of support points for each centroid. However, the experimenter is cautioned to ensure high accuracy levels in the measurement of ingredient levels.

References

Draper, N. R., & Pukelsheim, F. (1998). Mixture Models Based on Homogeneous Polynomials. *Journal of Statistical Planning and Inference, 71*, 303-311. https://doi.org/10.1016/S0378-3758(98)00012-3

Draper, N. R., Heilijers, B., & Pukelsheim, F. (2000). Kiefer Orderinf of Simplex Designs for Mixture Models with Four Ingredients. *Annals of Statistics, 28*, 578-590. https://doi.org/10.1214/aos/1016218231

Henry, S. (1963). The Simplex-Centroid Design for Experiments with Mixtures. *J. Roy. Statist. Soc. Ser., 25*, 235-257. https://doi.org/10.1111/j.2517-6161.1963.tb00506.x

Klein, T. (2004). Optimal Designs for Second-Degree Kronecker Model Mixture Experiments. *Journal of Statistical Planing and Inference, 123*, 117-131. https://doi.org/10.1016/S0378-3758(03)00145-9

Murty, V. N., & Studden, W. J. (1972). Optimal Designs for Estimating the Slope of a Polynomial Regression. *Journal of America Statistics Association, 67*, 869-873. https://doi.org/10.1080/01621459.1972.10481308

Ott, L., & Mendenhall, W. (1972). Designs for Estimating the Slope of a Second Order Linear Model. *Technometrics, 14*, 341-353. https://doi.org/10.1080/00401706.1972.10488920

Pukelsheim, F. (1993). *Optimal Designs of Experiments.* New York: wiley.

Scheffe, H. (1958). Experiments with Mixtures. *J. Roy. Statist. Ser, B20*, 344-360. https://doi.org/10.1111/j.2517-6161.1958.tb00299.x

Sung, H. P., Hyang, S. J., & Rabindra, D. (2009). Rotatability of Second Order Response Surface Regression Models with Corrected Error. *Quality Technology & Quantitative Management, 6*(4), 471-492. https://doi.org/10.1080/16843703.2009.11673211

T., K. (2002). Invariant Symmetric Block Matices for Design of Mixture Experiments. *Journal of Statistical Planning and Inference*, 178-196.

T., K. (2004). Optimal Designs for Second-Degree Kronecker Model ixture Experiments. *Journal of Statistical Planning and Inference, 123*, 117-131. https://doi.org/10.1016/S0378-3758(03)00145-9

International Journal of Statistics and Probability; Vol. 9, No. 2; March 2020
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Bayesian Estimation of Parameters of Weibull Distribution Using Linex Error Loss Function

Josphat. K. Kinyanjui[1] & Betty. C. Korir[2]

[1] Department of Mathematics, Statistics and Actuarial Science, Karatina University, P O Box 1957-10101, Karatina, Kenya

[2] Department of Mathematics and Computer Science, University of Eldoret, P O Box 1125-30100, Eldoret, Kenya

Correspondence: Josphat. K. Kinyanjui, Department of Mathematics, Statistics and Actuarial Science, Karatina University, P O Box 1957-10101, Karatina, Kenya. E-mail: jkinyanjui@karu.ac.ke

Received: April 18, 2019 Accepted: May 10, 2019 Online Published: February 29, 2020

doi:10.5539/ijsp.v9n2p38 URL: https://doi.org/10.5539/ijsp.v9n2p38

Abstract

This paper develops a Bayesian analysis of the scale parameter in the Weibull distribution with a scale parameter θ and shape parameter β (known). For the prior distribution of the parameter involved, inverted Gamma distribution has been examined. Bayes estimates of the scale parameter, θ, relative to LINEX loss function are obtained. Comparisons in terms of risk functions of those under LINEX loss and squared error loss functions with their respective alternate estimators, viz: Uniformly Minimum Variance Unbiased Estimator (U.M.V.U.E) and Bayes estimators relative to squared error loss function are made. It is found that Bayes estimators relative to squared error loss function dominate the alternative estimators in terms of risk function.

Keywords: Bayes estimates, squared error loss function, LINEX loss function, U.M.V.U.E.

1. Introduction

Sometimes, in practical situations either from past experience or from some reliable sources, one may have a guessed estimate of the parameter which can be treated as a prior information. Thompson (1968a, b) introduced the idea of shrinking usual estimators towards point as well as interval guess value to get the improved estimators. In some situations, in place of point estimation or interval guess value, the prior information may be available in the form of prior distribution. Applying Bayesian approach, prior information available in form of prior distribution may be utilized in the estimation of parameters. In Bayesian estimation, the loss function and prior distribution play important role.

The symmetric loss function viz squared error loss function has been widely used by several authors including Berger (1980), Box and Tiao (1973), Martz and Waller (1982), Sinha and Kale (1980). The researchers such as Aitchson and Dunsmore (1975), Berger (1980), Fergusson (1967), Varian (1975) and Zellner and Gielsel (1968), have pointed out that in some situations use of symmetric loss functions may be inappropriate. Actually, we may come across the situations where a given negative error may be more serious than a given positive error or vice-versa, e.g. in dam construction, overestimation of peak of water level is more serious than underestimation. In the same way, in estimation of reliability function, use of symmetric loss function may be inappropriate as recognized by Canfield (1970). Here, overestimation of reliability function or average failure time is more serious than underestimation. [Feymann (1987)]. Varian (1975) proposed a very useful assymetric loss function known as LINEX loss function which rises exponentially on one side of zero and almost linearly on the other side of zero.

Weibull distribution play an important role in many fields of application. It has two parameters α and β where 'α' is referred to as shape parameter and 'β' as scale parameter. This distribution was used by a Swedish scientist Weibull in (1951) to describe experimentally observed variation in fatigue resistance of steel, its elastic limits, e.t.c. It has also been used to study the variation of the length of service of radio service equipment. Finally, it has also been successfully used in reliability theory.

Theoretically, it arises as the limiting distribution as $n \to \infty$, of the smallest of n independent random variables with the same distribution. The use of the Weibull distribution as a model analyzing lifetime data, quality and reliability analyses dates back the late nineteenth century. Berger and Sun (1993) gives extensive survey of its uses in the context of lifetime data. Menderhall and Hader (1958) and Cox (1959) are among the first authors who addressed competing risks in survival analysis. Cox (1959) gave other examples where this type of distribution arises. For other references

concerning competing risks, see David and Moeschberger (1978), Basu and Klein (1982).

Bhattacharya (1967) considered the estimation of both the shape and scale parameters using an inverted Gamma prior probability density function. Canavos and Tsokos (1970) developed a fully Bayesian analysis of both the scale and shape parameters assuming independent prior distributions.

2.1 LINEX Loss Function

In some estimation and prediction problems, use of symmetric loss function may be inappropriate, as has been recognized in the literature – see, for example, Ferguson (1967), Zellner and Geisel (1968), Aitchson and Dunsmore (1975), Varian (1975) and Berger (1980).

The authors mentioned above, except for Varian, have considered symmetric linear loss functions. Varian (1975) proposed a very useful assymetric loss function known as LINEX loss function which rises approximately exponentially on one side of zero and almost linearly on the other side of zero in his applied study of real estate assessment. Underassessment results in an approximately linear loss of revenue whereas overassessment often results in appeals with attendant, substantial litigation and other costs.

Attention is directed herein at establishing properties of estimation and prediction procedures based on LINEX loss functions.

Let $\Delta = \hat{\theta} - \theta$ denote the scalar estimation error in using $\hat{\theta}$ to estimate θ. Varian (1975) introduced the following convex loss function:

$$L(\Delta) = be^{a\Delta} - c\Delta - b, \qquad a, c \neq 0,\ b > 0 \tag{2.1}$$

It is seen that $L(0) = 0$. Also, for a minimum to exist at $\Delta = 0$, we must have $ab = c$, and thus equation (2.1) can be reexpressed as

$$L(\Delta) = b\left[e^{a\Delta} - a\Delta - 1\right], \qquad a \neq 0,\ b > 0 \tag{2.2}$$

There are two parameters, *a* and *b*, involved in equation (2.2) with *'b'* serving to scale the loss function and *'a'* serving to determine its shape.

2.1.1 Obtaining Bayes Estimators Using Linex Loss Function

Let $h(\theta / D)$ denote the posterior density function of θ, where D denotes the sample and prior information, $(x_1, x_2, \cdots, x_n, \theta)$. Let $\underset{\theta}{E}$ denote the posterior expectation with respect to $h(\theta / D)$. The posterior risk is defined as the $\underset{\theta}{E}[L(\Delta)]$ and is given by

$$\underset{\theta}{E}[L(\Delta)] = b\left[e^{a\hat{\theta}}\ \underset{\theta}{E}\left(e^{-a\theta}\right) - a\left(\hat{\theta} - \underset{\theta}{E}(\theta)\right) - 1\right] \ldots \tag{2.3}$$

Theorem 2.1

The value of $\hat{\theta}$ that minimizes (2.7) is

$$\hat{\theta}_B = \frac{-1}{a}\log\left[\underset{\theta}{E}\left(e^{-a\theta}\right)\right] \tag{2.4}$$

provided, of course, $\underset{\theta}{E}\left(e^{-a\theta}\right)$ exists and is finite. This involves evaluation of moment generating function for posterior density.

Proof

The conditions for relative minimum are

(i) $\frac{\partial}{\partial\hat{\theta}}\underset{\theta}{E}[L(\Delta)] = 0$ and (ii) $\frac{\partial^2}{\partial\hat{\theta}^2}\underset{\theta}{E}[L(\Delta)] > 0$ at the minimum value.

(i) implies that

$$b\left[ae^{a\hat{\theta}}\ \underset{\theta}{E}\left(e^{-a\theta}\right) - a\right] = 0 \quad \text{that is} \quad ab\left[e^{a\hat{\theta}}\ \underset{\theta}{E}\left(e^{-a\theta}\right) - 1\right] = 0$$

Since $ab \neq 0$, it implies that

$$e^{a\hat{\theta}}\ \underset{\theta}{E}\left(e^{-a\theta}\right) - 1 = 0$$

Taking logs and simplifying, we get

$$\hat{\theta}_B = \frac{-1}{a}\log\left[\underset{\theta}{E}\left(e^{-a\theta}\right)\right]$$

(ii) implies that

$$\frac{\partial^2}{\partial\hat{\theta}^2}\underset{\theta}{E}\left[L(\Delta\right] = \frac{\partial}{\partial\hat{\theta}}\left[\frac{\partial}{\partial\hat{\theta}}\underset{\theta}{E}\left[L(\Delta)\right]\right]$$
$$= a^2 b\, e^{a\hat{\theta}}\underset{\theta}{E}\left[e^{-a\hat{\theta}}\right] > 0$$

Since $\hat{\theta}_B$ satisfies conditions (i) and (ii), it follows that $\hat{\theta}_B$ is the minimum value.

3. Bayesian Estimation of Parameters in Case of Weibull Distribution

3.1 Introduction

In this section, we develop a Bayesian analysis for the Weibull distribution with respect to the usual life-testing procedures. It is divided into two parts: in the first, we consider the Bayes estimates of the parameters of Weibull distribution using squared error loss function while in the second part we consider the Bayes estimates of the parameters of Weibull distribution using LINEX error loss function. These estimates obtained based on the two loss functions are later compared in order to show the corresponding efficiencies.

3.2 Weibull Bayesian Distribution

The Weibull distribution is given by the probability density function

$$f(x) = \begin{cases} \frac{\beta}{\theta}x^{\beta-1}e^{\frac{-x^\beta}{\theta}}, & 0 < x < \infty, \quad \theta, \beta > 0 \\ 0, \textit{ Otherwise} \end{cases} \tag{3.1}$$

The parameters θ and β are called the scale and shape parameters respectively. Because the shape parameter is known in some cases, we treat the scale parameter θ as the random variable. We derive a fully Bayesian solution by assuming independent prior distribution of θ. Specifically, we consider an inverted gamma,

$$\lambda(\theta) = \begin{cases} \frac{(\mu/\theta)^{v+1}e^{\frac{-\mu}{\theta}}}{\mu\Gamma(v)}, & 0 < \theta < \infty, \quad \mu, v > 0 \\ 0, \textit{ Otherwise} \end{cases} \tag{3.2}$$

The reasons for considering the inverted gamma prior density (3.2) are that it is flexible enough to capture almost any kind of prior experience, and it also possesses the attractive property that the posterior distribution of the parameter after the sample has been observed is also of the inverted gamma type. A family of prior densities which gives rise to posteriors belonging to the same family is very useful inasmuch as the mathematical tractability is maintained, and this 'nice' property has been termed 'closure under sampling' by Wetherill (1961). For densities which admit sufficient statistics of fixed dimensionality, Raiffa and Schlaifer have considered a method of generating prior densities on the parameter space that possess this desirable property. A family of such densities has been called by them a 'natural conjugate family', and for Weibull density (3.1), the inverted gamma prior forms such a family.

For our distribution given in (3.1), we have

$$f\left(\frac{x}{\theta,\beta}\right) = L(x,\theta,\beta) = \prod_{i=1}^{n} f(x_i,\theta,\beta) = \left(\frac{\beta}{\theta}\right)^n \prod_{i=1}^{n} x_i^{\beta-1} e^{\frac{-1}{\theta}\sum_{i=}^{n} x_i^\beta} \tag{3.3}$$

Now substituting the assumed value of $\lambda(\theta)$ given in (3.2) and the value of $f\left(\frac{x}{\theta,\beta}\right)$ obtained in (3.3) to (3.1), we obtain

$$f(\underline{x},\theta,\beta)=\frac{\beta^n\mu^v\prod_{i=1}^{n}x_i^{\beta-1}e^{\frac{-1}{\theta}\sum_{i=}^{n}\left(x_i^{\beta}+\mu\right)}}{\theta^{n+v+1}\,\Gamma(v)} \tag{3.4}$$

From (3.4), the marginal probability density function of X is

$$\begin{aligned} f(\underline{x})&=\int_{\theta}f(\underline{x},\theta,\beta)d\theta=\int_{\theta=0}^{\infty}\frac{\beta^n\mu^v\prod_{i=1}^{n}x_i^{\beta-1}e^{\frac{-1}{\theta}\sum_{i=}^{n}\left(x_i^{\beta}+\mu\right)}}{\theta^{n+v+1}\,\Gamma(v)}d\theta \\ &=\frac{\beta^n\mu^v\prod_{i=1}^{n}x_i^{\beta-1}\Gamma(n+v)}{\left(\mu^*\right)^{n+v}\Gamma(v)} \end{aligned} \tag{3.5}$$

where

$$\mu^*=\sum_{i=1}^{n}x_i^{\beta}+\mu \tag{3.6}$$

The posterior density function of θ given $X=\underline{x}$ is given by

$$h\left(\theta/\underline{x}\right)=\frac{f\left(\underline{x}/\theta\right)\lambda(\theta)}{\int_{\theta}f\left(\underline{x}/\theta\right)\lambda(\theta)d\theta} \tag{3.7}$$

The values obtained in (3.4) and (3.5) are substituted in (3.7) to obtain;

$$h\left(\theta/\underline{x}\right)=\frac{\left(\mu^*\right)^{n+v}e^{\left(\frac{-1}{\theta}\right)\mu^*}}{\theta^{n+v+1}\,\Gamma(n+v)} \tag{3.8}$$

which is the required posterior density function and μ^* is as given in (3.6).

3.3 Bayes Estimator and Bayes Risk Using Squared Error Loss Function

Theorem 3.1

If the loss function is the squared error, $L\left(\theta,\hat{\theta}_M\right)=\left(\theta-\hat{\theta}_M\right)^2$, then the Bayes estimator with respect to the prior distribution $\lambda(\theta)$ of θ is given by:

$$\hat{\theta}_M=E\left(\theta/\underline{x}\right)=\frac{\mu^*}{n+v-1} \tag{3.9}$$

Proof

From (3.9);

$$\hat{\theta}_M=E\left(\theta/\underline{x}\right)=\int_{\theta=0}^{\infty}\theta\,h\left(\theta/\underline{x}\right)d\theta \tag{3.10}$$

Now substituting the value obtained in (3.8) to (3.10), we get;

$$\hat{\theta}_M=\frac{\left(\mu^*\right)^{n+v}}{\Gamma(n+v)}\int_{\theta=0}^{\infty}\frac{e^{\left(\frac{-1}{\theta}\right)\mu^*}}{\theta^{n+v+1}}d\theta$$

which upon simplification gives;

$\hat{\theta}_M = \frac{\mu^*}{n+v-1}$ which is the required Bayes estimator of θ.

Theorem 3.2

If the loss function is the squared error, $L(\theta,\hat{\theta}_M)=(\theta-\hat{\theta}_M)^2$, then the corresponding Bayes risk is given by:

$$ {}_B R_S(\hat{\theta}_M) = E\left[Var\left(\theta/\underline{x}\right)\right] = \frac{(\mu^*)^2}{(n+v-1)(v-1)(v-2)} \tag{3.11}$$

Proof

$$ Var\left(\theta/\underline{x}\right) = \int_\theta \left[\theta - E\left(\theta/\underline{x},\beta\right)\right]^2 h\left(\theta/\underline{x}\right)d\theta = E\left(\theta^2/\underline{x},\beta\right) - \left[E\left(\theta/\underline{x},\beta\right)\right]^2 \tag{3.12}$$

Now using (3.8), we get;

$$ E\left(\theta^2/\underline{x},\beta\right) = \int_{\theta=0}^{\infty} \theta^2 h\left(\theta/\underline{x}\right)d\theta = \frac{(\mu^*)^2}{(n+v-1)(n+v-2)} \tag{3.13}$$

Substituting the values obtained in (3.8) and (3.13) into (3.12), we get;

$$ Var\left(\theta/\underline{x}\right) = \frac{(\mu^*)^2}{(n+v-1)^2(n+v-2)};\ n+v>1 \tag{3.14}$$

Hence using (3.14) in (3.11), we get;

$$ {}_B R_S(\hat{\theta}_M) = \int\int\cdots\int \frac{\beta^n \mu^v \prod_{i=1}^n x_i^{\beta-1}\Gamma(n+v)}{(n+v-1)^2(n+v-2)(\mu^*)^{n+v-2}\Gamma(v)}d\underline{x}$$

$$ = \frac{\beta^n \mu^v \Gamma(n+v)}{(n+v-1)^2(n+v-2)\Gamma(v)} \int_{x_1}\int_{x_2}\cdots\int_{x_n} \frac{\prod_{i=1}^n x_i^{\beta-1}}{(\mu^*)^{n+v-2}}d\underline{x}$$

This is simplified to give

$$ {}_B R_S(\hat{\theta}_M) = \frac{\beta^n \mu^v \Gamma(n+v-2)}{(n+v-1)\Gamma(v)} I_n \tag{3.15}$$

where

$$ I_n = \int_{x_1}\int_{x_2}\cdots\int_{x_n} \frac{\prod_{i=1}^n x_i^{\beta-1}}{\left(\sum_{i=1}^n x_i^\beta + \mu\right)^{n+v-2}}d\underline{x} = \frac{\Gamma(v-2)}{\beta^n \mu^{v-2}\Gamma(n+v-2)} \tag{3.16}$$

Hence from (3.15) and (3.16), we obtain

$$ {}_B R_S(\hat{\theta}_M) = \frac{(v-3)!\,\mu^2}{(n+v-1)\Gamma(v)} = \frac{\mu^2}{(v-1)(v-2)(n+v-1)}$$

which is the required Bayes risk of $\hat{\theta}_M$ relative to squared error loss function.

3.4 Risk Function of $\hat{\theta}_M$ Using Linex and Squared Error Loss Function

In this section, we shall obtain the risk function of the Bayes estimator $\hat{\theta}_M$, using both squared and Linex error loss function.

Theorem 3.3

The risk function of $\hat{\theta}_M$ obtained using squared error loss function, $L(\theta,\hat{\theta}_M)=(\theta-\hat{\theta}_M)^2$, is given by:

$$R_S\left(\hat{\theta}_M\right)=\left[\frac{\mu-\theta(v-1)}{(n+v-1)}\right]^2+\frac{n\theta^2}{(n+v-1)^2}$$

Proof

$$R_S(\hat{\theta}_M)=\int_{x_1}\int_{x_2}\cdots\int_{x_n}\left(\hat{\theta}_M-\theta\right)^2 f(\underline{x}/\theta,\beta)\,d\underline{x} \tag{3.17}$$

where $f(\underline{x}/\theta,\beta)$ is the joint distribution.

From (3.3)

$$\begin{aligned}R_S(\hat{\theta}_M)&=\int_{x_1}\int_{x_2}\cdots\int_{x_n}\left(\hat{\theta}_M^2-2\theta\hat{\theta}_M+\theta^2\right)f(\underline{x}/\theta,\beta)\,d\underline{x}\\&=\int_{x_1}\int_{x_2}\cdots\int_{x_n}\hat{\theta}_M^2 f(\underline{x}/\theta,\beta)\,d\underline{x}-2\theta\int_{x_1}\int_{x_2}\cdots\int_{x_n}\hat{\theta}_M f(\underline{x}/\theta,\beta)\,d\underline{x}+\theta^2\int_{x_1}\int_{x_2}\cdots\int_{x_n}f(\underline{x}/\theta,\beta)\,d\underline{x}\end{aligned} \tag{3.18}$$

Substituting the value of $\hat{\theta}_M$ obtained in (3.10) to (3.18) we get

$$\begin{aligned}R_S\left(\hat{\theta}_M\right)&=\int_{x_1}\cdots\int_{x_n}\left(\frac{\mu^*}{n+v-1}\right)^2\left(\frac{\beta}{\theta}\right)^n\prod_{i=1}^n x_i^{\beta-1}e^{\left(\frac{-1}{\theta}\right)\sum_{i=1}^n x_i^\beta}\,d\underline{x}\\&-2\theta\int_{x_1}\cdots\int_{x_n}\left(\frac{\mu^*}{n+v-1}\right)\left(\frac{\beta}{\theta}\right)^n\prod_{i=1}^n x_i^{\beta-1}e^{\left(\frac{-1}{\theta}\right)\sum_{i=1}^n x_i^\beta}\,d\underline{x}+\theta^2\end{aligned} \tag{3.19}$$

Using μ^* as given in (3.6) in (3.19) gives

$$\begin{aligned}R_S\left(\hat{\theta}_M\right)&=\frac{\beta^n}{\theta^n(n+v-1)^2}\int_{x_1}\cdots\int_{x_n}\left(\sum_{i=1}^n x_i^\beta+\mu\right)^2\prod_{i=1}^n x_i^{\beta-1}e^{\left(\frac{-1}{\theta}\right)\sum_{i=1}^n x_i^\beta}\,d\underline{x}\\&-\frac{2\beta^n}{\theta^n(n+v-1)}\int_{x_1}\cdots\int_{x_n}\left(\sum_{i=1}^n x_i^\beta+\mu\right)\prod_{i=1}^n x_i^{\beta-1}e^{\left(\frac{-1}{\theta}\right)\sum_{i=1}^n x_i^\beta}\,d\underline{x}+\theta^2\end{aligned} \tag{3.20}$$

which upon simplification of integrals gives

$$R_S\left(\hat{\theta}_M\right)=\left[\frac{\mu-\theta(v-1)}{(n+v-1)}\right]^2+\frac{n\theta^2}{(n+v-1)^2}$$

Theorem 3.4

The risk function of $\hat{\theta}_M$ using Linex error loss function is given by:

$$R_L\left(\hat{\theta}_M\right)=b\left[\frac{e^{-a\left(\theta-\mu/(n+v-1)\right)/(n+v-1)}}{(1-a\theta/(n+v-1))^n}+\frac{a\theta(v-1)-a\mu}{(n+v-1)}-1\right] \tag{3.21}$$

Proof

Using Linex error loss function as given in (2.4), we get

$$E[L(\Delta)]=be^{-a\theta}\int_{x_1}\int_{x_2}\cdots\int_{x_n}e^{a\hat{\theta}_M}f(\underline{x}/\theta)\,d\underline{x}-ab\int_{x_1}\int_{x_2}\cdots\int_{x_n}\hat{\theta}_M e^{a\hat{\theta}_M}f(\underline{x}/\theta)\,d\underline{x}+b(a\theta-1) \tag{3.22}$$

Substituting the value of $\hat{\theta}_M$ obtained in (3.10) to (3.18) we get

$$R_L\left(\hat{\theta}_M\right)= be^{-a\theta}\int_{x_1}\cdots\int_{x_n} e^{a\mu^*/(n+v-1)}\left(\frac{\beta}{\theta}\right)^n\prod_{i=1}^{n}x_i^{\beta-1}\,e^{\left(\frac{-1}{\theta}\right)\sum_{i=1}^{n}x_i^{\beta}}\,d\underline{x}$$

$$-ab\int_{x_1}\cdots\int_{x_n}\left(\frac{\mu^*}{n+v-1}\right)\left(\frac{\beta}{\theta}\right)^n\prod_{i=1}^{n}x_i^{\beta-1}\,e^{\left(\frac{-1}{\theta}\right)\sum_{i=1}^{n}x_i^{\beta}}\,d\underline{x}+b(a\theta-1)$$

and using μ^* as given in (3.6), we get

$$\begin{aligned}R_L\left(\hat{\theta}_M\right)&= \frac{be^{-a\theta}\beta^n}{\theta^n}\int_{x_1}\cdots\int_{x_n} e^{a\left(\sum_{i=1}^{n}x_i^{\beta}+\mu\right)/(n+v-1)}\prod_{i=1}^{n}x_i^{\beta-1}\,e^{\left(\frac{-1}{\theta}\right)\sum_{i=1}^{n}x_i^{\beta}}\,d\underline{x}\\&-\frac{ab\beta^n}{\theta^n(n+v-1)}\int_{x_1}\cdots\int_{x_n}\left(\sum_{i=1}^{n}x_i^{\beta}+\mu\right)\prod_{i=1}^{n}x_i^{\beta-1}\,e^{\left(\frac{-1}{\theta}\right)\sum_{i=1}^{n}x_i^{\beta}}\,d\underline{x}+b(a\theta-1)\end{aligned}\tag{3.23}$$

which when the integrals are solved, the final value is obtained as

$$R_L\left(\hat{\theta}_M\right)= b\left[\frac{e^{-a\left(\theta-\mu/(n+v-1)\right)/(n+v-1)}}{\left(1-a\theta/(n+v-1)\right)^n}+\frac{a\theta(v-1)-a\mu}{(n+v-1)}-1\right]$$

Theorem 3.5

The Bayes risk of $\hat{\theta}_M$ using Linex error loss function is given by:

$${}_B R_L\left(\hat{\theta}_M\right)= \frac{be^{a\mu/(n+v-1)}\,\mu^{(v+2)/2}\,a^{v/2}}{\Gamma(v)}\left[\frac{2}{\mu}K_v\left(2\sqrt{a\mu}\right)\frac{2n}{n+v-1}\sqrt{a/\mu}\,K_{v-1}\left(2\sqrt{a\mu}\right)+\frac{n(n+1)a}{(n+v-1)^2}K_{v-2}\left(2\sqrt{a\mu}\right)\right]-b\tag{3.24}$$

where $K_v(z)$ is an integral representation of the modified Bessel function of the third kind of order v.

Proof

The prior risk function of $\hat{\theta}_M$ (denoted by $R(\lambda,\hat{\theta}_M)$) with respect to the prior distribution $\lambda(\theta)$ of θ is defined as the prior expectation of the risk function. That is

$$R(\lambda,\hat{\theta}_M)= E\left[R_L\left(\hat{\theta}_M\right)\right]=\int_{\theta}R_L\left(\hat{\theta}_M\right)\lambda(\theta)d\theta$$

where

$$R_L\left(\hat{\theta}_M\right)= E\left[L\left(\hat{\theta}_M,\theta\right)\right]$$

The prior risk function is also called the Bayesian risk or simply the Bayes risk. Thus

$${}_B R_L\left(\hat{\theta}_M\right)= E\left[R_L\left(\hat{\theta}_M\right)\right]=\int_{\theta=0}^{\infty}R_L\left(\hat{\theta}_M\right)\lambda(\theta)d\theta\tag{3.25}$$

Substituting the values of $\lambda(\theta)$ and $R_L\left(\hat{\theta}_M\right)$ given in (3.2) and (3.21) respectively to (3.25) we get

$$\begin{aligned}{}_B R_L\left(\hat{\theta}_M\right)&= b\int_{\theta=0}^{\infty}\frac{e^{a\mu^*/(n+v-1)}\,e^{-a\theta}\mu^v e^{\frac{-\mu}{\theta}}}{\left[1-a\theta/(n+v-1)\right]^n\theta^{v+1}\Gamma(v)}d\theta-b\int_{\theta=0}^{\infty}\frac{a\mu\mu^v e^{\frac{-\mu}{\theta}}}{(n+v-1)\theta^{v+1}\Gamma(v)}d\theta\\&+\frac{ab(v-1)}{(n+v-1)}\int_{\theta=0}^{\infty}\frac{\theta\mu^v e^{\frac{-\mu}{\theta}}}{\theta^{v+1}\Gamma(v)}d\theta-b\end{aligned}\tag{3.26}$$

which further gives;

$$ {}_{B}R_{L}\left(\hat{\theta}_{M}\right)=\frac{be^{a\mu^{*}/(n+v-1)}\,\mu^{v}}{\Gamma(v)}\int_{\theta=0}^{\infty}\frac{e^{-\left(a\theta+\frac{\mu}{\theta}\right)}}{\left[1-a\theta/(n+v-1)\right]^{n}\theta^{v+1}}d\theta-b \tag{3.27}$$

The integral value is solved as follows:

$$\int_{\theta=0}^{\infty}\frac{e^{-\left(a\theta+\frac{\mu}{\theta}\right)}}{\left[1-a\theta/(n+v-1)\right]^{n}\theta^{v+1}}d\theta=\int_{\theta=0}^{\infty}\frac{e^{-\left(a\theta+\frac{\mu}{\theta}\right)}}{\theta^{v+1}}\left[1-\frac{a\theta}{n+v-1}\right]^{-n}d\theta$$

Using the binomial expansion to expand $\left[1-\frac{a\theta}{n+v-1}\right]^{-n}$, we have;

$$\left[1-\frac{a\theta}{n+v-1}\right]^{-n}=1+\frac{na\theta}{n+v-1}+\frac{n(n+1)a^{2}\theta^{2}}{2(n+v-1)^{2}}+\cdots$$

To simplify our integration, we can consider the first three terms and assume that the rest are negligible. Since $\frac{a\theta}{n+v-1}<0$;

$$\begin{aligned}&\int_{\theta=0}^{\infty}\frac{e^{-\left(a\theta+\frac{\mu}{\theta}\right)}}{\theta^{v+1}}\left[1-\frac{a\theta}{n+v-1}\right]^{-n}d\theta=\int_{\theta=0}^{\infty}\frac{e^{-\left(a\theta+\frac{\mu}{\theta}\right)}}{\theta^{v+1}}\left[1+\frac{na\theta}{(n+v-1)}+\frac{n(n+1)a^{2}\theta^{2}}{2(n+v-1)^{2}}\right]d\theta\\&=\int_{\theta=0}^{\infty}\frac{e^{-\left(a\theta+\frac{\mu}{\theta}\right)}}{\theta^{v+1}}d\theta+\frac{na}{(n+v-1)}\int_{\theta=0}^{\infty}\frac{e^{-\left(a\theta+\frac{\mu}{\theta}\right)}}{\theta^{v}}d\theta+\frac{n(n+1)a^{2}}{2(n+v-1)^{2}}\int_{\theta=0}^{\infty}\frac{e^{-\left(a\theta+\frac{\mu}{\theta}\right)}}{\theta^{v-1}}d\theta\end{aligned} \tag{3.30}$$

The above integrals are evaluated by using an integral representation of $K_{v}(z)$, the modified Bessel function of the third kind of order v. (Erd'ely, et. al, formula 23 p82), and subsequent use of the same formula in conjunction with the fact that $K_{-v}(z)=K_{v}(z)$. Accordingly,

$$K_{v}(az)=\frac{1}{2a^{v}}\int_{t=0}^{\infty}e^{-\frac{1}{2}z\left(t+\frac{a^{2}}{t}\right)}t^{-v-1}\,dt$$

where $\mathrm{Re}(z)>0$, $\mathrm{Re}\left(a^{2}z\right)>0$, z is the variable and v is the order of the Bessel's function.

The function

$$K_{v}(z)=\frac{\pi}{2\sin(v\pi)}\left[\mathrm{I}_{-v}(z)-\mathrm{I}_{v}(z)\right]$$

is a solution of Bessel differential equation

$$z^{2}\,d^{2}w/dz^{2}+z\,dw/dz-\left(z^{2}+v^{2}\right)w=0$$

$$\mathrm{I}_{v}(z)=\sum_{m=0}^{\infty}\frac{x^{2m+v}}{2^{2m+v}\,m!\Gamma(m+v-1)}$$

Now, in our case;

$$\int_{\theta=0}^{\infty}\frac{e^{-\left(a\theta+\frac{\mu}{\theta}\right)}}{\theta^{v+1}}d\theta=\int_{\theta=0}^{\infty}\frac{e^{-a\left(\theta+\frac{\mu}{a\theta}\right)}}{\theta^{v+1}}d\theta$$

and by comparing with the relation

$$K_v\left(a^* z\right)=\frac{1}{2a^v}\int_{t=0}^{\infty}\frac{e^{-\frac{1}{2}z\left(t+\frac{a^2}{t}\right)}t^{-v-1}}{t^{v+1}}dt$$

we see that;

$t=\theta,\ \frac{1}{2}z=a$ which implies that $z=2a$.

$\left(a^*\right)^2=\frac{\mu}{a}$ which implies that $a^*=\sqrt{\frac{\mu}{a}}$. Therefore;

$$\int_{\theta=0}^{\infty}\frac{e^{-\left(a\theta+\frac{\mu}{\theta}\right)}}{\theta^{v+1}}d\theta=\left(\frac{2}{\left(a^*\right)^v}\right)K_v\left(a^* z\right)=\frac{2}{\left(\mu/a\right)^{v/2}}K_v\left(2\sqrt{a\mu}\right) \tag{3.31}$$

Similarly,

$$\int_{\theta=0}^{\infty}\frac{e^{-\left(a\theta+\frac{\mu}{\theta}\right)}}{\theta^{v}}d\theta=\frac{2}{\left(\mu/a\right)^{(v-1)/2}}K_{v-1}\left(2\sqrt{a\mu}\right) \tag{3.32}$$

and

$$\int_{\theta=0}^{\infty}\frac{e^{-\left(a\theta+\frac{\mu}{\theta}\right)}}{\theta^{v-1}}d\theta=\frac{2}{\left(\mu/a\right)^{(v-2)/2}}K_{v-2}\left(2\sqrt{a\mu}\right) \tag{3.33}$$

Thus substituting (3.31), (3.32) and (3.33) in (3.30), we obtain

$$\begin{aligned}&\int_{\theta=0}^{\infty}\frac{e^{-\left(a\theta+\frac{\mu}{\theta}\right)}}{\theta^{v+1}}\left[1-\frac{a\theta}{n+v-1}\right]^{-n}d\theta=\frac{2}{\left(\mu/a\right)^{v/2}}K_v\left(2\sqrt{a\mu}\right)+\frac{2na}{(n+v-1)\left(\mu/a\right)^{(v-1)/2}}K_{v-1}\left(2\sqrt{a\mu}\right)\\&+\frac{2n(n+1)a^2}{2(n+v-1)^2\left(\mu/a\right)^{(v-2)/2}}K_{v-2}\left(2\sqrt{a\mu}\right)\end{aligned} \tag{3.34}$$

Substituting (3.34) in (3.27), we get

$${}_B R_L\left(\hat{\theta}_M\right)=$$
$$\frac{be^{a\mu/(n+v-1)}\,\mu^{(v+2)/2}\,a^{v/2}}{\Gamma(v)}\left[\frac{2}{\mu}K_v\left(2\sqrt{a\mu}\right)+\frac{2n}{n+v-1}\sqrt{a/\mu}\,K_{v-1}\left(2\sqrt{a\mu}\right)+\frac{n(n+1)a}{(n+v-1)^2}K_{v-2}\left(2\sqrt{a\mu}\right)\right]-b$$

which is the required Bayes risk of $\hat{\theta}_M$ using Linex error loss function.

3.5 Risk Function and Bayes Risk of Uniformly Minimum Variance Unbiased Estimator of θ

In this section, we shall obtain the Uniformly Minimum Variance Unbiased Estimator of θ. This estimator will be used to obtain the risk function and Bayes risk of the estimator using both squared and Linex error loss function.

Theorem 3.6

Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from the Weibull distribution. Let $\hat{\theta}=t\left(x_1, x_2, \cdots, x_n\right)$ be an estimator of θ. Then $\hat{\theta}=\frac{\sum_{i=1}^{n}x_i^{\beta}}{n}$ is a unique U.M.V. Unbiased Estimator of θ.

Proof

The likelihood function of $X_1, X_2, \cdots, X_n$ is given by:

$$L(\underline{x},\theta,\beta)=\prod_{i=1}^{n}f(x_i,\theta,\beta)=\left(\frac{\beta}{\theta}\right)^n\prod_{i=1}^{n}x_i^{\beta-1}e^{\frac{-1}{\theta}\sum_{i=}^{n}x_i^{\beta}} \tag{3.35}$$

Let $t(x_1,x_2,\cdots,x_n)=\sum_{i=1}^{n}x_i^{\beta}$. Hence (3.35) becomes

$L(\underline{x},\theta,\beta)=g(t,\theta)h(x_1,x_2,\cdots,x_n)$ where

$g(t,\theta)=\left(\frac{\beta}{\theta}\right)^n e^{\frac{-t}{\theta}}$ and $h(x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}x_i^{\beta-1}$

By factorization theorem, $t(x_1,x_2,\cdots,x_n)=\sum_{i=1}^{n}x_i^{\beta}$ is a sufficient statistics for θ.

To show that $t(x_1,x_2,\cdots,x_n)=\sum_{i=1}^{n}x_i^{\beta}$ is a complete statistics, we need to also show that

$E[h(T)]=0$ implies that $h(T)=0$, for all values of T, where $h(T)$ is a function of T.

$E[h(T)]=\int_{t=0}^{\infty}h(t)\Pr(T=t)dt$ where $T=\sum_{i=1}^{n}x_i^{\beta}$

To obtain the p.d.f. of T

$T=\sum_{i=1}^{n}x_i^{\beta}$. Let $y=x^{\beta}$ it implies that $x=y^{\frac{1}{\beta}}$ and $\frac{dx}{dy}=\frac{1}{\beta}y^{\frac{1}{\beta}-1}$

Therefore;

$g(y)=f(x)\frac{dx}{dy}$ which reduces to $g(y)=\frac{1}{\theta}e^{\frac{-y}{\theta}}$

Since $y_1,y_2,\cdots,y_n$ are independent and identically distributed as Y, then the distribution of T is;

$$h(T)=\frac{1}{\theta^n}e^{\frac{-t}{\theta}},\quad 0<t<\infty$$

$E[h(T)]=0$, then $\sum_{t=0}^{\infty}h(t)\frac{1}{\theta^n}e^{\frac{-t}{\theta}}=0$, which on reducing becomes

$$\left[h(0)+h(1)e^{\frac{-1}{\theta}}+h(2)e^{\frac{-2}{\theta}}+\cdots+h(n)e^{\frac{-n}{\theta}}+\cdots\right]=0$$

Therefore;

$h(0)=0,\ h(1)=0,\ h(2)=0,\cdots,\ h(n)=0$, that is, $h(t)=0$ for all values of $t=0,1,2,\cdots$

Therefore $T=\sum_{i=1}^{n}x_i^{\beta}$ is a complete statistics.

Now;

$$E(T)=E\left(\sum_{i=1}^{n}x_i^{\beta}\right)=\sum_{i=1}^{n}E\left(x_i^{\beta}\right) \tag{3.36}$$

But;

$$E\left(x^{\beta}\right)=\int x^{\beta}\left(\frac{\beta}{\theta}\right)^{n} x^{\beta-1} e^{\frac{-1}{\theta}x^{\beta}} dx = \theta$$

Thus, (3.36) reduces to

$E(T)=\sum_{i=1}^{n}\theta = n\theta$ and therefore $E(T/n)=\theta$

Thus, $\hat{\theta}=h(T)=\frac{T}{n}=\frac{\sum_{i=1}^{n}x_i^{\beta}}{n}$ is the required unique U.M.V. Unbiased Estimator of θ.

Theorem 3.7

The risk function of $\hat{\theta}$ using squared error loss function is given by $R_S\left(\hat{\theta}\right)=\frac{\theta^2}{n}$, where $\hat{\theta}$ is the unique U.M.V. Unbiased Estimator of θ.

Proof

The risk function of $\hat{\theta}$ is obtained by using the squared error loss function.

$$R_S\left(\hat{\theta}\right)=E\left[L\left(\hat{\theta},\theta\right)\right]=E\left(\hat{\theta}-\theta\right)^2=Var\left(\hat{\theta}\right) \tag{3.37}$$

Substituting the value of $\hat{\theta}$ obtained in Theorem 3.6 in (3.37), we get

$$R_S\left(\hat{\theta}\right)=Var\left(\hat{\theta}\right)=Var\left(\frac{\sum_{i=1}^{n}x_i^{\beta}}{n}\right)=\frac{1}{n^2}Var\left(\sum_{i=1}^{n}x_i^{\beta}\right) \tag{3.38}$$

Since $x_1, x_2, \cdots, x_n$ are independent and identically distributed as X, (3.38) becomes

$$R_S\left(\hat{\theta}\right)=\frac{n}{n^2}Var\left(x^{\beta}\right)=\frac{1}{n}Var\left(x^{\beta}\right) \tag{3.39}$$

Now we need to obtain $Var\left(x^{\beta}\right)$.

By definition, $Var\left(x^{\beta}\right)=E\left(x^{2\beta}\right)-\left[E\left(x^{\beta}\right)\right]^2=E\left(x^{2\beta}\right)-\theta^2$ (3.40)

Now;

$$E\left(x^{2\beta}\right)=\int x^{2\beta}\left(\frac{\beta}{\theta}\right)^{n} x^{\beta-1} e^{\frac{-1}{\theta}x^{\beta}} dx = 2\theta^2 \text{ and hence } Var\left(x^{\beta}\right)=\theta^2 \tag{3.41}$$

The value obtained in (3.41) is substituted in (3.39) to get

$$R_S\left(\hat{\theta}\right)=\frac{\theta^2}{n},$$

which is the required risk function of the unique U.M.V. Unbiased Estimator of θ.

Theorem 3.8

The Bayes risk of $\hat{\theta}$ using squared error loss function is given by ${}_B R_S\left(\hat{\theta}\right)=\frac{\mu^2}{n(v-1)(v-1)}$, where $\hat{\theta}$ is the unique U.M.V. Unbiased Estimator of θ.

Proof

The prior risk function of $\hat{\theta}$ (denoted by $R(\lambda,\hat{\theta})$) with respect to the prior distribution $\lambda(\theta)$ of is defined as the prior expectation of the risk function. That is

$$R(\lambda,\hat{\theta}) = E\left[R_S(\hat{\theta})\right] = \int R_S(\hat{\theta})\lambda(\theta)\,d\theta \tag{3.42}$$

The prior risk function is also called the Bayesian risk or simply the Bayes risk

Thus;

$$_B R_S(\hat{\theta}) = \int_{\theta=0}^{\infty} R_S(\hat{\theta})\lambda(\theta)d\theta \tag{3.43}$$

The values of $\lambda(\theta)$ and $R_S\left(\hat{\theta}\right)$ given in (3.2) and in Theorem 3.7 respectively are used in (3.43) to obtain

$$_B R_S(\hat{\theta}) = \int_{\theta=0}^{\infty} R_S(\hat{\theta})\lambda(\theta)d\theta = \int_{\theta=0}^{\infty} \frac{\theta^2}{n}\cdot\frac{\left(\frac{\mu}{\theta}\right)^{v+1} e^{\frac{-\mu}{\theta}}}{\mu\Gamma(v)}d\theta = \frac{\mu^{v+1}}{n\mu\Gamma(v)}\int_{\theta=0}^{\infty}\left(\frac{1}{\theta}\right)^{v-1} e^{\frac{-\mu}{\theta}}d\theta$$

which after simplification becomes

$$_B R_S(\hat{\theta}) = \frac{\mu^{v+1}}{n\mu\Gamma(v)}\cdot\frac{\Gamma(v-2)}{\mu^{v-2}} = \frac{\mu^2}{n(v-1)(v-2)} \quad \text{as required.}$$

Theorem 3.9

The risk function of $\hat{\theta}$ using Linex error loss function is given by $R_L\left(\hat{\theta}\right) = b\left[\frac{e^{-a\theta}}{\left(1-\frac{a\theta}{n}\right)^n} - 1\right]$, where $\hat{\theta}$ is the unique U.M.V. Unbiased Estimator of θ.

Proof

Using Linex error loss function as given in (2.4), we get

$$\begin{aligned} R_L\left(\hat{\theta}\right) &= be^{-a\theta}\left(\frac{\beta}{\theta}\right)^n \int_{x_1}\cdots\int_{x_n} e^{\frac{a}{n}\left(\sum_{i=1}^n x_i^\beta\right)}\prod_{i=1}^n x_i^{\beta-1} e^{\left(\frac{-1}{\theta}\right)\sum_{i=1}^n x_i^\beta}\,d\underline{x} \\ &- \frac{ab\beta^n}{n\theta^n}\int_{x_1}\cdots\int_{x_n}\sum_{i=1}^n x_i^\beta\prod_{i=1}^n x_i^{\beta-1} e^{\left(\frac{-1}{\theta}\right)\sum_{i=1}^n x_i^\beta}\,d\underline{x} + b(a\theta-1) \end{aligned} \tag{3.44}$$

Using integration by parts to integrate the integral

$$\int_{x_1}\cdots\int_{x_n}\prod_{i=1}^n x_i^{\beta-1} e^{-\sum_{i=1}^n x_i^\beta\left[\frac{1}{\theta}-\frac{a}{n}\right]}d\underline{x} \text{ , we obtain } \int_{x_1}\cdots\int_{x_n}\prod_{i=1}^n x_i^{\beta-1} e^{-\sum_{i=1}^n x_i^\beta\left[\frac{1}{\theta}-\frac{a}{n}\right]}d\underline{x} = \left[\frac{n\theta}{\beta(n-a\theta)}\right]^n \tag{3.45}$$

Similarly, using integration by parts to integrate the integral

$$\int_{x_1}\cdots\int_{x_n}\sum_{i=1}^n x_i^\beta\prod_{i=1}^n x_i^{\beta-1} e^{-\frac{1}{\theta}\sum_{i=1}^n x_i^\beta}\,d\underline{x}, \text{ we get } \int_{x_1}\cdots\int_{x_n}\sum_{i=1}^n x_i^\beta\prod_{i=1}^n x_i^{\beta-1} e^{-\frac{1}{\theta}\sum_{i=1}^n x_i^\beta}\,d\underline{x} = \frac{n\theta^{n+1}}{\beta^n} \tag{3.46}$$

Hence (3.45) and (3.46) are used in (3.44) to obtain;

$$R_L\left(\hat{\theta}\right)= be^{-a\theta}\left(\frac{\beta}{\theta}\right)^n\left[\frac{n\theta}{\beta(n-a\theta)}\right]^n-\frac{ab\beta^n}{n\theta^n}\left[\frac{n\theta^{n+1}}{\beta^n}\right]+b(a\theta-1)$$

which reduces to

$$R_L\left(\hat{\theta}\right)= b\left[\frac{e^{-a\theta}}{\left(1-\frac{a\theta}{n}\right)^n}-1\right]$$

Theorem 3.10

The Bayes risk of $\hat{\theta}$ using Linex error loss function is given by ${}_B R_L\left(\hat{\theta}\right)=\frac{b\mu^{(v+2)/2}a^{(v-2)/2}}{\Gamma(v)}\left[2\left(\frac{a}{\mu}\right)K_V\left(2\sqrt{a\mu}\right)+2a\sqrt{\frac{a}{\mu}}K_{V-1}\left(2\sqrt{a\mu}\right)+\frac{(n+1)a^2}{n}K_{V-2}\left(2\sqrt{a\mu}\right)\right]-b$, where $\hat{\theta}$ is the unique U.M.V. Unbiased Estimator of θ.

Proof

The prior risk function is also called the Bayesian risk or simply the Bayes risk. Thus

$${}_B R_L\left(\hat{\theta}\right)= E\left[R_L\left(\hat{\theta}\right)\right]=\int_{\theta=0}^{\infty}R_L\left(\hat{\theta}\right)\lambda(\theta)d\theta \tag{3.47}$$

Substituting the values of $\lambda(\theta)$ and $R_L\left(\hat{\theta}\right)$ given in (3.2) and Theorem (3.9) respectively to (3.47) we get

$${}_B R_L\left(\hat{\theta}\right)= b\int_{\theta=0}^{\infty}\left[\frac{e^{-a\theta}}{\left[1-a\theta/n\right)\right]^n}-1\right]\frac{(\mu/\theta)^{v+1}e^{\frac{-\mu}{\theta}}}{\mu\Gamma(v)}d\theta=\frac{b\mu^v}{\Gamma(v)}\int_{\theta=0}^{\infty}\frac{e^{-(a\theta+\mu/\theta)}}{\theta^{v+1}}\left[1-\frac{a\theta}{n}\right]^{-n}d\theta-b \tag{3.48}$$

The integral is solved as in (3.27) which then gives the final value of the integral as;

$${}_B R_L\left(\hat{\theta}\right)=\frac{b\mu^{(v+2)/2}a^{(v-2)/2}}{\Gamma(v)}\left[2\left(\frac{a}{\mu}\right)K_V\left(2\sqrt{a\mu}\right)+2a\sqrt{\frac{a}{\mu}}K_{V-1}\left(2\sqrt{a\mu}\right)+\frac{(n+1)a^2}{n}K_{V-2}\left(2\sqrt{a\mu}\right)\right]-b$$

4. Computation of Relative Efficiency and Comparison in Terms of Risk Functions

4.1 Introduction

In this section, we make comparisons of the obtained estimators in terms of risk functions of those under Linex loss and squared error loss function. Once Bayes estimators under Linex loss function and squared error loss function have been obtained, comparisons in terms of their risk functions have been made, their relative efficiencies are computed. Thus some conclusions based on computations and graphs regarding relative efficiencies for some effective intervals will help us to know what estimators performs better than alternative estimators in terms of effective interval relative to Linex loss function than those relative to squared error loss function.

4.2 Computation of Relative Efficiencies and Comparison in Terms of Risk Function of Weibull Bayesian Distribution

In this section, we will use some of the results obtained section three. We have obtained that the risk functions, denoted by $R_S\left(\hat{\theta}_M\right), R_L\left(\hat{\theta}_M\right), R_S\left(\hat{\theta}\right)$ and $R_L\left(\hat{\theta}\right)$, where the subscript L denotes risk relative to Linex error loss function and S denotes risk relative to squared error loss function. These risk functions are given in Theorems 3.3, 3.4, 3.7, and 3.9 respectively.

Let us define relative efficiencies of the estimator $\hat{\theta}_M$ with respect to $\hat{\theta}$ under the Linex and squared error loss function as follows:

$$RE_L\left(\hat{\theta}_M,\hat{\theta}\right)= R_L\left(\hat{\theta}\right)\Big/R_L\left(\hat{\theta}_M\right) \text{ and } RE_S\left(\hat{\theta}_M,\hat{\theta}\right)= R_S\left(\hat{\theta}\right)\Big/R_S\left(\hat{\theta}_M\right)$$

These relative efficiencies (RE) are functions of a, μ, θ, n, and v. For some sets of values of a, μ, θ, n, and v, the graphs of the relative efficiencies, plotted against θ are shown in Figs. 4.1 and 4.2.

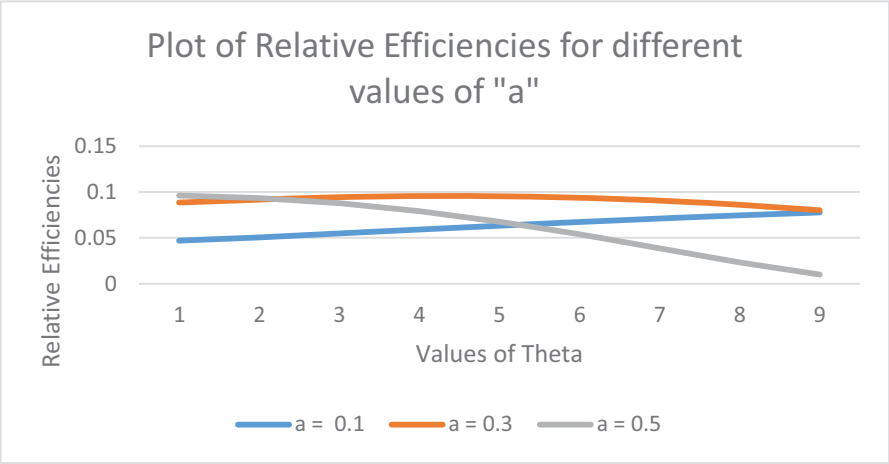

Figure 4.1. Plot of $RE_L(\hat{\theta}_M, \hat{\theta})$ for different values of "a" given n = 4, $\mu = 3$ and v = 0.5

From the above figure, we observe that for an increase in the magnitude of "a", ttytytyty

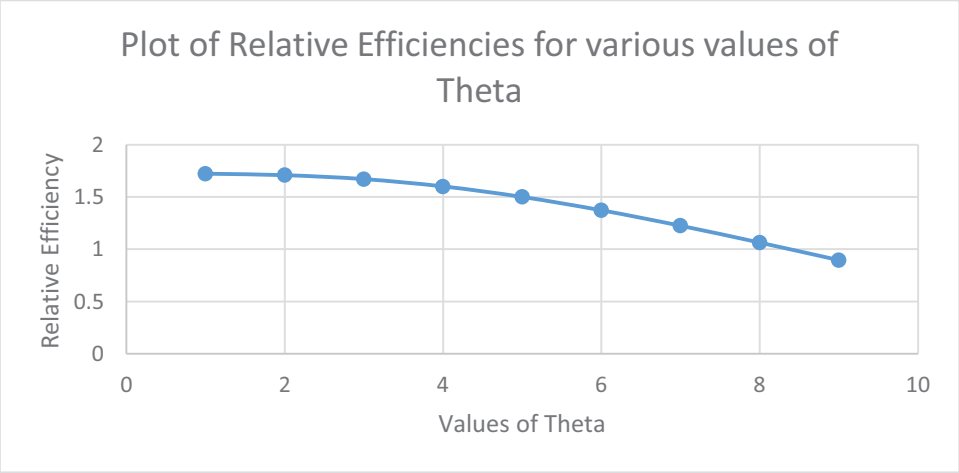

Figure 4.2. Plot of $RE_S(\hat{\theta}_M, \hat{\theta})$ for different values of "Theta" given n = 4, $\mu = 3$ and v = 0.5

From the above figure, for an increase in the values of θ, there is a decrease in the magnitude of RE.

Thus some conclusions based on graphs regarding effective interval reveals that for the Weibull distribution, $\hat{\theta}_M$ performs better than alternative estimators in terms of effective interval relative to squared error loss function than those relative to Linex error loss function.

5. Summary and Concluding Remarks

Asymmetric LINEX loss functions have been employed in the analysis of several central statistical estimation and prediction problems. Optimal estimators and predictors relative to LINEX loss and their associated risk functions have been derived. The analytical ease with which results can be obtained using asymmetric LINEX loss functions makes them attractive for use in applied problems and in assessing the effects of departures from assumed symmetric loss functions. For example, Pandey and Rai (1992), in the normal – mean problem, found that Bayes estimators relative to LINEX loss functions dominate the alternative estimators in terms of risk function and Bayes risk. They also found out that if σ^2 is unknown, the Bayes estimators are still preferable over alternative estimators. In the Weibull distribution, it was straight forward to derive an estimator that is optimal relative to LINEX loss function and to obtain its risk function and Bayes risk. Also, certain well known estimators, for example, the U.M.V. Unbiased Estimator of θ, that are admissible relative to squared error loss function were shown to be also admissible relative to LINEX loss function. As a referee Zellner (1973) has stated, "…the point that questions of admissibility may depend quite sensitively on

features of the loss function, such as symmetry, is not generally appreciated…" and implies that a lot more thought should be given to the choice of a loss function, rather than to blindly trust in squared error loss function", see Zellner (1973) for an analysis of the effects of errors in specifying loss functions on solutions to control problems. While the LINEX class of loss functions is convenient and useful, it is recognized that other asymmetric loss functions, for example, asymmetric linear and quadratic loss functions, are available and may be useful. Further study of the properties of alternative estimators relative to these and other types of asymmetric loss functions would be useful and is left for future research.

References

Aitchison, J., & Dunsmore, I. R. (1975). Statistical Prediction Analysis, *London: Cambridge University Press*. https://doi.org/10.1017/CBO9780511569647

Alexander, M. M. Introduction to the Theory of Mathematical Statistics, *Third ed. McGraw – Hill series in Probability and Statistics.*

Basu, A. P., & Ebrahim, N. (1988). Bayesian Approach to Life Testing and Reliability Estimation Using Asymmetric Loss Function. *Mathematical Science Technical Report No. 144. AFOSSR, Technical Report No. 88.*

Basu, A. P., & Klein, J. P. (1982). Some recent results in competing risks theory. *Lecture Notes-Monograph Series*, *2*, 216-229. https://doi.org/10.1214/lnms/1215464851

Berger and Sun (1993). Bayesian Analysis for the Poly – Weibull Distribution. *JASA, 88*(424). https://doi.org/10.1080/01621459.1993.10476426

Berger, J. O. (1980). Statistical Decision Theory: Foundations, Concepts and Methods, *New – York Springer – Verlag.* https://doi.org/10.1007/978-1-4757-1727-3_1

Bhattacharya, S. K. (1962). On a point Analogue used in a Life Test based on Weibull Distribution, *Australia Journal of Statistics*, *4,* 101-105. https://doi.org/10.1111/j.1467-842X.1962.tb00327.x

Bhattacharya, S. K. (1966). A Modified Bessel Function Model in Life Testing. *Metrika*, *11*, 133-144. https://doi.org/10.1007/BF02613584

Bhattacharya, S. K. (1967). Bayesian Approach to Life Testing and Reliability Estimation. *JASA, 62*(317), 48-62. https://doi.org/10.1080/01621459.1967.10482887

Bhattacharya, S. K., & Holla, M. S. (1965). On a Discrete Distribution with Special Reference to the Theory of Accident Proneness. *JASA, 60*, 1060-1066. https://doi.org/10.1080/01621459.1965.10480850

Canavos and Tsokos (1970). Bayesian Estimation of Life Parameters in the Weibull Distribution. *Operations Research, 45*, 24-31.

Canfield, R. V. (1970). A Bayesian Approach to Reliability Estimation Using a Loss Function, *I.E.E.E. Transaction on Reliability, 19*, 13-16. https://doi.org/10.1109/TR.1970.5216372

Cohen, A. C. (1965), Maximum Likelihood Estimation in the Weibull Distribution based on complete and Censored Samples, *Technometrics, 7*, 579-588. https://doi.org/10.1080/00401706.1965.10490300

Cox, D. R. (1959). The Analysis of Exponentially Distributed Lifetimes with two types of Failures. *The Journal of Royal Statistical Society, Ser. B., 21*, 411-421. https://doi.org/10.1111/j.2517-6161.1959.tb00349.x

David, H. A., & Moeschberger, M. L. (1978). The Theory of Competing Risks. *Griffin's Statistical Monographs and Courses,* No. 39, London: Charles. W. Griffin.

Drake, A. W. (1996). Bayesian Statistics for the reliability Engineer, *Proc. Annual Symposium on reliability,* 315-320.

Ehrenfeld, S. (1962). Some Experimental Design Problems in Life Testing. *JASA, 57*, 668-679. https://doi.org/10.1080/01621459.1962.10500555

Epstein, B., & Sobel, M. (1953). Life Testing. *JASA, 48*, 486-502. https://doi.org/10.1080/01621459.1953.10483488

Erdelyi, A. et.al. (1953). Higher Transcedental Functions. *Vol. II, McGraw Hill Book Company, inc. New – York.*

Ferguson, T. S. (1967). Mathematical Statistics: A decision Theoretical Approach, *New – York: Academic Press.*

Hogg, R. V., & Craig, A. T. (1956). Introduction to mathematical statistics, Pg.200-227.

Mendehall, W. (1958). A Bibliography on Life Testing and Related Topics. *Biometrika, 45*, 521-543. https://doi.org/10.1093/biomet/45.3-4.521

Mendenhall, W., & Hadel, R. J. (1958). Estimation of Parameters of Mixed Exponentially Distributed Failure Time

Distributions from Censored Life Test Data. *Biometrika, 45*, 504-520. https://doi.org/10.1093/biomet/45.3-4.504

Pandey, & Rai. (1992), Bayesian Estimation of Mean and Square of Mean of Normal Distribution using LINEX Loss Function, *Communication in Statistics, 21*(12), 3369-3391. https://doi.org/10.1080/03610929208830985

Searls, D. T. (1964). The Utilization of a Known Coefficient of Variation in the Estimation Procedure. *JASA,* 59, 1225-1226. https://doi.org/10.1080/01621459.1964.10480765

Sinha, S. K., & Kale, B. K. (1980). Life Testing and Reliability Estimation, *Wiley Eastern Limited, New Delhi.*

Tate, R. F. (1959). Unbiased Estimation: Functions of Location and Scale Parameters. *Ann. Math. Stat., 30*, 341-366. https://doi.org/10.1214/aoms/1177706256

Thompson, J. R. (1968). Some Shrinkage Techniques for Estimating the Mean. *JASA, 63*, 113-123. https://doi.org/10.2307/2283832

Thompson, J. R. (1968b). Accuracy Borrowing in the Estimation of Mean by Shrinkage to an Interval. *JASA, 63*, 953-963. https://doi.org/10.1080/01621459.1968.11009322

Varian, H. R. (1975). A Bayesian approach to real estate assessment. *Studies in Bayesian econometric and statistics in Honor of Leonard J. Savage*, 195-208.

Wetherill, G. B. (1961), Bayesian Sequential Analysis, *Biometrika, 48,* 281-292. https://doi.org/10.1093/biomet/48.3-4.281

Zellner, A. (1971).An Introduction to Bayesian Inference in Econometrics, *New – York, John Willey and sons, inc.*

Zellner, A. (1973). The quality of Quantitative Economic Policymaking When Targets and Costs of Change are Misspecified. *In Selected Readings in Econometrics and Economic Theory: Essays in Honour of Jan Tinbergen, ed. W. Sellekaerts, London: Macmillan,* 147-164. https://doi.org/10.1007/978-1-349-01936-6_7

Zellner, A. (1986), Bayes Estimation and Prediction Using Asymmetric Loss Functions. *JASA, 81*, 446-451. https://doi.org/10.1080/01621459.1986.10478289

International Journal of Statistics and Probability; Vol. 9, No. 2; March 2020
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Reviewer Acknowledgements

International Journal of Statistics and Probability wishes to acknowledge the following individuals for their assistance with peer review of manuscripts for this issue. Their help and contributions in maintaining the quality of the journal is greatly appreciated.

Many authors, regardless of whether *International Journal of Statistics and Probability* publishes their work, appreciate the helpful feedback provided by the reviewers.

Reviewers for Volume 9, Number 2

Abdullah A. Smadi, Yarmouk University, Jordan

Felix Almendra-Arao, UPIITA del Instituto Politécnico Nacional , México

Gane Samb Lo, University Gaston Berger, SENEGAL

Man Fung LO, Hong Kong Polytechnic University, Hong Kong

Noha Youssef, American University in Cairo, Egypt

Pablo José Moya Fernández, Universidad de Granada, Spain

Philip Westgate, University of Kentucky, USA

Vilda Purutcuoglu, Middle East Technical University (METU), Turkey

Vyacheslav Abramov, Swinburne University of Technology, Australia

Wei Zhang, The George Washington University, USA

Weizhong Tian, Eastern New Mexico University, USA

Wojciech Gamrot, University of Economics, Poland

Wendy Smith

On behalf of,

The Editorial Board of *International Journal of Statistics and Probability*

Canadian Center of Science and Education

ACADEMIC AND OPEN ACCESS

Canadian Center of Science and Education

CCSENET.ORG

➢ CALL FOR MANUSCRIPTS

International Journal of Statistics and Probability is a peer-reviewed journal, published by Canadian Center of Science and Education. The journal publishes research papers in all aspects of statistics and probability. The journal is available in electronic form in conjunction with its print edition. All articles and issues are available for free download online.

We are seeking submissions for forthcoming issues. All manuscripts should be written in English. Manuscripts from 3000–8000 words in length are preferred. All manuscripts should be prepared in LaTeX or MS-Word format, and submitted online, or sent to: ijsp@ccsenet.org

Paper Selection and Publishing Process

a) Submission acknowledgement. If you submit manuscript online, you will receive a submission acknowledgement letter sent by the online system automatically. For email submission, the editor or editorial assistant sends an e-mail of confirmation to the submission's author within one to three working days. If you fail to receive this confirmation, please check your bulk email box or contact the editorial assistant.

b) Basic review. The editor or editorial assistant determines whether the manuscript fits the journal's focus and scope. And then check the similarity rate (CrossCheck, powered by iThenticate). Any manuscripts out of the journal's scope or containing plagiarism, including self-plagiarism are rejected.

c) Peer Review. We use a double-blind system for peer review; both reviewers' and authors' identities remain anonymous. The submitted manuscript will be reviewed by at least two experts: one editorial staff member as well as one to three external reviewers. The review process may take four to ten weeks.

d) Make the decision. The decision to accept or reject an article is based on the suggestions of reviewers. If differences of opinion occur between reviewers, the editor-in-chief will weigh all comments and arrive at a balanced decision based on all comments, or a second round of peer review may be initiated.

e) Notification of the result of review. The result of review will be sent to the corresponding author and forwarded to other authors and reviewers.

f) Pay the article processing charge. If the submission is accepted, the authors revise paper and pay the article processing charge (formatting and hosting).

g) E-journal is available. E-journal in PDF is available on the journal's webpage, free of charge for download. If you need the printed journals by post, please order at http://www.ccsenet.org/journal/index.php/ijsp/store/hardCopies.

h) Publication notice. The authors and readers will be notified and invited to visit our website for the newly published articles.

More Information

E-mail: ijsp@ccsenet.org

Website: http://ijsp.ccsenet.org

Paper Submission Guide: http://ijsp-author.ccsenet.org

Recruitment for Reviewers: http://www.ccsenet.org/journal/index.php/ijsp/editor/recruitment

➢ JOURNAL STORE

To order back issues, please contact the journal editor and ask about the availability of journals. You may pay by credit card, PayPal, and bank transfer. If you have any questions regarding payment, please do not hesitate to contact the journal editor or editorial assistant.

Price: $40.00 USD/copy Shipping fee: $20.00 USD/copy

ABOUT CCSE

The Canadian Center of Science and Education (CCSE) is a private for-profit organization delivering support and services to educators and researchers in Canada and around the world.

The Canadian Center of Science and Education was established in 2006. In partnership with research institutions, community organizations, enterprises, and foundations, CCSE provides a variety of programs to support and promote education and research development, including educational programs for students, financial support for researchers, international education projects, and scientific publications.

CCSE publishes scholarly journals in a wide range of academic fields, including the social sciences, the humanities, the natural sciences, the biological and medical sciences, education, economics, and management. These journals deliver original, peer-reviewed research from international scholars to a worldwide audience. All our journals are available in electronic form in conjunction with their print editions. All journals are available for free download online.

Mission

To work for future generations

Values

Scientific integrity and excellence

Respect and equity in the workplace

CONTACT US

9140 Leslie St. Suite 110
Beaver Creek, Ontario, L4B 0A9
Canada
Tel: 1-416-642-2606
Fax: 1-416-642-2608
E-mail: info@ccsenet.org
Website: www.ccsenet.org

The journal is peer-reviewed
The journal is open-access to the full text
The journal is included in:

Aerospace Database
BASE (Bielefeld Academic Search Engine)
EZB (Elektronische Zeitschriftenbibliothek)
Google Scholar
JournalTOCs
Library and Archives Canada
LOCKSS
MIAR
PKP Open Archives Harvester
SHERPA/RoMEO
Standard Periodical Directory
Ulrich's

International Journal of Statistics and Probability

Bimonthly

Publisher	Canadian Center of Science and Education
Address	9140 Leslie St. Suite 110, Beaver Creek, Ontario, L4B 0A9, Canada
Telephone	1-416-642-2606
Fax	1-416-642-2608
E-mail	ijsp@ccsenet.org
Website	http://ijsp.ccsenet.org

ISSN 1927-7032