

## Original Article

### Discriminant Analysis involving Count Data

George Chinanu Mbaeyi\* and Chijioke Joel Nweke

Department of Mathematics and Statistics, Faculty of Physical Science, Alex Ekwueme Federal University Ndufu-Alike, 482131, Nigeria.

\*Corresponding author, email address: [george.chinanu@funai.edu.ng](mailto:george.chinanu@funai.edu.ng)

#### Abstract

One of the situations which give rise to violation of the normality assumption in discriminant analysis is that which involve count observations. For a two variable case involving count observations, this paper presents a new discriminant analysis approach when one variable is observed conditional on the other. Two cases involving Poisson-Binomial and Poisson-Poisson distributions were considered. The derived allocation rules are based on the resulting joint distribution of the two count variables. Applicability of the suggested allocation rules in discriminant analysis involving count data and its performance in comparison with Fisher linear discriminant rule was studied under different conditions. Results obtained shows promising applicability of the suggested allocation rules when compared with the Fisher linear discriminant method.

**Keyword:** Count data; discriminant analysis; error rate; allocation rules; Poisson distribution; Binomial distribution

#### 1. Introduction

Normality assumption is fundamental for useful inference in linear discriminant analysis. Unfortunately, most real-life situations are generated by mechanism that violates this assumption. One of such real-life phenomena are those involving discrete distribution that are amenable to count data. The commonest probability distributions often used in analyzing count data is the Poisson and negative Binomial distributions (Witten, 2011). In analyzing count data, Poisson variates are mostly faced with the problem of under- (or over) dispersion (Inouye, Yang, Allen, &

30 Ravikumar, 2017). With Poisson variates in discriminant analysis, the authors are of the opinion  
31 that it may no longer be appropriate to use the linear discriminant analysis approach whose  
32 allocation rule is derived based on normality assumption. An optimal approach for discriminant  
33 analysis in this situation would be to derive the allocation rule based on the originating distribution  
34 (Mbaeyi & Nweke, 2021), and not the assumed normal distribution. Some studies concerning  
35 classification analysis involving count data has been noted in literature. Examples are; Poisson  
36 linear discriminant analysis (Witten 2011), Negative Binomial linear discriminant analysis (Dong,  
37 Zhao, Tong, & Wan, 2016), zero inflated Poisson logistic discriminant analysis (Zhou, Wan,  
38 Zhang, & Tong, 2018), zero inflated negative Binomial logistic discriminant analysis (Zhu, Yuan,  
39 Shu, Liao, Zhao, & Zhou, 2021) and decision tree model for count data (Wah, Nasaruddin, Voon,  
40 & Lazim, 2012). These methods are either based on no (or very weak) assumption, or some  
41 inconsistent transformations. In addition, conditional relationship between variables were not  
42 adequately considered.

43 A typical scenario which this work attempts to consider is, for example, road accident  
44 occurrence which may lead to one or more casualties. The casualties are characterized by the extent  
45 of physical injury or fatality. Let  $X_i$  be the number of accidents in a given location for a given  
46 interval.  $X_i$  is assumed to follow Poisson distribution with parameter  $\theta$ . Suppose that the variable  
47  $Y_i$  assumes the value 1 with probability  $p$  if the  $i$ th accident is fatal, and the value 0 with probability  
48  $q = 1 - p$  if the  $i$ th accident is not fatal, then  $Y = Y_1 + Y_2 + \dots + Y_X$  represents the number of fatal  
49 accidents out of a total of  $X$  accidents. In this case,  $Y_i$  is better represented by a Binomial  
50 distribution with parameter  $p$ . It is expected that  $Y \leq X$  and the bivariate distribution  $f(x, y)$   
51 represents the joint distribution of number of accidents and cases of fatal accident. One may also  
52 consider jointly the number of accidents  $X_i$  and the corresponding number of casualties  $Z_i$  in which

53 case both  $X_i$  and  $Z_i$  follows Poisson distribution. The number of accidents together with the  
54 corresponding casualty indices can be classified as having resulted from one of either a clear or a  
55 cloudy weather.

## 56 **2. Methodology**

57 Based on the typical scenario described in section I which this work attempts to focus, the  
58 Poisson and Binomial discrete distributions shall be considered. Joint distribution of Poisson-  
59 Binomial and Poisson-Poisson distribution shall be considered and, on the basis of the derived  
60 joint distribution, allocation rule for classifying observation  $w = (x, y)$  shall then be obtained and  
61 applied consequently.

### 62 **2.1 Poisson-Binomial**

63 Let  $X$  be a Poisson random variable with parameter  $\lambda$  and  $Y$  be a Binomial random variable with  
64 parameter  $n, p$  and  $q$ . That is,

$$65 \quad f(x) = \frac{e^{-\lambda}\lambda^x}{x!}; \quad \lambda > 0, x = 0,1,2, \dots \quad (1)$$

66 and

$$67 \quad f(y) = \binom{n}{y} p^y q^{n-y}; y = 0,1,2, \dots, n \quad (2)$$

68 where  $n$  is the number of times event of interest was observed,  $p$  is the probability of success out  
69 of  $n$  observed events,  $q$  is the probability of failure out of the  $n$  observed events and  $\lambda$  is the mean  
70 number of events.  $y$  is a realization of observations with defined attribute when observing  $x$  such  
71 that  $P(y_i = 1) = p$  whenever the attribute is present in  $x$  and  $P(y_i = 0) = q = 1 - p$  whenever  
72 the attribute is not present in  $x$ . It follows that  $y$  observation is made given that  $x$  is observed

73 already. Hence, we can now define the conditional distribution of  $y$  given  $x$  as (Ramachandran &  
74 Tsokos, 2009)

$$75 \quad f(y|x) = \binom{x}{y} p^y q^{x-y}; y = 0,1,2, \dots, x \quad (3)$$

76 Combining (1) and (3), the joint distribution of  $x$  and  $y$  is given as

$$77 \quad f(x, y) = \frac{e^{-\lambda} \lambda^x p^y q^{x-y}}{(x-y)! y!}; x = 0,1,2, \dots; y = 0,1,2, \dots, x \quad (4)$$

78 **The marginal of (4) is a discrete Poisson probability mass function (See Appendix for proof)**

79 Following Kendall & Stuart (1967), it can easily be shown that for (3) and (4),

$$80 \quad E(y|x) = xp \quad (5)$$

$$81 \quad Var(y|x) = xpq \quad (6)$$

$$82 \quad E(x, y) = p\lambda(\lambda + 1) \quad (7)$$

$$83 \quad Cov(x, y) = p\lambda \quad (8)$$

$$84 \quad \rho(x, y) = +\sqrt{p} \quad (9)$$

85 The maximum likelihood estimates of the parameters of (4) are  $\lambda = \bar{x}$  and  $p = \frac{\bar{y}}{\bar{x}}$ .

86 For the purpose of classification, optimal allocation rule for classifying observation  $\mathbf{w} = (x, y)$   
87 into group  $D_1$  and group  $D_2$  can be derived using (4). Let  $L_1(x, y, \lambda_1, p_1)$  and  $L_2(x, y, \lambda_2, p_2)$  be  
88 the likelihood function of (4) for group  $D_1$  and group  $D_2$  respectively. According to Anderson  
89 (1958), observation  $\mathbf{w} = (x, y)$  can be classified as belonging to group  $D_1$  if  $L_1(x, y, \lambda_1, p_1) \geq$   
90  $L_2(x, y, \lambda_2, p_2)$ . That is, if

91 
$$R: \left(\frac{\lambda_1}{\lambda_2}\right)^x \left(\frac{p_1}{p_2}\right)^y \exp(\lambda_2 - \lambda_1) \geq 1 \quad (10)$$

92 By taking logarithm of (10) and simplifying, we have that

93 
$$R: x \ln \left(\frac{\lambda_1}{\lambda_2}\right) + y \ln \left(\frac{p_1}{p_2}\right) \geq (\lambda_1 - \lambda_2) \quad (11)$$

94 Let the prior probability of observation falling into group  $D_1$  and group  $D_2$  be  $\pi_1$  and  $\pi_2$   
 95 ( $\pi_1 + \pi_2 = 1$ ) respectively, then the bayes rule can obtained by comparing  $\pi_i L_i(x, y, \lambda_i, p_i)$ ,  $i =$   
 96 1,2 in which case we are to allocate observation  $\mathbf{w} = (x, y)$  to group  $D_1$  if

97 
$$R: x \ln \left(\frac{\lambda_1}{\lambda_2}\right) + y \ln \left(\frac{p_1}{p_2}\right) \geq \ln \left(\frac{\pi_2}{\pi_1}\right) + (\lambda_1 - \lambda_2) \quad (12)$$

98 Otherwise, observation  $\mathbf{w} = (x, y)$  is allocated to group  $D_2$ . Where equal prior probability is  
 99 assumed for group  $D_1$  and  $D_2$ , (12) remains as in (11).

## 100 2.2 Poisson-Poisson

101 Let  $k$  and  $r$  be two available variables for consideration in a discriminant analysis. Both  $k$  and  $r$   
 102 are independent Poisson variates with parameters  $\lambda$  and  $\theta$  respectively.

103 
$$g(k) = \frac{e^{-\lambda} \lambda^k}{k!}; \quad \lambda > 0, k = 0, 1, 2, \dots \quad (13)$$

104 and

105 
$$g(r) = \frac{e^{-\theta} \theta^r}{r!}; \quad \theta > 0, r = 0, 1, 2, \dots \quad (14)$$

106 Observation  $r$  are made given that  $k$  has been observed. Thus, the conditional distribution of  $r$   
 107 given that  $k$  has been observed is given as

108 
$$g(r|k) = \frac{e^{-\theta k} (\theta k)^r}{r!}; \quad r = 0, 1, 2, \dots \quad (15)$$

109 As above, combining (13) and (15), the joint distribution of  $r$  and  $k$  is given in (16) as

$$110 \quad g(k, r) = \frac{\lambda^k (\theta k)^r \exp\{-(\lambda + \theta k)\}}{k! r!}; \lambda, \theta > 0, k = 0, 1, 2, \dots; r = 0, 1, 2, \dots \quad (16)$$

111 As in (4), the marginal of (16) is also a univariate Poisson probability mass function (See Appendix  
112 for proof)

113 Some properties of (15) and (16) are readily given as

$$114 \quad E(r|k) = \theta k \quad (17)$$

$$115 \quad Var(r|k) = \theta k \quad (18)$$

$$116 \quad E(k, r) = \theta \lambda (\lambda + 1) \quad (19)$$

$$117 \quad Cov(k, r) = \theta \lambda \quad (20)$$

$$118 \quad \rho(k, r) = \sqrt{\frac{\theta}{\theta + 1}} \quad (21)$$

119 The maximum likelihood estimates of the parameters of (16) are also given as  $\lambda = \bar{k}$  and  $\theta = \frac{\bar{r}}{\bar{k}}$ .

120 To perform discriminant analysis using the values of the variables  $r$  and  $k$ , the optimal allocation  
121 rule is best derived using the joint distribution in (16). By defining  $L_1(k, r, \lambda_1, \theta_1)$  and  
122  $L_2(k, r, \lambda_2, \theta_2)$  to be the likelihood functions of (16) for group  $D_1$  and group  $D_2$  respectively, the  
123 allocation rule will be to allocate observation  $\mathbf{w} = (k, r)$  to group  $D_1$  be classified as belonging to  
124 group  $D_1$  if  $L_1(k, r, \lambda_1, \theta_1)/L_2(k, r, \lambda_2, \theta_2) \geq 1$ . That is, if

$$125 \quad R: \left(\frac{\lambda_1}{\lambda_2}\right)^k \left(\frac{\theta_1}{\theta_2}\right)^r \exp(\lambda_2 - \lambda_1) \geq 1 \quad (22)$$

126 By also taking logarithm of (22) and simplifying, we have that

127 
$$R: k \ln \left( \frac{\lambda_1}{\lambda_2} \right) + r \ln \left( \frac{\theta_1}{\theta_2} \right) \geq (\lambda_1 - \lambda_2) \quad (23)$$

128 Let the prior probability of observation falling into group  $D_1$  and group  $D_2$  be  $\pi_1$  and  $\pi_2$   
 129 ( $\pi_1 + \pi_2 = 1$ ) respectively, then the bayes rule can be obtained by comparing  $\pi_i L_i(k, r, \lambda_i, \theta_i)$ ,  $i =$   
 130 1,2 in which case we are to allocate observation  $\mathbf{w} = (k, r)$  to group  $D_1$  if

131 
$$R: k \ln \left( \frac{\lambda_1}{\lambda_2} \right) + r \ln \left( \frac{\theta_1}{\theta_2} \right) \geq \ln \left( \frac{\pi_2}{\pi_1} \right) + (\lambda_1 - \lambda_2) \quad (24)$$

132 Otherwise, observation  $\mathbf{w} = (k, r)$  is allocated to group  $D_2$ . Where equal prior probability is  
 133 assumed for group  $D_1$  and  $D_2$ , (24) remains as in (23).

134 **2.3 Fisher Linear Discriminant Analysis**

135 Fisher linear discriminant (FLD) analysis is a generalization of linear discriminant analysis, a  
 136 method commonly used to find a linear combination of attributes that characterizes or separates  
 137 two or more groups of objects. Given a set of independent multivariate observations, the FLD  
 138 assumes that observations are distributed multivariate normal with vector of means which varies  
 139 in each group but a covariance matrix that is common across all groups. FLD rule maximizes the  
 140 ratio between sum of squares between and sum of squares within and then finds a linear  
 141 combination of the predictors to predict group membership. Typically, given a vector  $\mathbf{x}$  of  
 142 observations assumed to be multivariate normal with mean  $\boldsymbol{\mu}_i (i = 1, 2)$  and covariance matrix  $\boldsymbol{\Sigma}$   
 143 coming from one of two groups and assuming equal a prior probability of group membership, the  
 144 FLD allocation rule is to classify observation  $\mathbf{x}$  as belonging to group  $G_1$  if

145 
$$\mathbf{x}' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \geq \frac{1}{2} (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$$

146 It has been argued that discriminant analysis is relatively robust to violations of normality and  
147 homoscedastic assumptions (Hardle & Simar, 2007). Hence, in line with the focus of this work,  
148 this study shall consider application of FLD to count data.

### 149 3. Analysis, Results and Discussion

150 In order to demonstrate the applicability of the allocation rules presented in section (2), both  
151 artificial and real-life data was considered. For the artificial data, random samples of Poisson and  
152 Binomial variates were generated under various sample sizes ( $n_1=n_2=20, 50, 100, 150, 200, 300,$   
153  $500, 750, 1000, 1500, 2000$ ) for a two-group discriminant analysis. In order to evaluate the  
154 allocation rules in (12) and (24),  $\lambda_1 = 1.2, \lambda_2 = 0.8, \theta_1 = 0.42, \theta_2 = 0.31, p_1 = 0.42$  and  $p_2 =$   
155  $0.31$  were used in generating Poisson and Binomial data.  $\lambda_1$  and  $\lambda_2$  are Poisson parameters for  
156 variable  $x_1$  in groups  $D_1$  and  $D_2$ . Similarly,  $\theta_1$  and  $\theta_2$  are Poisson parameters for variable  $x_2$  in  
157 groups  $D_1$  and  $D_2$  while  $p_1$  and  $p_2$  are Binomial parameters for variable  $x_2$  in groups  $D_1$  and  $D_2$   
158 respectively. As a way of introducing over-dispersion and under-dispersion into the dataset, 20%  
159 of each of  $n_1$  and  $n_2$  was generated by replacing  $\lambda_1$  and  $\lambda_2$  by  $9.5\lambda_1$  and  $9.5\lambda_2$  respectively to  
160 obtain over-dispersed data while under-dispersion was introduced by replacing  $\lambda_1$  and  $\lambda_2$  by  
161  $0.125\lambda_1$  and  $0.125\lambda_2$  respectively. The Poisson-Binomial (P-B) and Poisson-Poisson (P-P) based  
162 allocation rules were then applied to the generated data for a two-group discriminant analysis. For  
163 comparison purpose, the Fisher linear discriminant (FLD) analysis procedure was also applied to  
164 the generated data. Error rates resulting from the P-B, P-P and FLD allocation rules were obtained  
165 and presented when the data are (i) under-dispersed, (ii) undispersed, and (iii) over-dispersed.  
166 Error rate is measure of misclassification made by any given allocation rule. It is simply obtained  
167 by dividing the sum of misclassified observations in groups  $D_1$  and  $D_2$  by the total number of  
168 observations.



169 For the real-life data, record of accidents in the UK was extracted from  
170 [https://www.kaggle.com/datasets/benoit72/uk-accidents-10-years-history-with-many-](https://www.kaggle.com/datasets/benoit72/uk-accidents-10-years-history-with-many-variables?select=Vehicles0514.csv)  
171 [variables?select=Vehicles0514.csv](https://www.kaggle.com/datasets/benoit72/uk-accidents-10-years-history-with-many-variables?select=Vehicles0514.csv) and analyzed as in two-group discriminant analysis. The data  
172 contains many variables related to road accident in UK between 2005 and 2014 some of which  
173 include weather, location, latitude/longitude, accident severity, casualty severity, casualty type,  
174 age of casualty, age of driver, age of victim, etc. Number of accidents (x) and number of casualties  
175 (y) were used for the P-P approach while number of casualties (k) and incidences of fatal casualties  
176 (r) were used for the P-B approach. After extraction,  $n = 20$  observed cases of accidents classified  
177 under “Clear” and “Cloudy” weather for the period 2005 – 2014 were available for analysis. All  
178 analysis was performed in R (2020) programming environment and the results presented in Table  
179 1 and Table 2 below

180 Results presented in Table 1 and Table 2 above shows the applicability of the suggested allocation  
181 rules. Since the problem of dispersion is common with Poisson variates, it suffices to consider  
182 various forms of dispersion in the application. Generally, in terms of error rate as presented in  
183 Table 1, P-P performed better than FLD both when there is under-dispersion and no dispersion in  
184 the dataset. However, with over-dispersed data, FLD fairly outperformed P-P. Error rates resulting  
185 from FLD in Table 1 appears to show slight evidence of stability regardless the form of dispersion  
186 present in the datasets as the error rates were between approximately 0.4100 and 0.4900. From the  
187 results presented in Table 2, FLD has an overall better performance than the P-B approach. Again,  
188 a fairly case-wise stability in error rate was noted for both P-B and FLD. Hence, effect of changing  
189 sample sizes appears not to noticeably affect the error rate of both P-B and FLD but this appears  
190 not to be the case with the various forms of dispersion considered. Moreso, introduction of  
191 binomial variate for the P-B analysis appears to have slightly improved the error rates of FLD in

192 Table 2 when compared with those of FLD in Table 1 but such was not the case for the P-B  
193 approach. In Table 1 and Table 2, error rates from P-P, P-B and FLD showed no trend with respect  
194 to increasing/decreasing sample sizes, hence, it may not be provable to infer that  
195 increasing/decreasing sample size improves error rate of any of the approaches considered in this  
196 work. On a general note, the lower error rates from FLD may not be valid enough to justify its  
197 usage in discriminant analysis when available data are not normally distributed. In case the  
198 argument persists, the optimal answer would depend on a choice between having a lower error rate  
199 using an incorrect methodology or having a moderate error rate using a correct methodology.

200 For the real-life data, all the approaches considered performed almost equally in terms of their  
201 respective error rates. When the data involving number of accidents ( $x$ ) and number of casualties  
202 ( $y$ ) was analyzed, error rates were 0.4880 and 0.5000 for P-P and FLD respectively. Similarly,  
203 with the data involving number of casualties ( $k$ ) and incidences of fatal casualties ( $r$ ),  
204 corresponding error rate was 0.4920 and 0.5000 for P-B and FLD respectively. Unlike in the  
205 artificial datasets, error rates from FLD were not better than P-P and P-B. However, it can easily  
206 be observed that the former looks similar with the result in Table 1 when  $n = 20$  while the latter  
207 is somewhat not similar with that of  $n = 20$  in Table 2. Hence, whereas applicability of the various  
208 approaches has been demonstrated with a real-life dataset, more cases of real-life scenario need to  
209 be analyzed in order to make valid inference regarding these approaches and their application in  
210 real-world data analysis.

#### 211 **4. Conclusion**

212 This paper has presented allocation rule based on discrete distribution amenable to count data. The  
213 allocation rules suggested in this paper demonstrated straightforward applicability and ability to  
214 handle cases of under-dispersion and over-dispersion in Poisson variates. The suggested allocation

215 rules are amenable to simple error rate estimation procedures, copes with the problem of small  
216 available samples and its implementation are easy in any user-defined programme package. The  
217 performance of the suggested allocation rule in comparison with the Fisher linear discriminant  
218 analysis is an indication that appreciable level of accuracy can be gained when allocation rules are  
219 derived based on the originating distribution of the data.

## 220 **References**

- 221 Dong, K., Zhao, H., Tong, T., & Wan, X. (2016). Negative binomial linear discriminant analysis  
222 for RNA-seq data. *BMC Bioinformatics*. doi: 10.1186/s12859-016-1208-1
- 223 Hardle, W, & Simar L. (2007). *Applied Multivariate Statistical Analysis*. Berlin Heidelberg,  
224 Springer.
- 225 Inouye, D. I., Yang, E., Allen, G. I., & Ravikumar, P. (2017). A review of multivariate distributions  
226 for count data derived from the Poisson distribution. *Computational Statistics*, 9(3),  
227 e1398. <https://doi.org/10/1002/wics.1398>
- 228 Kendall, M. G., & Stuart, A. (1967). *The advanced theory of statistics*. London, UK: Griffin and  
229 Co.
- 230 Mbaeyi, G. C. & Nweke, C. J. (2021). Discriminant analysis with non normal variables.  
231 Communication in Statistics – Theory and Methods. doi: 10.1080/03610926.2021.1908563
- 232 Ramachandran, K. M. & Tsokos, C. P. (2009). *Mathematical Statistics with Applications*. UK:  
233 Elsevier Academic Press.
- 234 R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for  
235 Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>

236 Wah, Y. B., Nasaruddin, N., Voon, W. S. & Lazim, M. A. (2012). Decision tree model for count  
 237 data. *Proceedings of the World congress in Engineering*, 1, 330-335. Retrieved from  
 238 [http://www.iaeng.org/publication/WCE2012/WCE2012\\_pp330-335.pdf](http://www.iaeng.org/publication/WCE2012/WCE2012_pp330-335.pdf)  
 239 Witten, D. (2011). Classification and clustering of sequencing data using a Poisson model. *Annals*  
 240 *Applied Statistics*, 5, 2493-2518. doi: 10.1214/11-AOAS493.  
 241 Zhou, Y., Wan, X., Zhang, B. & Tong, T. (2018). Classifying next-generation sequencing data  
 242 using a zero-inflated Poisson model. *Bionformatics*, 34, 1329-1335. doi:  
 243 10.1093/bioinformatics/btx768  
 244 Zhu, J., Yuan, Z., Shu, L., Liao, W., Zhao, M. & Zhou, Y. (2021). Selecting classification methods  
 245 for small samples of next-generation sequencing data. *Frontiers in Genetics*, 12, doi:  
 246 10.3389/fgene.2021.6422

247

248 **Appendix**

249 **Marginal of P-B**

250 
$$f(x, y) = \frac{e^{-\lambda} \lambda^x p^y q^{x-y}}{(x-y)! y!} = \binom{x}{x!} \left( \frac{e^{-\lambda} \lambda^x p^y q^{x-y}}{(x-y)! y!} \right) = \frac{\lambda^x e^{-\lambda}}{x!} \binom{x}{y} p^y q^{x-y}$$

251 
$$f_X(x) = \frac{f(x, y)}{f(y|x)} = \frac{\frac{\lambda^x e^{-\lambda}}{x!} \binom{x}{y} p^y q^{x-y}}{\binom{x}{y} p^y q^{x-y}} = \frac{\lambda^x e^{-\lambda}}{x!}; x = 0, 1, 2, \dots$$

252 **Marginal of P-P**

253 
$$g_k(k) = \frac{g(k, r)}{g(r|k)} = \frac{\frac{\lambda^k (\theta k)^r \exp\{-(\lambda + \theta k)\}}{k! r!}}{\frac{e^{-\theta k} (\theta k)^r}{r!}} = \frac{\lambda^k \exp\{-\lambda\}}{k!}; k = 0, 1, 2$$

254

255 R Source Code

```
256 library(MASS)
257 ## Over-Dispersed P-P and FLD
258 Overdispersed<-function(n1,n2){
259   set.seed(200)
260   lambda1<-1.2
261   lambda2<-0.8
262   theta1<-0.42
263   theta2<-0.31
264   xx<-rpois(0.8*n1,lambda1)
265   x<-rpois(0.2*n1,lambda1*9.5)
266   x11<-c(x,xx)
267   x21<-rpois(n1,theta1)
268   XX<-rpois(0.8*n2,lambda2)
269   X<-rpois(0.2*n2,lambda2*9.5)
270   x12<-c(XX,X)
271   x22<-rpois(n2,theta2)
272   g1<-rep(1,n1)
273   g2<-rep(2,n2)
274   d1<-cbind(x11,x21,g1)
275   d2<-cbind(x12,x22,g2)
276   Grp1<-cbind(x11,x21)
277   Grp2<-cbind(x12,x22)
278   c<-Grp1[,1]
279   cc<-Grp1[,2]
280   C<-Grp2[,1]
281   CC<-Grp2[,2]
282   Data<-rbind(d1,d2)
283   g<-Data[,3]
284   x1<-Data[,1]
285   x2<-Data[,2]
286   Dt<-data.frame(g,x1,x2)
287   fit<-lda(formula=g~.,data=Dt)
288   tabel<-table(actual=Dt$g,predicted=predict(fit,Dt)$class)
289   pi1<-0.5
290   pi2<-0.5
291   D<-log(pi2/pi1)+(lambda1-lambda2)
292   a <- 0
293   for(i in 1:length(Grp1[,1])){
```

```

294  A<-cc[i]*(log(lambda1/lambda2)+theta2-theta1)
295  B<-c[i]*log(theta1/theta2)
296  w1 <- A+B
297  if(w1>=D) {
298  a <- a+1
299  }
300  }
301  b <- 0
302  for(i in 1:length(Grp1[,1])){
303  A<-cc[i]*(log(lambda1/lambda2)+theta2-theta1)
304  B<-c[i]*log(theta1/theta2)
305  w1 <- A+B
306  if(w1<D) {
307  b <- b+1
308  }
309  }
310  c <- 0
311  for(i in 1:length(Grp2[,1])){
312  A<-CC[i]*(log(lambda1/lambda2)+theta2-theta1)
313  B<-C[i]*log(theta1/theta2)
314  w2 <- A+B
315  if(w2>=D) {
316  c <- c+1
317  }
318  }
319  d <- 0
320  for(i in 1:length(Grp2[,1])){
321  A<-CC[i]*(log(lambda1/lambda2)+theta2-theta1)
322  B<-C[i]*log(theta1/theta2)
323  w2 <- A+B
324  if(w2<D) {
325  d <- d+1
326  }
327  }
328  tr <- ftable(tabel)
329  ErFLD <- (tr[1,2]+tr[2,1])/(n1+n2)
330  ErPP <- (b + c)/(n1+n2)
331  list(c(ErPP,ErFLD))
332  #END
333  }
334

```

335

336

```
337  ## Under-Dispersed P-P and FLD
338  Underdispersed<-function(n1,n2){
339    set.seed(200)
340    lambda1<-1.2
341    lambda2<-0.8
342    theta1<-0.42
343    theta2<-0.31
344    xx<-rpois(0.8*n1,lambda1)
345    x<-rpois(0.2*n1,lambda1/8)
346    x11<-c(x,xx)
347    x21<-rpois(n1,theta1)
348    XX<-rpois(0.8*n2,lambda2)
349    X<-rpois(0.2*n2,lambda2/8)
350    x12<-c(XX,X)
351    x22<-rpois(n2,theta2)
352    g1<-rep(1,n1)
353    g2<-rep(2,n2)
354    d1<-cbind(x11,x21,g1)
355    d2<-cbind(x12,x22,g2)
356    Grp1<-cbind(x11,x21)
357    Grp2<-cbind(x12,x22)
358    c<-Grp1[,1]
359    cc<-Grp1[,2]
360    C<-Grp2[,1]
361    CC<-Grp2[,2]
362    Data<-rbind(d1,d2)
363    g<-Data[,3]
364    x1<-Data[,1]
365    x2<-Data[,2]
366    Dt<-data.frame(g,x1,x2)
367    fit<-lda(formula=g~.,data=Dt)
368    tabel<-table(actual=Dt$g,predicted=predict(fit,Dt)$class)
369    pi1<-0.5
370    pi2<-0.5
371    D<-log(pi2/pi1)+(lambda1-lambda2)
372
373    a <- 0
374    for(i in 1:length(Grp1[,1])){
375      A<-cc[i]*(log(lambda1/lambda2)+theta2-theta1)
376      B<-c[i]*log(theta1/theta2)
377      w1 <- A+B
```

```

378  if(w1<D) {
379  a <- a+1
380  }
381  }
382  b <- 0
383  for(i in 1:length(Grp1[,1])){
384  A<-cc[i]*(log(lambda1/lambda2)+theta2-theta1)
385  B<-c[i]*log(theta1/theta2)
386  w1 <- A+B
387  if(w1>=D) {
388  b <- b+1
389  }
390  }
391  c <- 0
392  for(i in 1:length(Grp2[,1])){
393  A<-CC[i]*(log(lambda1/lambda2)+theta2-theta1)
394  B<-C[i]*log(theta1/theta2)
395  w2 <- A+B
396  if(w2>=D) {
397  c <- c+1
398  }
399  }
400  d <- 0
401  for(i in 1:length(Grp2[,1])){
402  A<-CC[i]*(log(lambda1/lambda2)+theta2-theta1)
403  B<-C[i]*log(theta1/theta2)
404  w2 <- A+B
405  if(w2<D) {
406  d <- d+1
407  }
408  }
409  tr <- ftable(tabel)
410  ErFLD <- (tr[1,2]+tr[2,1])/(n1+n2)
411  ErPP <- (b + c)/(n1+n2)
412  list(c(ErPP,ErFLD))
413  #END
414  }
415
416  ## Undispersed P-P and FLD
417  Undispersed<-function(n1,n2){
418  set.seed(200)
419  lambda1<-1.2

```



```

420 lambda2<-0.8
421 theta1<-0.42
422 theta2<-0.31
423 x11<-rpois(n1,lambda1)
424 x21<-rpois(n1,theta1)
425 x12<-rpois(n2,lambda2)
426 x22<-rpois(n2,theta2)
427 g1<-rep(1,n1)
428 g2<-rep(2,n2)
429 d1<-cbind(x11,x21,g1)
430 d2<-cbind(x12,x22,g2)
431 Grp1<-cbind(x11,x21)
432 Grp2<-cbind(x12,x22)
433 c<-Grp1[,1]
434 cc<-Grp1[,2]
435 C<-Grp2[,1]
436 CC<-Grp2[,2]
437 Data<-rbind(d1,d2)
438 g<-Data[,3]
439 x1<-Data[,1]
440 x2<-Data[,2]
441 Dt<-data.frame(g,x1,x2)
442 fit<-lda(formula=g~.,data=Dt)
443 tabel<-table(actual=Dt$g,predicted=predict(fit,Dt)$class)
444 pi1<-0.5
445 pi2<-0.5
446 D<-log(pi2/pi1)+(lambda1-lambda2)
447 a <- 0
448 for(i in 1:length(Grp1[,1])){
449 A<-cc[i]*(log(lambda1/lambda2)+theta2-theta1)
450 B<-c[i]*log(theta1/theta2)
451 w1 <- A+B
452 if(w1<D) {
453 a <- a+1
454 }
455 }
456 b <- 0
457 for(i in 1:length(Grp1[,1])){
458 A<-cc[i]*(log(lambda1/lambda2)+theta2-theta1)
459 B<-c[i]*log(theta1/theta2)
460 w1 <- A+B
461 if(w1>=D) {
462 b <- b+1
463 }
464 }
465 c <- 0
466 for(i in 1:length(Grp2[,1])){
467 A<-CC[i]*(log(lambda1/lambda2)+theta2-theta1)

```

```

468 B<-C[i]*log(theta1/theta2)
469 w2 <- A+B
470 if(w2>=D) {
471 c <- c+1
472 }
473 }
474 d <- 0
475 for(i in 1:length(Grp2[,1])){
476 A<-CC[i]*(log(lambda1/lambda2)+theta2-theta1)
477 B<-C[i]*log(theta1/theta2)
478 w2 <- A+B
479 if(w2<D) {
480 d <- d+1
481 }
482 }
483 tr <- ftable(tabel)
484 ErFLD <- (tr[1,2]+tr[2,1])/(n1+n2)
485 ErPP <- (b + c)/(n1+n2)
486 list(c(ErPP,ErFLD))
487 #END
488 }
489

```

```

490 ## Overdispersed P-B and FLD
491 Overdispersed<-function(n1,n2){
492 set.seed(200)
493 N<-10
494 lambda1<-1.2
495 lambda2<-0.8
496 p1<-0.42
497 p2<-0.31
498 xx<-rpois(0.8*n1,lambda1)
499 x<-rpois(0.2*n1,lambda1*9.5)
500 x1<-c(xx,x)
501 x21<-rbinom(n1,N,p1)
502 XX<-rpois(0.8*n2,lambda2)
503 X<-rpois(0.2*n2,lambda2*9.5)
504 x12<-c(XX,X)
505 x22<-rbinom(n2,N,p2)
506 g1<-rep(1,n1)
507 g2<-rep(2,n2)
508 d1<-cbind(x1,x21,g1)
509 d2<-cbind(x12,x22,g2)
510 C1<-cbind(x1,x21)
511 C2<-cbind(x12,x22)
512 c<-C1[,1]
513 cc<-C1[,2]

```

```

514 C<-C2[,1]
515 CC<-C2[,2]
516 Data<-rbind(d1,d2)
517 g<-Data[,3]
518 x1<-Data[,1]
519 x2<-Data[,2]
520 Dt<-data.frame(g,x1,x2)
521 fit<-lda(formula=g~.,data=Dt)
522 tabel<-table(actual=Dt$g,predicted=predict(fit,Dt)$class)
523 pi1<-0.5
524 pi2<-0.5
525 lambda1<-mean(x11)
526 lambda2<-mean(x12)
527 p1<-mean(x11)/mean(x21)
528 p2<-mean(x12)/mean(x22)
529 D<-log(pi2/pi1)+(lambda1-lambda2)
530 a <- 0
531 for(i in 1:length(C1[,1])){
532 A<-c[i]*log(lambda1/lambda2)
533 B<-cc[i]*log(p1/p2)
534 w1 <- A+B
535 if(w1>=D) {
536 a <- a+1
537 }
538 }
539 b <- 0
540 for(i in 1:length(C1[,1])){
541 A<-c[i]*log(lambda1/lambda2)
542 B<-cc[i]*log(p1/p2)
543 w1 <- A+B
544 if(w1<D) {
545 b <- b+1
546 }
547 }
548 c <- 0
549 for(i in 1:length(C2[,1])){
550 A<-C[i]*log(lambda1/lambda2)
551 B<-CC[i]*log(p1/p2)
552 w2 <- A+B
553 if(w2>=D) {
554 c <- c+1
555 }
556 }
557 d <- 0
558 for(i in 1:length(C2[,1])){
559 A<-C[i]*log(lambda1/lambda2)
560 B<-CC[i]*log(p1/p2)
561 w2 <- A+B

```

```

562  if(w2<D) {
563  d <- d+1
564  }
565  }
566  tr <- ftable(tabel)
567  ErFLD <- (tr[1,2]+tr[2,1])/(n1+n2)
568  ErPB <- (b + c)/(n1+n2)
569  list(c(ErPB,ErFLD))
570  #END
571  }
572
573  ## Underdispersed P-B and FLD
574  Underdispersed<-function(n1,n2){
575  set.seed(200)
576  N<-10
577  lambda11<-1.2
578  lambda22<-0.8
579  p11<-0.42
580  p22<-0.31
581  xx<-rpois(0.8*n1,lambda11)
582  x<-rpois(0.2*n1,lambda11/8)
583  x11<-c(xx,x)
584  x21<-rbinom(n1,N,p11)
585  XX<-rpois(0.8*n2,lambda22)
586  X<-rpois(0.2*n2,lambda22/8)
587  x12<-c(XX,X)
588  x22<-rbinom(n2,N,p22)
589  g1<-rep(1,n1)
590  g2<-rep(2,n2)
591  d1<-cbind(x11,x21,g1)
592  d2<-cbind(x12,x22,g2)
593  C1<-cbind(x11,x21)
594  C2<-cbind(x12,x22)
595  c<-C1[,1]
596  cc<-C1[,2]
597  C<-C2[,1]
598  CC<-C2[,2]
599  Data<-rbind(d1,d2)
600  g<-Data[,3]
601  x1<-Data[,1]
602  x2<-Data[,2]
603  Dt<-data.frame(g,x1,x2)
604  fit<-lda(formula=g~.,data=Dt)
605  tabel<-table(actual=Dt$g,predicted=predict(fit,Dt)$class)
606  pi1<-0.5
607  pi2<-0.5

```

```

608 lambda1<-mean(x11)
609 lambda2<-mean(x12)
610 p1<-mean(x11)/mean(x21)
611 p2<-mean(x12)/mean(x22)
612 D<-log(pi2/pi1)+(lambda1-lambda2)
613 a <- 0
614 for(i in 1:length(C1[,1])){
615 A<-c[i]*log(lambda1/lambda2)
616 B<-cc[i]*log(p1/p2)
617 w1 <- A+B
618 if(w1>=D) {
619 a <- a+1
620 }
621 }
622 b <- 0
623 for(i in 1:length(C1[,1])){
624 A<-c[i]*log(lambda1/lambda2)
625 B<-cc[i]*log(p1/p2)
626 w1 <- A+B
627 if(w1<D) {
628 b <- b+1
629 }
630 }
631 c <- 0
632 for(i in 1:length(C2[,1])){
633 A<-C[i]*log(lambda1/lambda2)
634 B<-CC[i]*log(p1/p2)
635 w2 <- A+B
636 if(w2>=D) {
637 c <- c+1
638 }
639 }
640 d <- 0
641 for(i in 1:length(C2[,1])){
642 A<-C[i]*log(lambda1/lambda2)
643 B<-CC[i]*log(p1/p2)
644 w2 <- A+B
645 if(w2<D) {
646 d <- d+1
647 }
648 }
649 tr <- ftable(tabel)
650 ErFLD <- (tr[1,2]+tr[2,1])/(n1+n2)
651 ErPB <- (b + c)/(n1+n2)
652 list(c(ErPB,ErFLD))
653 #END
654 }
655

```

```

656  ## Undispersed P-B and FLD
657  Undispersed<-function(n1,n2){
658  set.seed(200)
659  N<-10
660  lambda1<-1.2
661  lambda2<-0.8
662  p1<-0.42
663  p2<-0.31
664  x11<-rpois(n1,lambda1)
665  x21<-rbinom(n1,N,p1)
666  x12<-rpois(n2,lambda2)
667  x22<-rbinom(n2,N,p2)
668  g1<-rep(1,n1)
669  g2<-rep(2,n2)
670  d1<-cbind(x11,x21,g1)
671  d2<-cbind(x12,x22,g2)
672  C1<-cbind(x11,x21)
673  C2<-cbind(x12,x22)
674  c<-C1[,1]
675  cc<-C1[,2]
676  C<-C2[,1]
677  CC<-C2[,2]
678  Data<-rbind(d1,d2)
679  g<-Data[,3]
680  x1<-Data[,1]
681  x2<-Data[,2]
682  Dt<-data.frame(g,x1,x2)
683  fit<-lda(formula=g~.,data=Dt)
684  tabel<-table(actual=Dt$g,predicted=predict(fit,Dt)$class)
685  pi1<-0.5
686  pi2<-0.5
687  lambda1<-mean(x1)
688  lambda2<-mean(x2)
689  p1<-mean(x1)/mean(x2)
690  p2<-mean(x2)/mean(x2)
691  D<-log(pi2/pi1)+(lambda1-lambda2)
692  a <- 0
693  for(i in 1:length(C1[,1])){
694    A<-cc[i]*log(lambda1/lambda2)
695    B<-c[i]*log(p1/p2)
696    w1 <- A+B
697    if(w1>=D) {
698      a <- a+1
699    }
700  }
701  b <- 0
702  for(i in 1:length(C1[,1])){
703    A<-cc[i]*log(lambda1/lambda2)

```

```

704 B<-c[i]*log(p1/p2)
705 w1 <- A+B
706 if(w1<D) {
707 b <- b+1
708 }
709 }
710 c <- 0
711 for(i in 1:length(C2[,1])){
712 A<-CC[i]*log(lambda1/lambda2)
713 B<-C[i]*log(p1/p2)
714 w2 <- A+B
715 if(w2>=D) {
716 c <- c+1
717 }
718 }
719 d <- 0
720 for(i in 1:length(C2[,1])){
721 A<-CC[i]*log(lambda1/lambda2)
722 B<-C[i]*log(p1/p2)
723 w2 <- A+B
724 if(w2<D) {
725 d <- d+1
726 }
727 }
728 tr <- ftable(tabel)
729 ErFLD <- (tr[1,2]+tr[2,1])/(n1+n2)
730 ErPB <- (b + c)/(n1+n2)
731 list(c(ErPB,ErFLD))
732 #END
733 }

```

**Table 1.** Error rates of the P-P and FLD under various form of dispersion and sample sizes

<b>Sample size (n)</b>	<b>Over-dispersed</b>		<b>Under-dispersed</b>		<b>Undispersed</b>	
	P-P	FLD	P-P	FLD	P-P	FLD
20	0.3750	0.4000	0.2750	0.4000	0.2750	0.4250
50	0.5000	0.4800	0.2800	0.4900	0.3600	0.4200
100	0.5150	0.4900	0.2450	0.4550	0.3600	0.4200
150	0.4567	0.4867	0.3133	0.4533	0.3800	0.4267
200	0.4225	0.4400	0.2950	0.4350	0.3500	0.4150
300	0.4367	0.4433	0.2900	0.4217	0.3600	0.4200
500	0.4390	0.4790	0.3020	0.4350	0.3610	0.4200
750	0.4327	0.4500	0.3307	0.4333	0.3760	0.4387
1000	0.4380	0.4555	0.3320	0.4400	0.3895	0.4305
1500	0.4347	0.4590	0.3203	0.4350	0.3813	0.4253
2000	0.4303	0.4435	0.3215	0.4235	0.3840	0.4100



**Table 2.** Error rates of P-B and FLD under various form of dispersion and sample sizes

<b>Sample size (n)</b>	<b>Over-dispersed</b>		<b>Under-dispersed</b>		<b>Undispersed</b>	
	P-B	FLD	P-B	FLD	P-B	FLD
20	0.5000	0.3000	0.5000	0.3250	0.4750	0.325
50	0.5000	0.3333	0.5100	0.3600	0.4100	0.3700
100	0.4950	0.3800	0.4050	0.3300	0.5050	0.3300
150	0.4667	0.3567	0.4500	0.3433	0.4967	0.3533
200	0.4450	0.3100	0.4300	0.3275	0.4950	0.3200
300	0.4833	0.3167	0.4200	0.3333	0.4883	0.3217
500	0.4300	0.3530	0.4340	0.3500	0.4950	0.3370
750	0.4540	0.3440	0.4773	0.3460	0.4693	0.3447
1000	0.4615	0.3320	0.4210	0.3465	0.4900	0.3355
1500	0.4373	0.3393	0.3937	0.3440	0.4893	0.3380
2000	0.4228	0.3545	0.4290	0.3493	0.4935	0.3448