International Journal of Statistika and Mathematika, ISSN: 2277- 2790 E-ISSN: 2249-8605, Volume 4, Issue 2, 2012 pp 47-49

D – Optimal Design for Compound Poisson Regression Model

S. Joshua David[1*] and C. Santharam[2†]

[1,2] Department of Statistics, Loyola College, Chennai – 600034 (TN) INDIA.

Corresponding Addresses:

[*]6joshuaadavid@gmail.com, [†]santharamsdeep@yahoo.com

Research Article

Abstract: D- Optimality is one of the most commonly used design criteria for linear regression model. In industrial experiments binary or count data often arise, for example defective/non- defective or number of defects. For such data (GLMS) are appropriate Generalized Linear Models are especially useful for actuarial applications, since they allow estimate multiplicative models, and also allow forms of heteroscedasticity such as they are found frequently in actuarial problems, of Poisson-type, of gamma-type with a fixed coefficient of variation. An analogous D-optimality design criterion can be developed using asymptotic covariance matrix, for GLM, this matrix is a weighted version of the covariance matrix for the linear case, and the extension of existing D-optimality algorithm. We consider the problem of finding an optimal design under a compound Poisson regression model with a, any number of independent variables and a reciprocal link additive model linear predictor. Local D-optimality of a class of designs is established through use of a canonical form of the problem and a general equivalence theorem. The theorem is applied in conjunction with clustering techniques to obtain a fast method of finding designs that are robust to wide ranges of model parameter values.

Key words: Compound Poisson, Clustering, Locally optimal design, Inverse link models.

Introduction:

D-optimality is a standard optimal design for linear regression model, partly because it corresponds to minimizing the area of confidence region for unknown parameters, and partly because it is reasonably easy mathematically and computationally. In actuarial statistics non-normal response are often measured. For such a data Compound Poisson regression models is appropriate (see for example McCullagh and Nelder 1989 and Dobson 1983). Compound Poisson optimal design criterion have been developed for the exponential dispersion model are directly applicable to the compound Poisson distribution applications of this particular compound distribution, primarily in the form of generalized linear models [GLM], have been found in actuarial science (Smyth and J_rgensen 2002), animal studies (Perry 1981), assay analysis (Davidian 1990), botany studies (Dunn and Smyth 2005), survival analysis (Hougaard, Harvald, and Holm 1992), rainfall modeling (Dunn 2004) and _shery research (Shono 2008). An often neglected part of the analysis is the estimation of the unknown variance function, i.e., the index parameter p. This parameter has a signi_cant impact on hypothesis tests and predictive Uncertainty measures (Davidian and Carroll 1987; Peters, Shevchenko, and W uthrich 2009; Zhang 2012), which is of independent interest in many applications. One approach in estimating the Variance function is using the prole likelihood (Cox and Reid 1987). For the compound Poisson distribution, such an approach must be implemented based on the true likelihood rather than the extended quasi-likelihood (Nelder and Pregibon 1987). Unlike other compound distributions whose density must be approximated via the slow recursive approach (Klugman, Panjer, andWillmot 2008), methods that enable fast and accurate numerical evaluation of compound Poisson density function are available (Dunn and Smyth 2005, 2008). These methods are provided by the tweedie package (Dunn 2011). With the density approximation methods, we can carry out not only maximum likelihood estimation but also Bayesian inference using Markov chain Monte Carlo methods (Gelman, arlin,Stern, and Rubin 2003).

The D-Optimality Criterion for Compound Poisson Model

The motivation for a D-optimality criterion for GLMs can be seen as a simple extension of the linear case. One way of viewing the linear case is that a D-optimal design will yield the smallest confidence regions for the parameters of interest. In the GLM setting, confidence statements are often made on an asymptotic basic, via the information matrix. The inverse expected information matrix I^{-1} is the analogue of the covariance matrix of the coefficients in the linear model. Thus to minimize the size of confidence regions for GLM parameter estimates, I^{-1} is used in place of the covariance matrix $(X'X)^{-1}$. For members of the GLM family compound Poisson distribution follows Gamma distribution, the information matrix has an inverse form, which is the main reason that specific extension is considered. Before illustrating this form, we define some notation for Compound Poisson model. If Y is the response vector, and corresponding values of the predictor variables are in the matrix X, then the expected value of response Y^i is related to the predictors xi via the given " inverse link" function g:

$E[Y_i / x_i] = \mu_i = g(x_i'\beta) = g(\eta)$

Where $\mu = \frac{1}{\eta(x)}$, β is a vector of parameters to be estimated. The variance of the response is given by:

$\mathrm{Var}[Y_i / x_i] = V(\mu_i^2)$.

The expected information does not depend on the form of the distribution beyond the mean and variance relations stated above. Dobson (1983, Appendix 2) shows that for GLMs with the above specification, the jk^{th} element of the expected information matrix is given as follows:

$$I_{jk} = \sum_i^N \frac{x_{ij}\, x_{ik}}{V(\mu_i^2)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2$$

This can be written in matrix notation: I = $= X'WX$.

Where W is a diagonal matrix with elements:

$$wi = \left(\frac{\partial \mu i}{\partial \eta i}\right)^2 \frac{1}{V(\mu i)}$$

Thus the asymptotic variance covariance matrix $I^{-1} = (X'X)^{-1}$ is of a similar form to the covariance matrix $(X'X)^{-1}$.

The simplest form of Compound Poisson model use the inverse link function given as follows.

$$\mu = \frac{1}{\eta(x)} \text{ And Var } (\mu) = (\mu^p)$$

The linear predictor is define by

$$\eta(x) = \frac{1}{\beta_0 + \beta_j x_{ij}} \quad \mu_i = (x_i'\beta)^{-1}$$

The information matrix under the inverse gamma model, given an observation at x, is

$$I(\theta, x) = v(x) \left(\frac{\partial \eta(x)}{\partial(\theta)}\right) \left(\frac{\partial \eta(x)}{\partial(\theta)}\right)'$$

Where $\left(\frac{\partial \eta(x)}{\partial(\theta)}\right) = \begin{pmatrix} -1 \\ -x_{ij} \\ -\beta j \end{pmatrix}$

And $v(x) = \frac{1}{V(Y)} \left(\frac{\partial \mu(x)}{\partial \eta(x)}\right)^2 = \mu^p$

The important application for this model is to find the point of maximum or minimum response and to estimate the parameters. The optimal design is the choice of points and corresponding weights that optimize the selected criterion function; the criterion function is commonly related to precision of the parameter estimates, such as the size of a confidence region of the sum of the variance of the parameter estimators. The criterion function involves the standardized information matrix. The standardized information matrix given a particular design ξ is the weighted sum of the contributions from each of the n design points.

The D-Optimality is the most common criterion `which seeks to maximize $|x'x|$, the Determinant of the information matrix (*X'X*) of the design' (NIST/SEMATECH 2007,ch. 5.5.2). This means that the optimal design matrix *X*\* contains the *n* experiments which maximizes the determinant of (*X'X*) Or in other words, the *n* runs `span the largest volume possible in the experimental region' (Eriksson et al. 2000, p. 216).Selection of *X**out of all possible design matrices chosen from ϑN. This connection between the design matrix and the determinant also explains the use of the "D" in the term D-optimal designs. $|X^{*\prime} X^*| = \max(|X'X|)$

Maximizing the determinant of the information matrix (*X'X*) is equivalent to Minimizing the determinant of the dispersion matrix $(X'X)^{-1}$.

D-optimality criterion is just to maximize the determinant of the Fisher information matrix.

Using such an idea, the D-efficiency of an arbitrary design, X, is naturally defined as

$$Eff\ D = \frac{|I(x,\beta)|}{|I(x^*,\beta)|}$$

X* is the true optimal design.

$$M(\theta,\xi) = \sum_{i=1}^n w_i v(x_i) \left(\frac{\partial \eta(x_i)}{\partial \theta}\right) \left(\frac{\partial \eta(x_i)}{\partial \theta}\right)'$$

$$= \sum_{i=1}^n w_i\, \mu^2 \begin{pmatrix} -1 & x_{ij} & \beta_j \\ -x_{ij} & x_{ij}^2 & \beta_j x_{ij} \\ -\beta_j & \beta_j x_{ij} & \beta_j^2 \end{pmatrix}$$

There are several methods to the practise of determining the optimal design. These include algorithms, analytical, numerical and graphical methods.

The derivation of D-optimal designs will be illustrated for three parameter sets:

θA=(-3 , 1 0); θB=(0,1,0) and θC=(3,1,0)

Step 1: To find the symmetric design consisting of p=3 points with equal design weights is assumed.

$$\xi_3 = \begin{Bmatrix} x & 0 & -x \\ 1/3 & 1/3 & 1/3 \end{Bmatrix}$$

$$\xi_{3A} = \begin{Bmatrix} 0.833333 & 0 & -0.833333 \\ 1/3 & 1/3 & 1/3 \end{Bmatrix}$$

$$\xi_{3B} = \begin{Bmatrix} 0.9047621.58 & 0 & -0.904762 \\ 1/3 & 1/3 & 1/3 \end{Bmatrix}$$

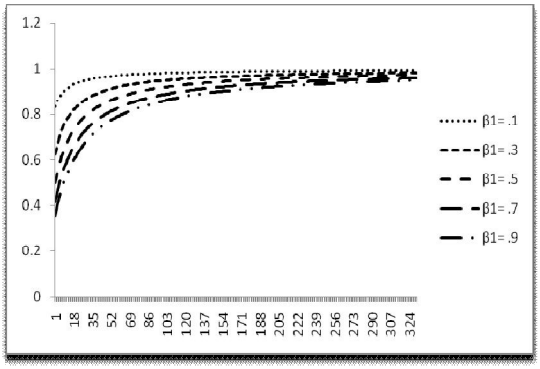
$$\xi_{3C} = \begin{Bmatrix} 1.95 & 0 & -1.95 \\ 1/3 & 1/3 & 1/3 \end{Bmatrix}$$

The optimal design depends on the values of β, so if they are known we can find the optimal design.

The linear predictor is $\eta(x)=\frac{1}{\beta_0+\beta_j x_{ij}}$ we plot the optimal design for varying β1. The curve represents linear predictors for the various mean values as a function of x for fixed β0. The slope β1 is varied to produce different linear predictors; we find that optimal design is increasing when the mean values are decreasing. The optimal design has equal allocation between [-1, 1].

Table: Support points of the locally D-optimal design for the Compound Poisson model for various values of μ_i

Mi	β1=.1	β1=.3	β1=.5	β1=.7	β1=.9
2	0.833333	0.625	0.5	0.416667	0.357143
1.818182	0.846154	0.647059	0.52381	0.44	0.37931
1.666667	0.857143	0.666667	0.545455	0.461538	0.4
1.538462	0.866667	0.684211	0.565217	0.481481	0.419355
1.428571	0.875	0.7	0.583333	0.5	0.4375
1.333333	0.882353	0.714286	0.6	0.517241	0.454545
1.25	0.888889	0.727273	0.615385	0.533333	0.470588
1.176471	0.894737	0.73913	0.62963	0.548387	0.485714
1.111111	0.9	0.75	0.642857	0.5625	0.5
1.052632	0.904762	0.76	0.655172	0.575758	0.513514
1	0.909091	0.769231	0.666667	0.588235	0.526316
0.952381	0.913043	0.777778	0.677419	0.6	0.538462
0.909091	0.916667	0.785714	0.6875	0.611111	0.55
0.869565	0.92	0.793103	0.69697	0.621622	0.560976
0.833333	0.923077	0.8	0.705882	0.631579	0.571429
0.8	0.925926	0.806452	0.714286	0.641026	0.581395
0.769231	0.928571	0.8125	0.722222	0.65	0.590909
0.740741	0.931034	0.818182	0.72973	0.658537	0.6
0.714286	0.933333	0.823529	0.736842	0.666667	0.608696
0.689655	0.935484	0.828571	0.74359	0.674419	0.617021
0.666667	0.9375	0.833333	0.75	0.681818	0.625



Conclusion:

The analytic contraction of D-optimal Compound Poisson model with inverse link function were derived to obtain locally D-optimal design, the solution to these formulae can be obtain numerically using R- software in contrast to using for examples algorithms which involve issues like finding a good initial design. Thus the locally D-optimal design depends only on the parameters β with varying Xi values it is the function of mean values. We conclude that when the optimal design with inverse link function have a inverse relationship between function of mean values and the predicted vales. When the function of mean values is decreasing the predictors are increasing the design involves asymptotic information matrix. The linear predictor attains maximal optimal values for the different β values it converges to one for various functional mean values.

References:

1. Silvey, D., 1980. Optimal Design, Chapman and Hall, London.
2. Bai, D.S., Chung, S.W.: An optimal design of accelerated life test for exponential distribution.Reliab. Eng. Syst. Saf. 31, 57–64 (1991).
3. Hugh A. Chipman and William J. Welch. D-optimal Design for Generalized Linear Models, Chicago (1996).
4. Holger Dette, Viatcheslav B. Melas and Weng Kee Wong: Locally D-Optimal Design for Exponential Regression Models. Los Angeles. Statistica Sinica 16(2006).
5. Santharam, C., Ponnusamy, K.N and Chandrasekar, B.,Universal optimality of nearest neighbour balanced block designs using ARMA models. Biometrical.J. 32,725 – 730, 1996.
6. RUBAN RAJA. B, SANTHARAM. C and RAMESHKUMAR, (2012) MV-Optimality of Nearest Neighbour Balanced Block Designs using First order and Second order correlated models, International Journal of Statistika and M athematika, (ISSN: 2277-2790 E-ISSN: 2249-8605)