行政院國家科學委員會專題研究計畫成果報告

多變量逆高斯分佈的進階研究(2/2)

計畫編號：NSC 90-2118-M-009-014
執行期限：2001 年 8 月 1 日至 2002 年 7 月 31 日
主持人：李昭勝博士　　國立交通大學統計研究所

一、中文摘要

在這個計畫中，我們考慮多維逆高斯分配的統計推論方法。其主要目的是要解決競爭模型中的問題。在這個計畫中，由於資料都是不完整的競爭資料，各種常見的方法都不太適用。因此，我們採用了複雜的 EM 演算法。除了估計外，我們還得到大樣本時的估計分配與信賴區間。另外，我們也考慮了可信度函數的估計。在小樣本時，我們比較了我們由大樣本得到的方法與有母數重與無母數重複抽樣的方法。最後，我們考慮了一些實際的例子與一些模型模擬資料的分析。

關鍵詞：多維逆高斯分配，競爭模型，EM 演算法，信賴區間，可信度函數，無母數重複抽樣

Abstract

We consider inference for the poly-inverse Gaussian model, which arises in competing risk scenarios when the risks have independent inverse Gaussian distributions. This article deals with maximum likelihood estimation of the parameters of the poly-inverse Gaussian distribution. Due to the complexity of the likelihood, direct maximization is difficult. An EM type algorithm is provided for the maximum likelihood estimation of the poly-inverse Gaussian distribution. They are then applied to obtain the reliability function. Next, we study the confidence intervals of the parameters and the reliability function basing on the large sample theory and the bootstrap method. Furthermore, we also explore the influences after changing some competing risks of the unit. The maximum likelihood estimation of the parameters and that of reliability function are investigated, and we apply the results on the appliance data of Nelson,W. (Applied life data analysis (1982), chapter 5)

Keywords:. poly-inverse Gaussian model, competing risk model, EM type algorithm, confidence interval, reliability function, nonparametric bootstrap method.

二、緣由與目的

Competing risk lifetime data arise in medical, engineering, and many other contexts when death or failure of an individual or unit is classified into one of a variety of types or causes. In the classical competing risks framework, a subject may only fail from one of several distinct causes. The analysis of competing risks data goes to heart of modern preoccupations in survival and lifetime analyses, touching on many of the major areas of importance in the subject. The exertion expended in fitting and interpreting competing risks model of some sort, apparent from a glance at the current literature, recommends them as a worthy object of study.

While there have been many investigations, often from the point of view of non-parameter and semi-parameter estimate, basing on large sample theory. However, we may get inappropriate conclusion with small sample size in above models. And some other articles assume the individual or unit with mixture models, and then do inference. Furthermore, there have been some papers, they assume each risk has individual distribution like log-normal, Weibull, exponential, Gamma, etc, to do inference. They consider inference for the poly-Weibull model, which arises in competing risk scenarios when the risks have independent Weibull distributions. However, Takagi recommended the mean of random samples from a log-normal distribution is not lognormally distribution, it cannot, deal with the changes in the distribution of time-weighted average values when the averaging time varies. Furthermore, Mudholkar and Natarajan (2002) described the inverse Gaussian family is strikingly analogous to Gaussian 2 family in terms of having simple inference solutions, which use the family chi, Gamma, and F distributions, for a variety of basic problem. Consequently, the inverse Gaussian distribution, which is similar to the log-normal distribution, offers the advantage of reproducibility. Furthermore, the advantage of developing an inverse Gaussian distribution based model is that it can be used for diverse quality variables ranging from highly skewed to almost symmetrical. Estimation of competing risks using the inverse Gaussian model plays a prominent role in the analysis of competing risks and addresses the data-analytic questions of interest.

The interpretation of the inverse Gaussian random variable as the first passage time suggests its potentially useful applications in studying lifetime or number of event occurrences for a wide range of fields. When early occurrences such as product failures or repairs are dominant in a lifetime distribution, its failure rate is expected to be non-monotonic, first increasing and later decreasing. In that situation, the inverse Gaussian distribution provides a suitable choice for a lifetime model.

We consider inference for the poly-inverse Gaussian model, which arises in competing risk scenarios when the risks have independent inverse Gaussian distributions. This article deals with maximum likelihood estimation of the parameters of the poly-inverse Gaussian distribution. Due to the complexity of the likelihood, direct maximization is difficult. An EM type algorithm is provided for the maximum likelihood estimation of the poly-inverse Gaussian distribution. They are then applied to obtain the reliability function. Next, we study the confidence intervals of the parameters and the reliability function basing on the large sample theory and the bootstrap method. Furthermore, we also explore the influences after changing some competing risks of the unit. The maximum likelihood estimation of the parameters and that of reliability function are investigated, and we apply the results on the appliance data of Nelson, (1982).

三、結果與討論

From the inverse Gaussian model, we get the likelihood function of the series system. Unfortunately, we observe that the likelihood function is difficult to maximize directly. Consequently, we try the EM algorithm to find the MLE. Since estimating the variance of the MLE is an important issue when we use EM algorithm, various authors have devised different methods. Louis (1982) derives a useful identity relating the log-likelihood function of the observed data to the log-likelihood function of the complete data. Using this identity we obtain the following covariance matrix of the MLE. Therefore, when the sample size n and the number of replications are large, the approximation is good enough.

When the sample size is small, in our paper, we found that the bootstrap technique (Efron 1979a) is a good method for understanding properties about our estimator. Also, we consider the estimation of reliability function.

The MLE of the reliability is given which can be obtained by substituting the estimates of the parameters. In order to obtain the asymptotic variance of the MLE of the reliability, we applied the Taylor expansion approximation. Using the above result, we obtain a good approximation of confidence intervals for reliability function.

All analyses we mention above require that the cause of each failure be identified. However, many areas of application, such as engineering and neurophysiology studies, the cause of each failure may be independent similar distribution but not identical. Therefore, we want to study how to estimate the parameters of causes distributed as inverse Gaussian with different parameters. Consequently, we can analyze competing risk data sets which allows for censoring and immune individuals.

For the generation of inverse Gaussian distribution, Michael et al. (1976) gave a method of generating random variables using a transformation with multiple roots. Their basic approach is to find a transformation of the random variable of interest, and then use the multinomial probabilities associated with the multiple roots of the transformation to choose one root for the random observation. In the simulation study, we only observe the minimum values of the lifetime of the causes. The results are summarized in our tables.

From the simulation results, when the sample sizes are relatively small, we found that the bootstrap methods are better than the traditional confidence interval based on large sample theory. From our tables, when the mode 1 is improved, we found that the estimates for mode 1 are suitable. Further, the estimates for other models after improving mode 1 are almost equal to the estimates of original experiment. From our figures, we find the system is improved obviously after improving mode 1.

The data (chapter 5 of "Applied life data analysis", Nelson (1982)) comes from the life testing of 52 units at various stages in a development program. For each unit, the data consist of the number of cycles it ran to failure or to removal from test and its failure code (one of 18 causes). We assume all failure modes are log-inverse Gaussian distribution. And from group 5, we have shown some good results in our paper which include the estimations of parameters, confidence intervals and the reliability functions. We believe that all these results will be useful for the competing risk models.

四、計畫成果自評

相信本計畫的研究結果對於競爭模型的研究有一定的貢獻。尤其是我們考慮資料經過某種特殊轉換後，然後再賦予一個好的逆高斯分配的模型。這種看法有其特殊的貢獻。模型也有一定的廣度。相信一定可以廣泛的被統計學家所使用。

五、參考文獻

Azcarate C. and Mallor F. (1999). A nonparametric estimator from autopsy data. *Communications in Statistics-Theory and Methods.* 28(2):479-494.

Barnodorff-Nielsen, O.E. (1983). On a formula for the distribution of themaximum likelihood estimator. *Biometrika* 70, 343-365.

Chhikara, R.S. andd Folks, J.L (1974). Estimation of the inverse Gaussian distribution function. *J.Amer. Statist. Ass*.,69:250-254

Chhikara, R.S. andd Folks, J.L (1977). The inverse Gaussian distribution as a lifetime model. *Technometrics* 19, 461-468.

Cooke R.M. (1996). The design of reliability data bases.1. Review of standard design concepts. *Reliability Engineering & System Safety.* 51(2):137-146.

Efron, B., & R.J. Tibshirani (1993). *An Introduction to the Bootstrap.* New York: Chapman & Hall.

Folk, J. L. and Chhikara, R.S. (1978). The inverse Gaussian distribution and its statistical application-a review. *J. Roy. Stat. Soc.*, B 40:263-289.

Gupta R.C. and Akman H.O.(1995). On the reliability studies of a weighted inverse Gaussian model. *Journal of Statistical Planning and Inference*,48 (1): 69-83.

Kalbfleisch J.D. and Prentice R.L. (1980). *The Statistical Analysis of Failure Time Data.* Wiley, New York.

Luo X.L. and Turnbull B.W. Comparing two treatments with multiple competing risks endpoints. *Statistica Sinica.* 9(4):985-997.

Jiang R., Murthy D.N.P. and Ji P. (2001). Models involving two inverse Weibull distributions. *Reliability Engineering & System Safety.* 73(1):73-81.

Maller R.A. and Zhou X.(2002). Analysis of parametric models for competing risks. *Statistica Sinica*, 12 (3): 725-750.

Mudholkar G.S. and Natarajan R. (2002). The inverse Gaussian models: Analogues of symmetry, skewness and kurtosis. *Annals of the Institute of Statistical Mathematics.*54(1): 138-154.

Nelson,W. (1982). *Applied Life Data Analysis.* Wiley, New York.

Sato S. and Inoue J. (1994). Inverse Gaussian distribution and its application. *Electronics and Communications in Japan*, Part3, Vol. 77(1), No. 1:32-42

Satten G.A. and Datta S. (1999). Kaplan-Meier representation of competing risk estimates. *Biomedical and Environmental Sciences.* 11(4):331-335.

Shuster, J. J. (1968). On the inverse Gaussian distribution function. J. *Amer. Statist. Ass*.,63:1514-1516.

Smith C.E. and Lansky P. (1994). A reliability application of a mixture of inverse Gaussian distributions. *Applied Stochastic Models and Data Analysis.* 10(1):61-69.

Tweedie, M.C.K. (1956). Some statistical properties of inverse Gaussian distribution. *Virginia Journal of Science* 7, 160-165.

Tweedie, M.C.K. (1957). Statistical properties of inverse Gaussian distribution, I. *Annals of Math. Stat*. 28, 362-377.

附件：封面格式

行政院國家科學委員會補助專題研究計畫成果報告

※※※※※※※※※※※※※※※※※※※※※※※※※※

※　　　　　　　　　　　　　　　　　　　　　　　　※

※　　多變量逆高斯分佈的進階研究(2/2)　　※

※　　　　　　　　　　　　　　　　　　　　　　　　※

※※※※※※※※※※※※※※※※※※※※※※※※※※

計畫類別：□個別型計畫　　□整合型計畫

計畫編號：NSC　90－2118－M－009－014

執行期間：2001 年 8 月 1 日至　2002 年 7 月 31 日

計畫主持人：李昭勝

共同主持人：

計畫參與人員：

本成果報告包括以下應繳交之附件：

□赴國外出差或研習心得報告一份

□赴大陸地區出差或研習心得報告一份

□出席國際學術會議心得報告及發表之論文各一份

□國際合作研究計畫國外研究報告書一份

執行單位：國立交通大學統計所

中　華　民　國　91　年　10　月　28　日