Journal of Modern Applied Statistical Methods

Volume 6 | Issue 1 Article 29

5-1-2007

Estimation of Risk for Developing Cardiac Problem in Patients of Type 2 Diabetes as Obtained by the Technique of Density Estimation

Ajit Mukherjee
Division of Reproductive Health and Nutrition, Indian Counsel of Medical Research, New Delhi, India

Ajit Mathur
Division of Reproductive Health and Nutrition, Indian Counsel of Medical Research, New Delhi, India

Rakesh Mittal
Division of Reproductive Health and Nutrition, Indian Counsel of Medical Research, New Delhi, India

Recommended Citation

Mukherjee, Ajit; Mathur, Ajit; and Mittal, Rakesh (2007) "Estimation of Risk for Developing Cardiac Problem in Patients of Type 2 Diabetes as Obtained by the Technique of Density Estimation," *Journal of Modern Applied Statistical Methods*: Vol. 6 : Iss. 1 , Article 29.
DOI: 10.22237/jmasm/1177993680
Available at: http://digitalcommons.wayne.edu/jmasm/vol6/iss1/29

Journal of Modern Applied Statistical Methods
May, 2007, Vol. 6, No. 1, 320-323

1538 – 9472/07/$95.00

Estimation of Risk for Developing Cardiac Problem in Patients of Type 2 Diabetes as Obtained by the Technique of Density Estimation

Ajit Mukherjee Ajit Mathur Rakesh Mittal
Division of Reproductive Health and Nutrition, Indian Council of Medical Research
New Delhi, India

High levels of cholesterol and triglyceride are known to be strongly associated with development of cardiac problem in patients of type 2 diabetes. In a hospital-based study, patients showing ECG positive were compared with those who were not. The observations on cholesterol and triglyceride were considered for estimation of risk for developing the cardiac problem. The technique of density estimation employing Epanechnikov kernel was used for estimating bivariate probability density functions with respect to observations on cholesterol and triglyceride of the two groups. Using the odds form of Bayes' rule, the estimates of posterior odds were computed.

Key words: Density estimation, kernel, logistic regression, probability density function.

Introduction

The technique of Density Estimation is a non-parametric approach and involves no assumptions as it deals directly with the experimental data. The method of density estimation describes the probability distribution of people with respect to the parameter under investigation. This technique has found favour with many applied statisticians in the past. Scott, Gotto, Cole, and Gorry (1978) used density estimation for assessing plasma lipids as collateral risk factors in coronary artery disease. Bithell (1990) gave an application of this technique in Geographical Epidemiology. Mukherjee, Kumar, Mittal, and Saxena (2002) used density estimation for estimating risk of developing goiter in an endemic area. Silverman (1986) provided an excellent account of various approaches to Density Estimation in his book.

High levels of cholesterol and triglyceride are known to be strongly associated with development of cardiac problem in patients of type 2 diabetes. However, the extent of risk posed by elevated levels of these two risk factors in patients of type 2 diabetes has not been studied extensively. The present article describes an alternative methodology whereby risk of developing cardiac problem in patients of type 2 diabetes can be estimated using cholesterol and triglyceride as risk factors.

Ajit Mukherjee, Ph.D. in Mathematical Statistics, is Deputy Director General in Division of Reproductive Health and Nutrition, Indian Council of Medical Research, Ansari Nagar, Post Box 4911, New Delhi-110029, India. He has over 30 publications to his credit in both national and international journals and is the holder of a prestigious National Award E-mail: mukherjeeajit@hotmail.com. Ajit Mathur, M.Sc., is also Deputy Director General in Division of Reproductive Health and Nutrition. He has 25 publications to his credit. E-mail:ajiticmr@hotmail.com. Rakesh Mittal, M. D., is Deputy Director General (Senior Grade), also in Division of Reproductive Health and Nutrition. He has about 30 publications and books to his credit.E-mail:mittrak@hotmail.com

Methodology

In a hospital-based study conducted by Indian Council of Medical Research in 1989-92, 4637 patients of Non Insulin Dependent Diabetes Mellitus (NIDDM) also known as Type2 Diabetes were enrolled. Various bio-chemical investigations and electrocardiogram (ECG) were carried out at regular intervals. The 311 patients showing ECG positive and thereby indicating coronary artery disease (CAD +), formed the first group. The remaining patients numbering 4326 formed the second group. The Epanechnikov kernel, which is known to have 100% efficiency in terms of mean integrated square error (Silverman, 1986), was employed in the technique of density estimation for estimating probability density functions of the patients falling in the two groups with respect to their cholesterol and triglyceride levels. Using the odds form of Bayes' rule, the estimate of odds ratio (OR) was obtained. A simulation study was undertaken and 100 estimates of OR were generated using the approach of density estimation giving a mean estimated odds ratio and an estimate of standard deviation.

Let the number of patients in the first group be denoted by N_{CAD+} and that in the second group by N_{NID}. Let x and y in general denote the observations on cholesterol and triglyceride of the patients with x_i and y_i being the observations on the ith patient. Then the bivariate kernel density estimator for the first group is given by

$$f_{CAD+} \equiv f_{CAD+}(x, y)$$
$$= \frac{1}{N_{CAD+} h_x h_y} \sum_{i=1}^{N_{CAD+}} K\left(\frac{x - x_i}{h_x}\right) K\left(\frac{y - y_i}{h_y}\right)$$

the quantities h_x and h_y are called bandwidths of the function f_{CAD+} and are appropriately chosen.

The function K(z) which is known as Epanechnikov kernel, is defined as follows:

$$K(z) = \begin{cases} \frac{3}{4}\left(1 - \frac{1}{5}z^2\right)/\sqrt{5}, & if\,|z| < \sqrt{5} \\ 0, & otherwise \end{cases}$$

Similarly, the Epanechnikov kernel density estimator for the other group namely, $f_{NID} \equiv f_{NID}(x, y)$ can also be worked out.

The Likelihood Ratio (LR) will then be given by

$$LR = \frac{f_{CAD+}}{f_{NID}}.$$

Further, the odds form of Bayes' rule states that

$$O'(D) = O(D) * LR,$$

where $O'(D)$ is the posterior odds and $O(D)$ is prior odds and is given as $O(D) = \frac{N_{CAD+}}{N_{NID}}$.

For the present set of data, $N_{CAD+} = 311$ and $N_{NID} = 4326$. For estimating f_{CAD+}, the optimum values of h_x and h_y were obtained by objectively starting the process of smoothing with $h_x = h_y = 5$ giving an increment of 5 until $h_x = h_y = 35$. Thereafter, the process of smoothing was continued by giving a unit increment. The estimates of density stabilized with values of h_x and h_y at 40, 41,42, 43 to five decimal places. Hence, $h_x = h_y = 40$ was accepted as an optimum value of h_x and h_y.

Similarly, the values of h_x and h_y for estimating f_{NID} were also obtained to be 40 each. The estimate of probability density function for the first group i.e., f_{CAD+} is as depicted in Figure 1.

Computation of OR

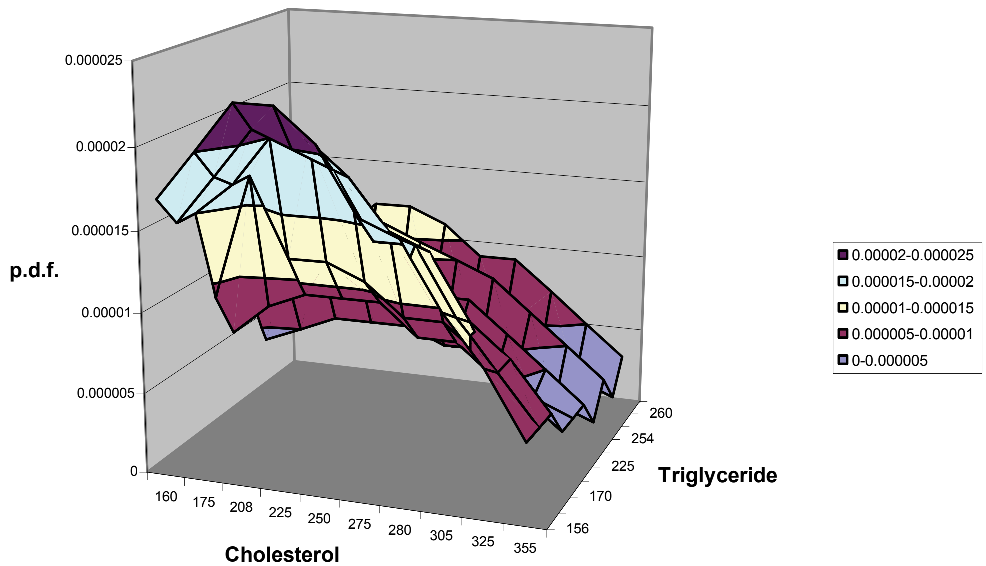
Consider the following transformation

$$P = \frac{1}{1 + e^{-Y}}, \text{ where}$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + - - - + \beta_k X_k + \varepsilon.$$

It can be shown that

Fig.1. Bivariate Probability Density Function of patients with CAD+ at different values of Cholesterol and Triglyceride using Epanechnikov Kernel



$$\log\left(\frac{P}{1-P}\right)=Y$$

or

$$\log\left(\frac{P}{1-P}\right)=\beta_0+\beta_1X_1+\beta_2X_2+---+\beta_kX_k+\varepsilon$$

.

Thus β_i gives average quantum of change in log(odds) per unit change in X_i, $i=1,2,.....,k$ and e^{β_i} gives the odds ratio with respect to X_i keeping other predictors at constant levels.

Results

Keeping cholesterol fixed at 250 and varying the value of triglyceride from 209 to 254, the posterior odds at three pairs of values of cholesterol and triglyceride viz., (250,209), (250,254) and (250, 260) were worked out to be respectively 0.0629, 0.08047 and 0.08549. Thus, keeping cholesterol fixed at 250 and increasing triglyceride by a margin of 45 and 51 units from 209, led to respectively 1.28 and 1.34 times increase in odds for developing a cardiac problem. Further, considering first two of the above three pairs of values of cholesterol and triglyceride, the following would be obtained:

$$\beta_Y=\frac{\log(0.08047)-\log(0.0629)}{45}$$

or

$$\beta_Y=0.005474$$

Therefore OR is given by $e^{\beta Y} = 1.005489$. Drawing a simple random sample of 100 consecutive pairs of values of cholesterol and triglyceride and using the above methodology, 100 estimates of OR were obtained with a mean value of OR as 1.0025 and S.D. of 0.0027 giving 95% C.I. as {1.0020, 1.0031}. The OR as estimated by Logistic Regression model was 1.0029 with a 95% C.I. of {0.9984, 1.0074}.

Conclusion

It is seen that with the technique of density estimation employing Epanechnikov kernel, it is possible to obtain an estimate of the probability density function of the patients of type 2 diabetes falling in the two groups with respect to their cholesterol and triglyceride levels. It has also been demonstrated in the present article, how the posterior odds vary with increasing levels of triglyceride keeping cholesterol at a constant high level, which ultimately led to an estimate of odds ratio (Table 1).

From table 1, it is clear that the estimate of odds ratio as obtained by the method of density estimation is in close proximity to the estimate as obtained by the method of logistic regression. Thus, the risk of developing a cadiac problem can also be alternatively estimated by using the technique of density estimation.

References

Bithell, J. F (1990). An application of density estimation to geographical epidemiology. *Statistics in Medicine*, *9*, 691-701.

Hennekens, C. H. & Burning, J. E. (1987). In (S. L. Mayrent, Ed.) *Epidemiology in medicine*. Boston, MA: Little, Brown and Co.

Mukherjee, A., Kumar, S., Mittal, R. & Saxena, N. C. (2002). An application of the technique of density estimation in geographical epidemiology of goitre in the Muzaffarpur district of Bihar state of India. *Statistics in Medicine*, *21*, 2403-2407.

Scott, D. W., Gotto, A. M., Cole, J.S., & Gorry, G. A. (1978). Plasma lipids as collateral risk factors in coronary artery disease: A study of 371 males with chest pain. *Journal of Chronic Diseases, 31*, 337-345.

Silverman, B. W (1986). *Density estimation for statistics and data analysis*. London: Chapman and Hall.

Table 1. Estimates of Odds Ratio as obtained by the application of Epanechnikov kernel in Density Estimation and Logistic Regression

Method	**Odds Ratio**	95% C.I.
Logistic Model	1.0029	0.9984-1.0074
Density Estimation: Epanechnikov Kernel	1.0025	1.002-1.0031