Ann. Inst. Statist. Math.
31 (1979), Part A, 271-277

A CONTINUOUS FORM OF POST-STRATIFICATION

DAVID R. BRILLINGER

(Received Dec. 4, 1978; revised Mar. 10, 1979)

Summary

The problem of estimating a given integral of a regression function is considered. The proposed estimate may be viewed as a continuous analog of the post-stratified mean of sample survey theory. The asymptotic distribution of the estimate is derived, under regularity conditions, and an estimate of its variance suggested.

1. Introduction

In seismology, travel time tables provide estimates of the time required for an earthquake wave to travel to an observatory as a function of the (angular) distance between the epicenter of the earthquake and the observatory. In practice the tables are constructed from data involving events whose azimuth angles, from the observatory, are not uniformly distributed. However the seismologist does not generally wish any interval of angles to be weighted more heavily than any other. This situation leads to the consideration of the following problem: let Y be a random variate whose distribution depends on a real-valued quantity X and in particular let

$$\mathrm{E}\{Y|X\}=\mu(X)\,, \qquad \mathrm{var}\{Y|X\}=\sigma^2(X)\,.$$

Suppose that it is desired to estimate the value

$$\theta=\int \mu(x)g(x)dx \tag{1.1}$$

for some given function $g(x)$. (In the seismological example mentioned $g(x)=1/(2\pi)$ for $0<x<2\pi$.) Suppose that the observations (X_i, Y_i), $i=1, \cdots, n$ are available. One means of proceeding is to form $\hat{f}(x)$ an estimate of the density of the X's at x and then to form

$$\hat{\theta}=n^{-1}\sum_1^n Y_i g(X_i)/\hat{f}(X_i)\,, \tag{1.2}$$

a weighted mean of the Y's, as an estimate of the desired value θ.

The connection of the estimate (1.2) with the traditional post-stratified estimator of sample survey (discussed for example in Cochran [4], Section 5A.8) may be seen as follows. Consider a random sample of values $Y_1, \cdots, Y_n$ selected from a finite population of size N. Suppose the population is known to contain N_h values in stratum h, $h=1, \cdots, H$. The post-stratified estimate of the mean of the population is given by

$$\sum_h \bar{Y}_h N_h/N \tag{1.3}$$

with $\bar{Y}_h$ the mean of the sample values coming from stratum h. To see the connection set $X_i=h$ if Y_i comes from stratum h and set $\delta\{X\}=1$ if X is 0, $=0$ otherwise. Then

$$\bar{Y}_h=\sum_i Y_i\delta\{X_i-h\}/\sum_i \delta\{X_i-h\}$$

and (1.3) is seen to be of the form (1.2) with

$$g(X)=N_h/N$$

and

$$\hat{f}(X)=\sum_i \delta\{X_i-h\}/n$$

if $X=h$. The procedure of the present paper is referred to as a continuous form of post-stratification because it may be viewed as corresponding to allowing the stratum label h to take on all values in a continuum.

Returning to the problem of constructing an estimate of the value (1.1), Priestley and Chao [9] propose

$$\hat{\mu}(X)=\sum Y_i(X_i-X_{i-1})\,W_n(X-X_i)$$

as an estimate of $\mu(X)$ where $W_n(x)$ is a weight function concentrated near $x=0$. This suggests as an estimate of θ

$$\sum Y_i(X_i-X_{i-1})\int g(x)\,W_n(x-X_i)dx=\sum Y_i(X_i-X_{i-1})g(X_i)$$

an expression of the form of (1.2) with $\hat{f}(x)$ a nearest neighbor type density estimate. Quite a broad array of estimates of regression functions, $\mu(X)$, have now been proposed and certain of their properties investigated. (See for example Watson [15], Nadaraya [8], Rosenblatt [11], Priestley and Chao [9], Stone [13], Singh [12], Benedetti [1], Clark [3], Stone [14].) Those properties while suggestive, do not lead immediately to the interesting properties of the statistic (1.2).

The concern of this paper is to develop certain useful results concerning the estimate (1.2). Generally the concern is with the, apparently more relevent, case of fixed X's, however some comments are made concerning the case of stochastic X's.

2. Initial results and assumptions

It is assumed that $X_1, X_2, \cdots$ are fixed values. It is assumed that $Y_1, Y_2, \cdots$ are independent random variables with $\mathrm{E}\{Y_i|X_i\}=\mu(X_i)$, $\mathrm{var}\{Y_i|X_i\}=\sigma^2(X_i)$. This implies immediately that

$$\mathrm{E}\,\hat{\theta}=n^{-1}\sum \mu(X_i)g(X_i)/\hat{f}(X_i) \tag{2.1}$$

$$\mathrm{var}\,\hat{\theta}=n^{-2}\sum \sigma^2(X_i)g(X_i)^2/\hat{f}(X_i)^2 \tag{2.2}$$

and that, under a minimal further condition (given for example in Feller [5], p. 256), $\hat{\theta}$ is asymptotically normal.

Before serious thought can be given to the estimate $\hat{\theta}$ though, it is necessary to investigate whether the quantity of expression (2.1) is near θ of (1.1) and to indicate an estimate of the variance (2.2). If $\hat{\mu}(X)$ is an estimate of the regression function $\mu(X)$, then an obvious estimate to consider for (2.2) is

$$n^{-2}\sum |Y_i-\hat{\mu}(X_i)|^2 g(X_i)^2/\hat{f}(X_i)^2\,. \tag{2.3}$$

This statistic will be investigated later. At this moment further conditions are set down concerning the X_i and $g(\cdot)$, $\mu(\cdot)$ and $\sigma(\cdot)$.

Let $F_n(x)$ denote the empiric distribution function of the X's, that is

$$F_n(x)=\{\text{number of } X_i\leqq x;\ i=1, \cdots, n\}/n\,. \tag{2.4}$$

ASSUMPTION I. There exists a function $F(x)$ with the properties (i)

$$d_n=\sup_x |F_n(x)-F(x)|\to 0\,, \qquad \text{as } n\to\infty, \tag{2.5}$$

(ii) $F(x)$ has a bounded third derivative, (iii) its first derivative $f(x)$ is such that $f(x)\geqq\varepsilon>0$ for some ε and all x.

If the X's are a sequence of independent identically distributed random variables, then (2.5) holds almost surely with $d_n=(2n)^{-1/2}(\log\log n)^{1/2}$, see Chung [2]. Assumption I forces the X's to have an asymptotic density.

Suppose next that the density estimate, $\hat{f}(x)$, appearing in the estimate (1.2) takes the form

$$\hat{f}(x)=n^{-1}\sum_i W_n(x-X_i)=\int W_n(x-u)dF_n(u)$$

where $W_n(x)=b_n^{-1}W(b_n^{-1}x)$ with b_n a binwidth parameter tending to 0 as $n\to\infty$ and with $W(\cdot)$ satisfying,

ASSUMPTION II. $W(x)$, $-\infty<x<\infty$, satisfies $W(-x)=W(x)$, $\int W(x)dx=1$,

$$\int |W(x)|dx, \int |x|^2|W(x)|dx, \int |W'(x)|dx<\infty ,$$

where $W'(x)$ denotes the derivative of $W(x)$.

Define

$$f^w(x)=\int W_n(x-u)f(u)du .$$

With Assumptions I and II one can measure the nearness of $\hat{f}(\cdot)$ to $f^w(\cdot)$ and to $f(\cdot)$. Specifically, by Taylor expansion it is clear that

$$\sup_x |f^w(x)-f(x)|\leqq\frac{1}{2}\sup|f''|\int |u|^2|W(u)|du\, b_n^2 \tag{2.6}$$

where f'' denotes the second derivative of f. Next one has

$$\begin{aligned}\hat{f}(x)-f^w(x)&=\int W_n(x-u)d[F_n(u)-F(u)]\\&=\int [F_n(u)-F(u)]W_n'(x-u)du\end{aligned}$$

after an integration by parts and so

$$\sup_x |\hat{f}(x)-f^w(x)|\leqq b_n^{-1}d_n\int |W'(x)|dx . \tag{2.7}$$

Combining (2.6) and (2.7) one has

$$\sup_x |\hat{f}(x)-f(x)|=O(b_n^2)+O(b_n^{-1}d_n) . \tag{2.8}$$

The expression will tend to 0 provided b_n, $b_n^{-1}d_n\to 0$.

Expression (2.8) may be used to measure the nearness of $\mathrm{E}\,\hat{\theta}$ of (2.1) to the desired θ of (1.1). First it is necessary to set down,

ASSUMPTION III. The functions $g(x)$, $\mu(x)$, $-\infty<x<\infty$, are such that

$$\int |\mu(x)g(x)|dx<\infty$$

and the function $\mu(x)g(x)$ is of bounded variation.

Now one has

$$
\begin{aligned}
&|\mathrm{E}\,\hat{\theta}-n^{-1}\textstyle\sum \mu(X_i)g(X_i)/f(X_i)| \\
&\quad\leqq n^{-1}\textstyle\sum |\mu(X_i)||g(X_i)||\hat{f}(X_i)-f(X_i)|/\hat{f}(X_i)f(X_i) \\
&\quad\leqq n^{-1}(\inf \hat{f}(x))^{-1}\sup|\hat{f}(x)-f(x)|\textstyle\sum |\mu(X_i)||g(X_i)|/|f(X_i)| \\
&\quad= O(b_n^2)+O(b_n^{-1}d_n)
\end{aligned}
$$

using (2.8) and the Lemma in the Appendix. That Lemma also implies that

$$
n^{-1}\textstyle\sum \mu(X_i)g(X_i)/f(X_i)=\int \mu(x)g(x)dx+O(d_n)
$$

and so one has

$$
\mathrm{E}\,\hat{\theta}=\int \mu(x)g(x)dx+O(d_n)+O(b_n^2)+O(b_n^{-1}d_n)\,. \tag{2.9}
$$

Were this paper simply an exercise in numerical integration, investigating the degree of approximation of the integral (1.1) by the finite sum (1.2), with no errors present (that is $\sigma(X)\equiv 0$), then expression (2.9) would give the answer desired.

Because stochastic errors have been assumed present it is of interest to investigate the variance estimate (2.3). That statistic is based upon an estimate, $\hat{\mu}(X)$, of the regression function $\mu(X)$. As indicated earlier, quite a number of estimates have been proposed. In connection with the one used suppose,

ASSUMPTION IV. The estimate $\hat{\mu}(X)$ has mean-squared error

$$
\tau_n^2(X)=\mathrm{E}\,|\hat{\mu}(X)-\mu(X)|^2
$$

satisfying $\tau_n^2(X)\leqq\varepsilon_n\sigma^2(X)$ with $\varepsilon_n\to 0$ as $n\to\infty$.

A final requirement is,

ASSUMPTION V. The functions $f(\cdot)$, $g(\cdot)$, $\sigma(\cdot)$ are such that

$$
\int \sigma^2(x)g(x)^2/f(x)dx<\infty\,,
$$

$\sigma^2(\cdot)g(\cdot)^2/f(\cdot)$ is of bounded variation and $f(\cdot)$ is bounded.

With these assumptions and arguing as before

$$
\mathrm{E}\,\{n^{-2}\textstyle\sum |Y_i-\hat{\mu}(X_i)|^2g(X_i)^2/\hat{f}(X_i)^2\}\sim \mathrm{var}\,\hat{\theta}\sim n^{-1}\int \sigma^2(x)g(x)^2/f(x)dx\,.
$$

The variance estimate is asymptotically unbiased.

In summary, an estimate of the integral (1.1) has been constructed

that is asymptotically unbiased, consistent, normal and an estimate of its variance has been investigated.

3. Further remarks

The discussion following Assumption III indicates that the more nearly constant are the functions $f(x)$ and $\mu(x)g(x)$ the less biased the estimate $\hat{\theta}$ may be expected to be. This suggests that prior to computing the estimate, the X values should be transformed, whenever possible, to make these functions more nearly constant.

In the case that $g(x) \equiv f(x)$, the problem is seen to be that of estimating $\mathrm{E}\, Y$. The estimate under discussion in the paper is one making use of supplementary information concerning a variate X, rather than the simple mean of the Y's.

X has been taken to be real-valued. This is no real restriction, the development for multidimensional X follows quite directly. The Lemma of the Appendix may be replaced by one of Zaremba [16]. The previously mentioned references on non-parametric regression function estimation, generally provide estimates for the multidimensional case.

It has been assumed throughout that the X's are fixed, rather than stochastic. This seems the more appropriate inferential basis. Were the asymptotic distribution desired for the case of (X_i, Y_i), $i=1, \cdots, n$, a sample from a multidimensional distribution, this could be found using the asymptotic representation of Révész [10] for a multidimensional empiric distribution.

In order to save computations, one might choose to estimate the variance of $\hat{\theta}$ via the jackknife procedure, (Mosteller and Tukey [7], Chapter 8). This however will lead to an overestimate of the variance as it is appropriate to the aforementioned case of stochastic X.

Acknowledgements

I would like to thank Professor B. A. Bolt and Dr. R. Uhrhammer, of the Seismographic Station, University of California, Berkeley for discussions leading up to the formulation of the problem considered in this paper.

Appendix

LEMMA. *Let $h(x)$ be a function with variation $V(h)$. Let $F(x)$ be a distribution function on $(-\infty, \infty)$ and given $X_1, \cdots, X_n$ let $F_n(x)$ be given by* (2.4) *and let d_n be given by* (2.5). *Then*

$$\left| n^{-1} \sum h(X_i) - \int h(x) dF(x) \right| \leqq V(h) d_n .$$

PROOF. Follows directly from the corresponding result for the uniform distribution given in Koksma [6].

THE UNIVERSITY OF CALIFORNIA, BERKELEY

REFERENCES

[1] Benedetti, J. K. (1977). On the nonparametric estimation of regression functions, *J. R. Statist. Soc.*, B, **39**, 248-253.

[2] Chung, K. L. (1949). An estimate concerning the Kolmogorov limit distribution, *Trans. Amer. Math. Soc.*, **67**, 36-50.

[3] Clark, R. M. (1977). Non-parametric estimation of a smooth regression function, *J. R. Statist. Soc.*, B, **39**, 107-113.

[4] Cochran, W. G. (1963). *Sampling Techniques*, J. Wiley, New York.

[5] Feller, W. (1966). *An Introduction to Probability Theory and its Applications*, Vol. II, J. Wiley, New York.

[6] Koksma, J. F. (1942). A general theorem from the theory of uniform distribution modulo 1, *Mathematica Zutphen* B, **11**, 7-11.

[7] Mosteller, F. and Tukey, J. W. (1977). *Data Analysis and Regression*, Addison-Wesley, Reading, Mass.

[8] Nadaraya, E. A. (1964). On estimating regression, *Theory Prob. Appl.*, **9**, 141-142.

[9] Priestley, M. B. and Chao, M. T. (1972). Non-parametric function fitting, *J. R. Statist. Soc.*, B, **34**, 385-392.

[10] Révész, P. (1976). On strong approximation of the multidimensional empirical process, *Ann. Prob.*, **4**, 729-743.

[11] Rosenblatt, M. (1969). Conditional probability density and regression estimates, in *Multivariate Analysis*-II (ed. P. R. Krishnaiah), Academic, New York, 25-31.

[12] Singh. R. S. (1977). Applications of estimators of a density and its derivatives to certain statistical problems, *J. R. Statist. Soc.*, B, **39**, 357-363.

[13] Stone, C. (1975). Nearest neighbor estimators of a nonlinear regression function, in *Proc. Computer Sci. and Statistics: 8th Annual Symp. on the Interface*, Los Angeles, Univ. of Calif., 413-418.

[14] Stone, C. (1977). Consistent nonparametric regression, *Ann. Statist.*, **5**, 595-620.

[15] Watson, G. S. (1964). Smooth regression analysis, *Sankhyā*, A, **26**, 359-372.

[16] Zaremba, S. K. (1968), Some applications of multidimensional integration by parts, *Ann. Polon. Math.*, **21**, 85-96.