“Non-Parametric Regression with a Functional Independent Variable”

Chuck Fuhrer
The Segal Company

38th Annual Actuarial Research Conference
August 9, 2003
University of Michigan
Ann Arbor, Michigan

Abstract

Pricing health insurance presents numerous statistical problems. One of these is that the usage of health care providers varies depends on the benefit plan's reimbursement level. This reimbursement level is a function of the amount of charges incurred for health care. Another problem is that the distribution of charges has a heavy tail with a variance that far exceeds the mean. I develop a non-parametric method for handling these problems. This method involves kernel density estimation of the charge level distribution and also kernel smoothing between the reimbursement function derivatives. I derive some properties of the estimates so obtained. I illustrate the method with some dental data and compare the results to some other simpler estimation techniques.

Outline

- The Reimbursement Function
- Health Coverages Characteristics
- The Problem
- A Solution
- Further Research
- Help

The Reimbursement Function

- *r(x)*, the amount paid to the insured for allowable charges *x*.
- Basic Properties
  1. continuous
  2. non-decreasing
  3. $0 \le r(x) \le x$
  4. $0 \le r(x_2) - r(x_1) \le x_2 - x_1$
- Usually
  1. piecewise linear
  2. $r(x) = c(x-d)^+ + \left(1-c)\left(x-d-\frac{l}{1-c}\right)^+\right)$ for constants: $0 < c < 1$, $d > 0$, and $l > 0$. $x^+ = \max\{x, 0\}$.

Health Coverages Characteristics

Amount of data is huge, but needs to be subdivided by:

1. Area
2. Age/Sex
3. Plan

Claims are:

1. Assumed never negative
2. Highly skewed.
3. Thick tailed

The Problem

- Charges random variable depends on plan (reimbursement function *r*).

- Need to estimate $E[r(X_r)]$.

- Data Available

Set of observations on *m* different plans. Let $x_{i,j}$ be the *i*'th observation on ($i = 1,2,3,\ldots n_j$) reimbursement function r_j ($j = 1,2,3,\ldots m$).

A Solution

Estimate the density $f(x,r)$ of X_r with the Kernel estimate on both parameters:

$$f(x,r) = \frac{1}{m}\sum_{j=1}^{m}\frac{1}{n_j}\sum_{i=1}^{n_j}\frac{1}{h_{i,j}}K_d\left(\frac{x - x_{i,j}}{h_{i,j}}\right)\frac{1}{k}K_r\left(\frac{r'(x) - r'_j(x_{i,j})}{k}\right)$$

For suitable Kernel functions: K_d and K_r.

Notes on Kernel functions:

1. Symmetric so $\int_{-\infty}^{\infty} xK(x)dx = 0$
2. $K(x) \geq 0$
3. $\int_{-\infty}^{\infty} K(x)dx = 1$.

Choose the bandwidths $h_{i,j}$ using an adaptive method so wider in the tail.

Set: $E[r(X_r)] = \dfrac{\int_0^{\infty} f(x,r)r(x)dx}{\int_0^{\infty} f(x,r)dx}$

Further Research

To dos
1. Test on some data
2. Estimate MSE

Other Problems
1. Multidimensional random variable
   a. Too many dimensions and sparse data
   b. One dimension's distribution depends on reimbursement in another one.
   c. Choice between options
2. Distribution is actually compound of a discrete distribution and reimbursement functions depend on discrete distribution also (e.g. copays).

HELP!

1. Discuss solution above and alternatives
2. Has this been done before? I don’t have access to statistical journals.

Bibliography

Reimbursement Functions

Fuhrer, Charles S. and Shapiro, Arnold F., "Modeling Flexible Benefit Selection", *Transactions of the Society of Actuaries*: v.44 (1992)

Density and Non-parametric Regression

Devroye, Luc, *A Course in Density Estimation*, Birkhäuser: 1987

Härdle, Wolfgang, *Applied Non-parametric Regression,* Cambridge: 1990.

Scott, David W., *Multivariate Density Estimation*, Wiley: 1992.

Silverman, B.W., *Density Estimations for Statistics and data Analysis*, Chapman and Hall: 1986