

Nonparametric Maximum Likelihood Estimation on Bivariate Censored Data Through Graph Theory

ABSTRACT

In the health or medicine, the incubation period is the time from the infection of agent exposure until the signs and symptoms of the disease appear or from the entry of germs/viruses into the human body until early symptoms of disease appear. The incubation period of HIV to AIDS is defined as the time between HIV infection and AIDS. According to Maathuis (2003), the lower limit of the incubation period of HIV to AIDS is 6 years and the upper limit is 13 years. If X represents the time interval of infection i.e $(x_L, x_R]$ and Y represents the time interval of onset i.e $(y_L, y_u]$, then as well as estimating the joint distribution function of the time of HIV infection and AIDS onset time. We use non-parametric method because no specific assumptions that must be met, the free distribution it has no close form shape, and the results with the numerical approach.

NPMLE calculation in the case of bivariate interval censored data is composed of two parts, the first is determined the intersection area of rectangles and the second is find the maximum likelihood. The results obtained in this study is the intersection graph algorithm complexity $O(n^2)$ and the complexity of obtaining estimator with rooted tree algorithm $O(n^3)$. If the value of first derivative of the maximum likelihood function is less than or equal to the rank of clique matrix \mathbf{A} and the second derivative (Hessian matrix) is negative definite then the likelihood function is a strictly concave function. The estimator $\hat{\mathbf{p}}$ is unique. The main advantages of finding the rooted tree algorithm is the simplification of NPMLE computation. Estimator obtained is self consistency estimator and every NPLME is self consistency estimator.

Keywords: NPMLE, bivariate censored data, rooted tree, incubation period, self consistency estimator.

***Nonparametric Maximum Likelihood Estimation* pada Data Tersensor Bivariat Melalui Teori Graf**

INTISARI

Dalam bidang kesehatan atau kedokteran, masa inkubasi adalah waktu dari saat paparan agen menular sampai tanda-tanda dan gejala penyakit muncul atau masuknya kuman/virus ke dalam tubuh manusia sampai timbul gejala awal penyakit. Contoh waktu inkubasi dari HIV menjadi AIDS didefinisikan sebagai waktu antara terinfeksi HIV dan terjangkitnya AIDS. Menurut Maathuis (2003), batas bawah waktu inkubasi HIV menjadi AIDS adalah 6 tahun dan batas atasnya adalah 13 tahun. Misalkan X menyatakan interval waktu terinfeksi (x_{Left}, x_{Right}] dan Y menyatakan interval waktu terjangkit (y_{Lower}, y_{Up}], maka untuk menyelesaikan permasalahan estimasi distribusi waktu inkubasi sama halnya mengestimasi fungsi distribusi gabungan dari waktu terinfeksi dan waktu terjangkit. Metode yang digunakan untuk menyelesaikan yaitu metode nonparametrik, karena tidak ada asumsi tertentu yang dipenuhi, bebas distribusi, tidak mempunyai bentuk yang *close form*, dan hasilnya dengan pendekatan numerik.

Perhitungan *Nonparametric Maximum Likelihood Estimation* (NPMLE) pada kasus data tersensor interval bivariat terdiri dari dua bagian, pertama menentukan daerah irisan persegi panjang dan kedua mencari *maximum likelihood*. Hasil yang diperoleh dalam penelitian ini adalah kompleksitas waktu algoritma graf irisan adalah $O(n^2)$ Adapun kelemahan dari algoritma graf irisan adalah titik koordinat harus merupakan bilangan bulat. Kompleksitas waktu memperoleh estimator dengan algoritma pohon berakar adalah $O(n^3)$ dan keunggulan dari algoritma pohon berakar adalah proses penyederhanaan dalam perhitungan untuk memperoleh estimator. Estimator yang diperoleh merupakan *self consistency* estimator, yaitu setiap NPLME merupakan *self consistency* estimator, tetapi tidak sebaliknya.

Kata kunci: NPMLE, data tersensor bivariat, pohon berakar, waktu inkubasi, *self consistency* estimator.