Least Squares solution for fitting beta distribution to empirical distrbution

Source: http://sci.tech−archive.net/Archive/sci.stat.math/2005−02/0540.html

---

clemenr_at_wmin.ac.uk
Date: 02/23/05

```
Date: 23 Feb 2005 02:45:18 −0800
```

Hi.

I would like to look at the following case. I have a set of data that I can use as a probability distribution using kernel density estimation. I would like to find the best possible (I'm thinking at the moment, least squares) beta distribution to approximate this empirical distribution.

It would be easy for me to find the parameters for the beta distribution by some form of stochastic search, such as a genetic algorithm or simulated annealing*

*in actual fact I'd use another technique, but I don't want to get into those arguments now.

However, I'm wondering if there is an analytic solution to this.

My history of finding analytic solutions is that about a year or so ago I had a (successful) go at deriving maximum likelihood estimates for the mean and sd of the normal distribution using the mathematical software Maxima.

Assuming that b(x, s,f) is the pdf for the beta distribution, and that k(x, data) is a kernel function returning the estimated density, both for a value x, 0 <= x <= 1, then I can define the squared error as:

squared_error(s, f) = SUM (x in data) (b(x, s, f) − k(x, data))
^ 2

A least squares estimate for s and f could then be found by differentiating the equation, and solving for d squared error d s,f =
0.

What I would like to ask is this: Is this likely to work out. Which, as far as I can see means "will I be able to solve the differential equation and will there be a single minima?" Or, is there a better solution? Or, is there a reference I could look at to find a well−known solution? If it's solvable, but not the kind of thing that is printed in books or papers, then I'd ask people not to solve it for me as I'd like to try doing so myself.

I do realise that choosing the kernel function is likely to be tricky, as (i) different kernel functions may make it more or less difficult (or impossible) to solve the differential equation and/or may mean that there are more or fewer minima. I am wondering if the same kernel function with different parameters could affect the number of minima, making it difficult or impossible to find an analytic solution. I don't intend to d this by hand, but to use Maxima or similar software.

Any hints?

Cheers,

Ross−c