Nat. Hazards Earth Syst. Sci. Discuss., 3, C2455–C2456, 2015
www.nat-hazards-earth-syst-sci-discuss.net/3/C2455/2015/

Natural Hazards and Earth System Sciences
Discussions

Interactive comment on "Exploration of diffusion kernel density estimation in agricultural drought risk analysis: a case study in Shandong, China" *by* C. Wen et al.

Anonymous Referee #1

Received and published: 30 November 2015

This is a very good illustration of the effects of boundary bias in kernel density estimation on a real dataset with importance in policy making.

I am not an expert on droughts and hydrology (that part of the paper appears reasonable to me), so I will restrict my comments only to the kernel density estimation technique used in this study.

My main point, demanding some additional comments in the paper, is listed as number one below. The rest of the points are minor.

1. The authors need to be aware that the bandwidth needed to estimate the cumu-

lative distribution function (cdf) is not the same as that needed to estimate the probability density function (pdf).

Thus, estimating the pdf with the optimal bandwidth, (t_X, t_Y), and then simply taking the cumulative sum to obtain a cdf yields a slightly oversmoothed (from the theoretical optimal) estimate for the cdf.

To put it another way, the optimal bandwidth for estimating the pdf is typically slightly larger than the optimal bandwidth for estimating the cdf. This is because the empirical cdf, already being much smoother than the empirical pdf, requires a smaller bandwidth (less smoothing) than the pdf to achieve optimal estimation.

For example, in one dimension the optimal bandwidth t for pdf estimation decays to zero at the rate $O(N^{-2/5})$, but the optimal bandwidth for the cdf estimation decays at the faster rate of $O(N^{-2/3})$.

In two dimensions the bandwidth decay rates will be modified from $O(N^{-1/3})$ for pdf to $O(N^{-1/2})$ for cdf.

I think the authors should mention this in the theoretical description.

I do not think there is a need to redo the density estimates with the modified bandwidth for the cdf, because the figures will be little changed. Nevertheless, this issue has to be mentioned.

2. Something is wrong with formula (9), because we have a useless index j. The authors should consult the original source.

3. just after equation (10) – $\psi_{i,j}$ is not a di-gamma function, but simply notation used to denote the integral in (11)

4. A small sample of typos I noticed: just after equation (10) – 'evaluated at $\hat{t}_{i,j}$'

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., 3, 6757, 2015.