

Statistics 240 Lecture Notes

P.B. Stark www.stat.berkeley.edu/~stark/index.html

Revised 9 November 2005

1 Assignment 3

1. Show that the collection of all sets of the form $(-\infty, x] \times (-\infty, y]$ comprise a Vapnik-Cervonenkis class (V-C class) over the plane.
2. Show that intersections and finite unions of V-C classes are V-C classes. Show that countable unions of V-C classes need not be V-C classes.
3. The file www.stat.berkeley.edu/~stark/Java/Data/lomaPrieta.dat contains 221 observations of the times of putative aftershocks of the 17 October 1989 earthquake in Loma Prieta, California. There are 222 lines in the file. The first is 0, the main shock, which occurred at 4:15:43pm. The other lines are the times in days from the main event to the aftershocks, defined as earthquakes determined to have magnitude 3.0 and above, focal depth of 0–20km, and epicenter within 40km of the epicenter of the Loma Prieta earthquake. The data are from the UC Berkeley Seismographic Stations, courtesy of Dr. Bob Uhrhammer.
 - (a) According to one theory, aftershocks follow the modified Omori Law. If so, the data are an iid sample of size 221, sorted into increasing order, from the density $C/(a+t)^b$, where C is a normalizing constant, a and b are parameters, and t is time in the interval 0 to 805 days.
 - i. There are natural restrictions on a and b for the density to peak at 0 and decrease monotonically. (Aftershocks generally are most frequent immediately after the main shock, then decrease in frequency.) What are those restrictions?

- ii. Find the MLE for a and b from the Loma Prieta data.
- iii. Compute the observed information matrix of the parameter estimates.
- iv. Is the MLE biased in this application? Is the observed information matrix a good approximation to the variance-covariance matrix of a and b ?

Hint: see See pp. 109–116 and Labs 10 and 11 in Freedman (2005), *Statistical Models: Theory and Practice*.

- (b) Estimate the probability density of aftershocks of the Loma Prieta earthquake for the time interval 0 to 805 days, conditional on the event that there were 221 aftershocks during that period. (Assume that, conditional on the total number of aftershocks in the interval, the times of the aftershocks are iid random variables with common density f , whose functional form is unknown; the data are those times sorted into increasing order.)

Use the following estimators:

- i. Histogram (use several choices of bin width).
- ii. The naive estimator (use several bin widths).
- iii. A kernel density estimate (use several bandwidths) The kernel you use is up to you, but say why you chose it.
- iv. The nearest neighbor estimate (use several neighborhood sizes).

Compare and contrast the estimates. You might want to plot the density estimates on a semi-logarithmic scale (linear in time, but logarithmic in the density: the rate of aftershocks is extremely high in the first few days). Please implement the algorithms yourself—do not use “canned” routines from a package. I think you will get a better feel for the methods if you have to code them. Feel free to use a canned package to check your results.