Problems with finding bandwidth via MLE

Calculating the bandwidth via maximum likelihood estimation (MLE) is a procedure for which there are not many literary references. Although a seemingly simple concept, the implementation proved to be more complicated than one would suspect. Using the R statistical programming application, the function optim was utilized for finding the optimal bandwidth using MLE. The function being maximized evolved throughout the procedure as problems, mathematically, programmatically, and contextually, were discovered.

Evolution of the negative log-likelihood function

Beginning with the form of the log-likelihood function for a point process model, and using the exponential distribution for the kernel density leads to the negative log-likelihood function given in Equation #. In the equation h is the bandwidth being estimated, r_i is the minimum distance from the point (x_i, y_i) to border, and $d_{(i,j)}$ is the distance between the point (x_i, y_i) in the modeling data set to the each fire in the background data set. The indicator variable in the function, denoted as $1_{d_{(i,j)}}$ is assigned a value of one for those points where a fire in the background data set is closer than the county boundary line. For each fire in the modeling data set, each fire in the background data set is compared to the boundary and assigned an indicator value, that is the indicator is actually a 556 row by 670 column matrix consisting of either ones or zeros. It is important to note that the distance to the border took into consideration only the minimum and maximum values of both the x and y axes, not each point along the county line.

$$\int \lambda\left(x_i, y_i\right) = \sum_{i=1}^{670} 2\pi \left\{1 - \left[\exp\left(-hr_i\right)\left(1 + hr_i\right)\right]\right\}$$

$$\sum \log \lambda\left(x_i, y_i\right) = \sum_{j=1}^{556} \log \sum_{i=1}^{670} h^2 \exp\left(-hd_{(i,j)}\right) 1_{d_{(i,j)}}$$

$$-\log L(\theta) = \int \lambda\left(x_i, y_i\right) - \sum \log \lambda\left(x_i, y_i\right)$$

For this function, it is more manageable to separate the equation into the summation and the integral portions, and then to put these respective pieces back together, once calculated, to form the negative log-likelihood equation to be minimized. Additionally, this separation allows easier and quicker verification of results, as each piece can be verified instead of only having the negative log-likelihood to assess the procedure. The optim function was monitored by outputting the summation portion of the negative log-likelihood, the integral of lambda, the value of the negative log-likelihood, and the parameter estimates at each iteration of the optim algorithm. In addition, there were a range of values that were anticipated which enabled the comparison of the results to the expected values so as to gauge the validity of the results returned by optim.

By monitoring the optim function, it was discovered that the negative log-likelihood function was still flawed, as it was discovered through optim returning unstable estimates of the bandwidth, that is the results of optim drastically varied depending on the initial value of the bandwidth provided to optim. For values of bandwidth greater than 4000, optim simply returned the initial bandwidth after a couple of iterations, while for values less than 4000 it returned a wide range of estimates for the bandwidth, never settling on a consistent estimate.

The problem with the negative log-likelihood function was that the indicator variable only took into consideration four points on the county boundary line. Los Angeles County is not shaped like a square or rectangle, so this caused a problem as the juts in and out of the county are not being accounted for. In fact, by using the four points, the bandwidth included the Pacific Ocean, which we know does not experience wildfires, and surrounding counties, which affects the results of the maximum likelihood estimation by giving inaccurate estimates of the density of fires in Los Angeles County.

To address the problem of the bandwidth reaching beyond the county lines, all points along the county line were used to determine the appropriate value of the indicator variable. Slightly more mathematically involved, the distance of each fire in the modeling data set to each of the points along the boundary was calculated, Equation #. Then, the minimum distance to the boundary was found for each fire in the modeling data set, and used for comparison to the distance from the background fires for the indicator variable. That is, for each point in the modeling data set, the minimum distance to the county line is found. This distance for each point is compared to the distance from each point in the background data set, and if the fire in the background data set is closer than the boundary the indicator is assigned a value of one and zero if the boundary is closer. Just as before, the indicator variable is a 556 row by 670 column matrix.

$$b_{(x,y)} = \sqrt{\left(fire76_{x_i} - boundary_{x_j}\right)^2 + \left(fire76_{y_i} - boundary_{y_j}\right)^2}$$

By defining the indicator variable in this manner, the bandwidth was forced to be within the county lines, as those values where the county line was closer than a background fire was simply given a background rate of zero as opposed to allowing the bandwidth to grow until a background fire was included, which resulted in a bandwidth reaching beyond the county line.

This change in the indicator variable solved the problem of overreaching bandwidths being estimated by the optim function. Once the negative log-likelihood function was determined, optim could be utilized to find the optimal bandwidth for the kernel smoothing of the data.

New section

To begin, the optim function was limited to looking at only positive values for the optimal bandwidth, as it is clearly a requirement of both the kernel smoothing procedure and the data that the bandwidth be non-negative. However, this was rather quickly discovered to not be the only restriction necessary for the optim function. Without a limit for a maximum value for the bandwidth, optim returned the largest value it could find for the 100 iterations it ran during the optimizing process. A bandwidth that is too large will increase the density throughout the area under consideration, and overestimate the influence of the background fires.

Once the limit was bounded, it was discovered that there were 23 fires in the modeling data set which were closer to the county boundary line than any fire in the background data set. This caused a mathematical problem, as it led to taking the log of zero. To address this problem, a small constant was added, calculated as number of fires in the modeling data set divided by the area of Los Angeles County. This constant is the background rate for those areas within Los Angeles County which did not have a fire nearby from 1951 through 1975.

The new form of the negative log-likelihood equation is given in Equation #, where *c* is the constant, *h* is the bandwidth, and *a* is the area of Los Angeles County. With this new equation, both the bandwidth and the constant are being optimized by the optim function. By including the small constant, there is no longer a mathematical error, and because the constant is so small, on the order of 10^{-12}, it does not affect the results of optimization of the bandwidth. The addition of the constant did add an additional constraint to the optim function, as it is required to be a positive value. (ENTER THE STARTING VALUE OF THE CONSTANT)

$$\int \lambda(x_i, y_i) = ca + \sum_{i=1}^{670} 2\pi \left\{1 - \left[\exp(-hr_i)(1 + hr_i)\right]\right\}$$

$$\sum \log \lambda(x_i, y_i) = \sum_{j=1}^{556} \log\left[\sum_{i=1}^{670} c + h^2 \exp\left(-hd_{(i,j)}\right) 1_{d_{(i,j)}}\right]$$

$$-\log L(\theta) = \int \lambda(x_i, y_i) - \sum \log \lambda(x_i, y_i)$$

These manipulations solved mathematical problems of the maximum likelihood method for our data, however little was it known that the optim function itself would pose a problem as well. By design, the optim function tries to optimize the specified parameters on a roughly integer scale. This arises as an issue for both the constant parameter, *c*, and the bandwidth, *h*, though more so for the constant since it is of smaller magnitude.

Due to this function design, optim was simply returning the initial values for both parameters after a single iteration. Essentially, the parameters optim was attempting to optimize were too small for its algorithm, and optim could not attempt values other than the initial values. In order to make optim effective, mathematical manipulation of the parameters was necessary. Rather than simply inputting the parameters directly into the negative log-likelihood equation, the parameters were changed to be calculated prior to the calculation of the negative log-likelihood. That is, for the constant *c*, rather than inputting an initial value to optim on the order of 10^{-12}, the initial value was on the order of 10^{12}, and within the function the constant is calculated by dividing it into 1, Equation #. The same is done for the bandwidth, Equation #. This allows optim to optimize on an integer scale, and still allows for a mathematically correct negative log-likelihood. This change proved critical, and allowed for optim to find the optimal constant and bandwidth for the negative log-likelihood after a number of iterations.

$$c = 1/c_1 \text{ where } c_1 = \text{ is being optimized}$$

$$h = 1/h_1 \text{ where } h_1 = \text{ is being optimized}$$

Once optim had returned parameter estimates, they were verified and validated using two image plots of Los Angeles County which display the smoothing estimate and one with the location of fires in the background data set and the other with the fires in the modeling data set. The first image plot, Figure #, allows us to check the density estimates for the kernel smoothing against the fires in the background data set. For this plot, the density of the smoothing should correspond fairly well to the density of the points plotted, since the smoothing estimate is based on these data points. It appears that the bandwidth suggested by optim is reasonable, as the densities match up fairly well. Areas with a low background rate correspond to areas with few or no fires.

The second image plot, Figure #, allows one to examine the background rate, i.e. smoothing density, against the location of fires in the modeling data set. There are some areas in the plot where there were no fires in the background data set, and a somewhat high frequency of fires in the modeling data set, which is expected. The background rate was calculated with respect to the background data set, not the modeling data set, so the rate should correspond more accurately to the background data set. The background rate is merely an attempt to enumerate the incidence of fire in a specific location within Los Angeles County, and is not required to match up with the modeling data set. By looking at the plot, it is fairly easy to see where new fires have arisen, that is where there are a number of fires with a low background rate.