統計計算 HW3

105024519 許閔翔

1. Suppose we observed the following counts data (sample size $n = 100$):

failure counts (X_i)	0	1	2	3	4	5	6	7	10
occurrence	19	5	10	19	17	17	9	3	1

Obviously, Poisson model does not fit the data well, especially for the situations with very low failure counts ($X_i = 0$ or 1). A modified model is considered:

$$\text{zero-inflated Poisson:} \quad X_i = \begin{cases} \text{POI}(\lambda), & \text{with probability } p, \\ 0, & \text{with probability } 1-p, \end{cases}$$

in which $\lambda > 0$ and $p \in [0, 1]$.

(a) Write the marginal likelihood of (λ, p) given observed data $\{X_i : i = 1, 2, ..., 100\}$, and solve the MLE directly. (coding for the optimization)

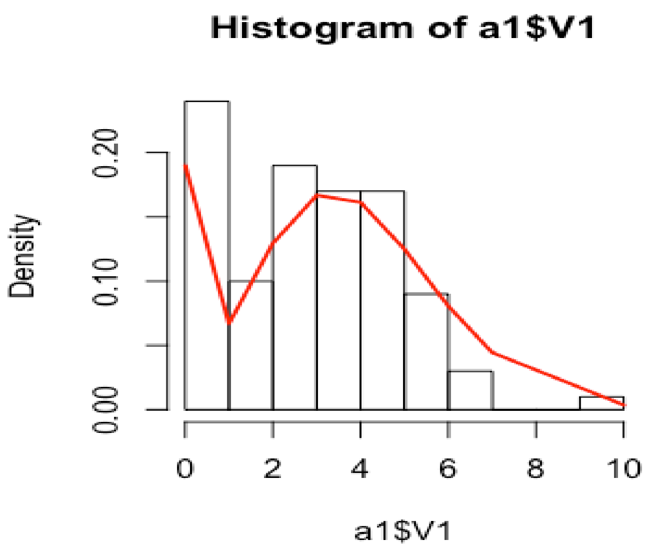
(b) Write the complete-data likelihood by adding latent membership indicators. Solve the MLE using EM algorithm. You need to derive the E-step and M-step theoretically, and implement the algorithm. (coding for the EM algorithm, showing the convergence in plot)

(c) In the same histogram plot above, add the pdf of the fitted zero-inflated Poisson and make some comments.

(a) MLE = (3.8680677, 0.8272886)

(b) MLE = (3.8679227, 0.8272895)

(c) .



2. To analyze data with outliers, trimmed mean estimate is often considered and more robust for estimating the population mean μ. The α-trimmed mean ($\alpha \in [0, 0.5)$) is defined as the average among observations but excluding the left-most (lowest) α proportion and the right-most (highest) α proportion of data.

   (a) Given the data in "hw3_problem2.txt", estimate the population mean μ using three mean estimates: the sample mean, 5%-trimmed mean, and 10%-trimmed mean. Provide the 95% confidence interval for μ based on the above three estimators. In particular, the confidence interval for μ based on the trimmed mean estimators can be obtained via bootstrap method. (State the bootstrap procedure and demonstrate the bootstrap results.)

   (b) Use bootstrap method to quantify the variance and bias for the 5%-trimmed mean estimator.

(a) 0%-trimmed mean: CI (-0.4954436, -0.0310249)
5%-trimmed mean: CI (-0.29216827, 0.05893311)
10%-trimmed mean: CI (-0.3030417, 0.0479672)

(b) Bias: -0.004219598
Variance: 0.008023818

3. Consider the following model for an image on a $K \times K$ grid. Let (i, j) indicate the (row,col) position on the grid, and Y_{ij} be the gray level associated with the (i, j)-th pixel. The image is contaminated with noise. We would like to remove the noise and extract the signal $\{\theta_{i,j} : i = 1, 2, ..., K;\ j = 1, 2, ..., K\}$ smoothed over the space by setting the following optimization problem:

$$\min_{\{\theta_{ij}\}} \left\{ \sum_{i=1}^{K} \sum_{j=1}^{K} (Y_{ij} - \theta_{ij})^2 + \lambda \sum_{i=1}^{K} \sum_{j=2}^{K} (\theta_{i,j} - \theta_{i,j-1})^2 + \lambda \sum_{j=1}^{K} \sum_{i=2}^{K} (\theta_{i,j} - \theta_{i-1,j})^2 \right\}, \qquad (*)$$

where λ is a given constant which controls the smoothness of the signal estimates.

As an example, for $K = 2$, we have the following:

Data

Y_{11}	Y_{12}
Y_{21}	Y_{22}

Signal

θ_{11}	θ_{12}
θ_{21}	θ_{22}

In this special case, the last two terms in the minimization becomes

$$\lambda \left\{ (\theta_{12} - \theta_{11})^2 + (\theta_{22} - \theta_{21})^2 + (\theta_{21} - \theta_{11})^2 + (\theta_{22} - \theta_{12})^2 \right\}.$$

You could start with this very simple example, write down the coordinate descent algorithm for (*). Then, move forward to solve the more general problem given below:

(a) Derive a coordinate descent algorithm for solving $\{\theta_{ij}\}$ for a general K.

(b) Apply your algorithm to the image data in "hw3_problem3.csv" ($K = 200$). Try different λ to see its impact on the signal estimate.

(a)

Define cost_function(t, ti, i, y, lambda, n)…(*)

```
function(n, lambda)
  Initial = rep(0.1, n^2)
  y = goal
  for i in range(1, n^2)
    theta_original = theta_new
    theta_new[i] = optimize(cost_function)
    error = norm(theta_new - theta_original, "2")
      if error >= 10^(-4)
  return theta_new
      if error <= 10^(-4)
  output list(theta_new)
```

$$\min_{\{\theta_{ij}\}} \left\{ \sum_{i=1}^{K}\sum_{j=1}^{K}(Y_{ij} - \theta_{ij})^2 + \lambda \sum_{i=1}^{K}\sum_{j=2}^{K}(\theta_{i,j} - \theta_{i,j-1})^2 + \lambda \sum_{j=1}^{K}\sum_{i=2}^{K}(\theta_{i,j} - \theta_{i-1,j})^2 \right\}, \qquad (*)$$

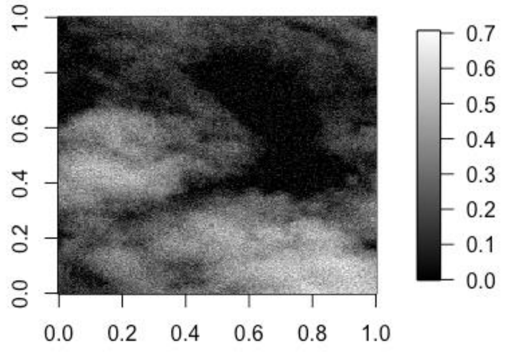
(b)

從上式之中，可以看出我們在做一個最佳化的問題，而 lambda 所乘的式子則是令我們在最佳化的過程有所顧忌。

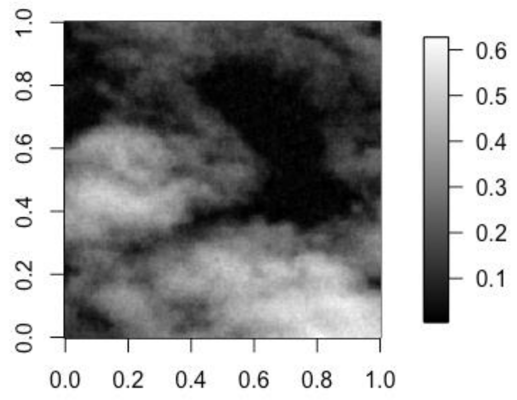
lambda 大代表你在最佳化 y 及 theta 的同時，越需要注意你與前一項也就是左下角的數字差距，導致你的每一次更新遞迴的差距不會太大，以圖片的角度來講，你相鄰的座標之間的像素差距會被限制，表示你的圖片會比較平滑。

lambda 小代表你可以盡情地去最佳化你的 y 跟 theta，所以與前一項也就是左下角的一項之間的差距可以很大比較不受限制。在圖片來講就表示你可以更佳清楚地表示你的圖片像素，也就會更清新。

我們也可以透過控制 Lambda 的大小，去有效地把圖片之中的雜訊消除。



原圖，包含許多雜訊。



lambda = 1

加入小 Lambda，也就是透過一點限制之後，所得到的最佳化結果會使圖片的雜訊變少。

lambda = 1000

則導致兩兩座標之間的差距不會太大，圖片更佳平滑。