Las Vegas Dice Trick

Ziniu Li

1. Problem Introduction

In Las Vegas games, here are two dices. One is normal, which means that each point occur equally probably (i.e.$\mathrm{P}(\mathrm{x} = \mathrm{i}) = 1/6$). However, another dice is unfair, which means that each point occur unequally probably (i.e. $\mathrm{P}(\mathrm{x} = 6) \gg \mathrm{P}(\mathrm{x} = \mathrm{i})$. In fact, we cannot judge whether the unfair is used or not, because the dealer is so cunning. Intuitively, it is safe to say that when the point of six occurs, the possibility of unfair dice is used is large than the possibility of fair dice is used. Now, our question is that how much is the probability of unfair dice is used and how much is the probability of each points occur under fair or unfair dice.

2. Problem Analysis

Let's formalize this problem mathematically. We will use x^i to denote the i_{th} example we observe and $x^i \in \{1, 2, 3, 4, 5, 6\}$. Besides, we will use z_k to denote whether k_{th} dice is used and π_k denotes the probability of occurring (i.e. $\mathrm{P}(z_k = 1) = \pi_k$, here k=1 means fair dice and k = 2 means unfair dice). Finally, we will use w_{ik} denotes the probability of point of i occurs under k_{th} dice. From the definition of probability, π_k satisfies $\sum_{k=1}^{2} \pi_k = 1$ and w_{ik} satisfies $\sum_{i=1}^{6} w_{ik} = 1$.

Suppose that we have data $x^{(1)} \cdots x^{(n)} \cdots x^{(N)}$, here the upper script n represent n_{th} data. We build our model in the following:

$$\mathrm{P}\left(x^{(n)} = i \middle| z^k = 1\right) = w_{ik}$$

$$\mathrm{P}(z_k = 1) = \pi_k$$

We can derivate:

The probability of the point of i occurs, which is the variable we can observe.

$$\mathrm{P}\left(x^{(n)} = i\right) = \sum_{k=1}^{2} \mathrm{P}\left(x^{(n)} = i \middle| z_k = 1\right)\mathrm{P}(z^k = 1)$$

$$= \sum_{k=1}^{2} \pi_k * w_{ik}$$

The posterior probability of i, which is the variable we don't know and need to estimate.

$$\begin{aligned}
\mathrm{P}\left(z_k = 1 \middle| x^{(n)} = i\right) &= \frac{P(z_k = 1, x^{(n)} = i)}{P(x^{(n)} = i)} \\
&= \frac{P\left(x^{(n)} = i \middle| z_k = 1\right)\mathrm{P}(z^k = 1)}{\sum_{k=1}^{2} \mathrm{P}\left(x^{(n)} = i \middle| z_k = 1\right)\mathrm{P}(z^k = 1)} \\
&= \frac{\pi_k * w_{ik}}{\sum_{k=1}^{2} \pi_k * w_{ik}} \qquad (1) \\
&= z_{ik} \textit{(New notation represents the upper result)}
\end{aligned}$$

3. Problem Solution

Our goal is to find the optimal parameter to make use the possible large probability of these observed data occurring. Using the principle of maximum like hood, we can get

$$Likehood = log \prod_{n=1}^{N} P(x^{(n)}) = \sum_{n=1}^{N} log \sum_{k=1}^{2} \pi_k * w_{ik}$$

$$= \sum_{n=1}^{N_1} log \sum_{k=1}^{2} \pi_k * w_{1k} + \sum_{n=1}^{N_2} log \sum_{k=1}^{2} \pi_k * w_{2k} + \cdots \sum_{n=1}^{N_6} log \sum_{k=1}^{2} \pi_k * w_{6k}$$

$$= \sum_{n=1}^{N_1} log \sum_{k=1}^{2} \pi_k * w_{1k} + \sum_{n=1}^{N_2} log \sum_{k=1}^{2} \pi_k * w_{1k} + \cdots$$

$$+ \sum_{n=1}^{N_6} log \sum_{k=1}^{2} \pi_k * (1 - w_{1k} - w_{2k} - \cdots - w_{5k})$$

Next get the partial derivative to optimize like hood:

$$\frac{\partial L}{w_{1k}} = \sum_{n=1}^{N_1} \frac{\pi_k}{\sum_{k=1}^{2} \pi_k * w_{1k}} + \sum_{n=1}^{N_6} \frac{-\pi_k}{\sum_{k=1}^{2} \pi_k (1 - w_{1k} - w_{2k} - \cdots - w_{5k})} = 0$$

$$\sum_{n=1}^{N_1} \frac{\pi_k}{\sum_{k=1}^{2} \pi_k * w_{1k}} + \sum_{n=1}^{N_6} \frac{-\pi_k}{\sum_{k=1}^{2} \pi_k * w_{6k}} = 0$$

$$\sum_{n=1}^{N_1} \frac{\pi_k w_{1k} w_{6k}}{\sum_{k=1}^{2} \pi_k * w_{1k}} = \sum_{n=1}^{N_6} \frac{\pi_k w_{1k} w_{6k}}{\sum_{k=1}^{2} \pi_k * w_{6k}}$$

$$\sum_{n=1}^{N_1} z_{1k} w_{6k} = \sum_{n=1}^{N_6} z_{6k} w_{1k}$$

$$N_1 z_{1k} w_{6k} = N_6 z_{6k} w_{1k}$$

$$\frac{w_{1k}}{w_{6k}} = \frac{N_1 z_{1k}}{N_6 z_{6k}}$$

With the same principle, it is easy to find that

$$w_{1k}: w_{2k}: w_{3k}: w_{4k}: w_{5k}: w_{6k} = N_1 z_{1k}: N_2 z_{2k}: N_3 z_{3k}: N_4 z_{4k}: N_5 z_{5k}: N_6 z_{6k}$$

Remember that $\sum_{i=1}^{6} w_{ik} = 1$, so

$$w_{1k} = \frac{N_1 z_{1k}}{N_1 z_{1k} + N_2 z_{2k} + N_3 z_{3k} + N_4 z_{4k} + N_5 z_{5k} + N_6 z_{6k}}$$

More generally,

$$w_{ik} = \frac{N_i z_{ik}}{\sum_{i=1}^{6} N_i z_{ik}} \tag{2}$$

As for π_k, we can construct the Lagrange multiplier as follows:

$$Lagrange = \sum_{n=1}^{N} log \sum_{k=1}^{2} \pi_k * w_{ik} - \lambda \sum_{k=1}^{2} (\pi_k - 1)$$

Next get the partial derivative to optimize like hood:

$$\frac{\partial L}{\partial \pi_k} = \sum_{n=1}^{N_1} \frac{w_{1k}}{\sum_{k=1}^{2} \pi_k * w_{1k}} + \sum_{n=1}^{N_2} \frac{w_{2k}}{\sum_{k=1}^{2} \pi_k * w_{2k}} \cdots + \sum_{n=1}^{N_6} \frac{w_{6k}}{\sum_{k=1}^{2} \pi_k * w_{6k}} - \lambda = 0$$

$$(\sum_{n=1}^{N_1} \frac{w_{1k}}{\sum_{k=1}^{2} \pi_k * w_{1k}} + \sum_{n=1}^{N_2} \frac{w_{2k}}{\sum_{k=1}^{2} \pi_k * w_{2k}} \cdots + \sum_{n=1}^{N_6} \frac{w_{6k}}{\sum_{k=1}^{2} \pi_k * w_{6k}})\pi_k = \lambda \pi_k$$

$$\sum_{n=1}^{N_1} z_{1k} + \sum_{n=1}^{N_2} z_{2k} + \cdots + \sum_{n=1}^{N_6} z_{6k} = \lambda \pi_k \qquad (3)$$

$$\sum_{k=1}^{2} (\sum_{n=1}^{N_1} z_{1k} + \sum_{n=1}^{N_2} z_{2k} + \cdots + \sum_{n=1}^{N_6} z_{6k}) = \sum_{k=1}^{2} \lambda \pi_k = \lambda$$

So

$$\lambda = \sum_{k=1}^{2} N_1 z_{1k} + N_2 z_{2k} + \cdots + N_6 z_{2k}$$

Take this result back to Lagrange,

$$\pi_k = \frac{N_1 z_{1k} + N_2 z_{2k} + \cdots + N_6 z_{6k}}{\lambda}$$

$$= \frac{N_1 z_{1k} + N_2 z_{2k} + \cdots + N_6 z_{6k}}{\sum_{k=1}^{2} N_1 z_{1k} + N_2 z_{2k} + \cdots + N_6 z_{2k}} \qquad (4)$$

Finally, we can solve this problem according to EM algorithm in the following form:

Algorithm

Initialization: initialize the w_{ik}, π_k and z_{ik} randomly

Solution:

Do {

Step 1: Compute the posterior probability z_{ik}^{new} based on w_{ik}^{old} and π_k^{old} according to (1)

Step 2: Compute the parameter w_{ik}^{new} and π_k^{new} based on z_{ik}^{old} according to (2) and (4)

Step 3: Replace the old parameter with new parameter

$w_{ik}^{old} := w_{ik}^{new}$

$\pi_k^{old} := \pi_k^{new}$

$z_{ik}^{old} := z_{ik}^{new}$

} until converge