EE 126 Probability and Random Processes
April 03, 2018 MAP, MLE II

1 Lecture

MAP / MLE

If p_i is the probability of cause i and q_i is the probability of the symptom given cause i, and there are N causes, then the **posterior probability** of cause i given the symptom is

$$\pi(i) = \frac{p_i q_i}{\sum_{j=1}^{N} p_j q_j}$$

The MAP rule is $i^* = \arg\max_{i \in \{1,\ldots,N\}} p_i q_i$. The MLE rule is $i^* = \arg\max_{i \in \{1,\ldots,N\}} q_i$ ("just find whichever cause has the largest q_i"). More generally,

$$\text{MAP}[X \mid Y = y] = \arg\max_x P(X = x \mid Y = y)$$

and

$$\text{MLE}[X \mid Y = y] = \arg\max_x P(Y = y \mid X = x)$$

Inference in Digital Communication

We send an input bit X over a noisy channel which then produces an observed bit Y. (The BSC is a simple example of such a channel.) We want to find the MLE and MAP estimates of X given Y.

In the case of the BSC "bit flip channel," we observe the bit we sent with probability $1 - p$ and the wrong bit with probability p. In other words, the channel flips the bit with probability p. Also, the prior probabilities are $P(X = 0) = 1 - \alpha$ and $P(X = 1) = \alpha$. By definition,

$$\text{MAP}[X \mid Y = b] = \arg\max_{i \in \{0,1\}} p_i q_i$$
$$\text{MLE}[X \mid Y = b] = \arg\max_{i \in \{0,1\}} q_i$$

where our symptom is $Y = b$. Say we observe a 0 ($Y = 0$). Then

$$p_0 = P(X = 0) = 1 - \alpha$$
$$q_0 = P(Y = 0 \mid X = 0) = 1 - p$$
$$p_1 = P(X = 1) = \alpha$$
$$q_1 = P(Y = 0 \mid X = 1) = p$$

In MAP, we choose 0 if $p_0 q_0 > p_1 q_1$, and 1 otherwise.

In other words, we compare

$$\frac{q_0}{q_1} = \frac{P(Y = b \mid X = 0)}{P(Y = b \mid X = 1)} \quad \text{and} \quad \frac{p_1}{p_0} = \frac{\alpha}{1 - \alpha}$$

The LHS [of the "and"] is the likelihood ratio $L(b)$ of b, and is a function of the observation. The RHS, our threshold, is a function of our prior belief. If the LHS is greater than the RHS, then our estimate is 0. If not, our estimate is 1.

In this case,

$$L(b) = \begin{cases} \frac{1-p}{p} & \text{if } b = 0 \\ \frac{p}{1-p} & \text{if } b = 1 \end{cases}$$

Under MAP: roughly speaking, the likelihood of our observation needs to overpower our prior belief.

Meanwhile, MLE simply declares our estimate of X given $Y = b$ to be b if $p < \frac{1}{2}$. This is the same as MAP with a uniform prior (i.e. where $\alpha = \frac{1}{2}$ and the $\frac{\alpha}{1-\alpha}$ threshold is 1).

Note: if $\alpha = \frac{1}{2}$ then the probability of error is p for the BSC. Considering p can be as high as 1, this can be really bad. How can we improve performance? *By preprocessing X before transmission.* Typically we add redundancy. For example, a repetition code "encodes" X by sending it n times. What's the error under this scheme?

$$\begin{aligned} L(b_1, ..., b_n) &= \frac{P(Y_1 = b_1, Y_2 = b_2, ..., Y_n = b_n \mid X = 0)}{P(Y_1 = b_1, Y_2 = b_2, ..., Y_n = b_n \mid X = 1)} \\ &= L(b_1)L(b_2) \cdots L(b_n) \end{aligned}$$

due to $Y_1, .., Y_n$ being conditionally independent given X (recall that the channel flips are i.i.d.).

Since sums are nicer than products, we can take logs of each $L(b_i)$. The **log-likelihood ratio** $LLR(b_i)$ is

$$LLR(b_i) = \begin{cases} \log\left(\frac{1-p}{p}\right) & \text{if } b_i = 0 \\ \log\left(\frac{p}{1-p}\right) & \text{if } b_i = 1 \end{cases}$$

so $LLR(b_1, ..., b_n) = \log L(b_1, ..., b_n)$ and the generalization of our MAP threshold test for the n-bit case becomes

$$\sum_{i=1}^{n} LLR(b_i) \underset{0}{\overset{1}{\lessgtr}} \log\left(\frac{\alpha}{1-\alpha}\right)$$

(The above operator says "declare 0 if the LHS is greater, and declare 1 otherwise.")

Let U = the number of 0's. Let V = the number of 1's. Then

$$\begin{aligned} \sum_{i=1}^{n} LLR(b_i) &= U \log\left(\frac{1-p}{p}\right) + V\left(-\log\left(\frac{p}{1-p}\right)\right) \\ &= (U - V)\log\left(\frac{1-p}{p}\right) \end{aligned}$$

so the overall MAP rule is

$$U - V \underset{0}{\overset{1}{\lessgtr}} \frac{\log\left(\frac{1-\alpha}{\alpha}\right)}{\log\left(\frac{1-p}{p}\right)}$$

i.e. when $\alpha = \frac{1}{2}$, we declare 0 if we observe more 0's and 1 if we observe more 1's.

The comparison simplifies to $U \underset{0}{\overset{1}{\lessgtr}} V$ (majority rule). Note: we are assuming that $p < \frac{1}{2}$.

However, it turns out that this is not practical because we need n to be too large for it to work. How large does n need to be to guarantee a probability of error of less than 0.1%? *To answer this, we use the MAP rule that the channel must flip less than $\frac{n}{2}$ bits to be successful. We can then estimate using the CLT (where each flip is Bernoulli).*

Example: AWGN (additive white Gaussian noise) channel. Here, Gaussian noise $Z \sim \mathcal{N}(0, \sigma^2)$ is added to the input. We are trying to send real values (voltages) over the channel. If we send -1, the output y is a Gaussian distribution around -1. If we send 1, the output y is a Gaussian distribution around 1. We observe that these two curves intersect at 0. Thus our MAP rule is as follows: if the value is positive, declare 1; if the value is negative, declare 0. If there is a bias, the threshold T will shift left or right:

$$\frac{f_{-1}(y)}{f_1(y)} \underset{-1}{\overset{1}{\lessgtr}} T$$

German Tank Problem

We have a large bin of balls, numbered serially. We draw randomly and get the ball numbered 43. What is the MLE of the number of balls in the bin? *43*. This is clear because $\frac{1}{43} > \frac{1}{44} > \frac{1}{45} > ....$ Of course, this is not a very good estimate since there's no bias.