We wish to give an alternative view of MAP detection. The derivation will be the same, but the point of view will lead to interesting results. In particular, we shall show an example where the sufficient conditions for the theorem on irrelevance (p.19 course reader) does not hold, yet one observation is still irrelevant. The culmination of this exercise will be an example where given two obsevations **Y** and **Z**, and an 'input' **X**, even though the probability of **Z** given **X** and **Y** does not equal the probability of **Z** given **Y**, Y is still irrelevant to the decision device in the MAP detector.

We rederive some of the previous results for review. Recall, we wish to maximize the probability that $\hat{X}$(R)=X given R=r is observed. We do this for the following reasons:

We wish to maximize the probability of having a correct decision, given by:

$$P_c = \sum_r P(\hat{X}(r) = X, R = r)$$

That is, we have a correct decision when we have the state that $\hat{X}$ (r)=X and R=r, otherwise we have made a false decoding. If we write the sum in another way, things become even more evident.

$$\begin{aligned} P_c &= \sum_r P(X = \hat{X}(r), r) \\ &\leq \sum_r \max_x P(X = x, r) \end{aligned}$$

With equality if and only if $\hat{X}$(r)=the maximizing x above. Finally, we note that X and R can be waveforms, vectors, or anything. The only thing that is important is that the statistical properties of the system are defined.

We now demonstrate the duality with the treatment in the book. We use Bayes Theorem as follows.

$$\begin{aligned} \max_x P(X = x \| R = r_{observed}) &= \max_x \frac{P(X = x, R = r_{observed})}{P(R = r_{observed})} \\ &\leftrightarrow \max_x P(X = x, r_{observed}) \end{aligned}$$

Since, as noted in the text, the observation is common to all conditional probabilities, and hence the P(R=r) drops out. So, essentially, we want to choose the most common joint observation, as we would intuitively expect. This leads to the very interesting result that if we have an observable quantity R which can be written as R=[Z Y], if P(Z‖Y,X)=P(Z‖Y) then the Z part of the observable quantity R is irrelevant. The book is pretty clear in showing that this observation does not affect the decision which intuitively leads to the fact that the performance does not change either. For those of you who would like a more formal argument, let's prove this: First, we will consider all the times when we have a correct decision. This occurs when the map detector correctly 'guesses' the result, x_{map}, from observed data y and z. This is equivalent to the times when x_{map}, y, and z occur together. So,

$$\begin{aligned} P_{correct} &= \sum_{(x,y,z)\ correct\ decisions} p(x,y,z) \\ &= \sum_{(x_{map}(y,z),y,z)} p(x,y,z) \end{aligned}$$

But, z is irrelevant (does not affect the decision rule) so we can sum over all z in **Z**(i.e. x_{map}(y,z)=x_{map}(y)). So,

$$\begin{aligned} P_{correct} &= \sum_{x_{map}(y),y} \sum_{z} p(x,y,z) \\ P_{correct} &= \sum_{x_{map}(y),y} p(x,y) \end{aligned}$$

which is the probability of having a correct decision using only y in the optimal detector.

Why is all this this relevant? Well, it gives you some intuition into the problem at hand. We would like to determine under what conditions z does not make 'too much' difference on the detector. This might occur when we are trying to decide on what variables are important to test for some detection problem. Moreover, might it be possible that certain variables seem to add information but are still irrelevant to the problem at hand. We contend that this can occur and the construct an example in the following way.

First, we want to choose an input distibution that relates some type of information. Generally, we want to make the input distribution uniform to send the highest amound of information. So, let **X,Y,Z** takes values X1,X2 Y1,Y2 and Z1,Z2 respectively with P(**X**=X1)=P(**X**=X2)=.5. Now, let us define the conditional distribution on **Y** alone, and then add in Z. The key idea here is that if X depends almost wholly on Y, then Z does not matter. So, we choose P(Y1‖X1)=P(Y2‖X2)=.8 and P(Y1‖X2)=P(Y2‖X1)=.2. So, we are making this a good channel with respect to the observation Y.

Now comes the key step. We want to make sure that Z does not tell us too much about the input. This is the same as making the conditional probability of **X** given **Y** and **Z** a coin flip. But, we do not want to do this since this will ensure that the sufficient conditions for irrelevance are met (try this). So, we make the distribution close to a coin flip, but not quite. Moreover, we choose the conditional distributions to ensure that the sufficient conditions cannot be met. i.e. for some **Z**=Z', **Y**=Y', make sure $P(Z'\|Y', X1) \neq P(Z'\|Y', X2)$. So, in a somewhat arbitrary manner:

$$P(Z1\|Y1, X1) = .4$$

$$P(Z1\|Y1, X2) = .6$$

$$P(Z2\|Y1, X1) = .6$$

$$P(Z2\|Y1, X2) = .4$$

$$P(Z1\|Y2, X1) = .4$$

$$P(Z2\|Y2, X1) = .6$$

$$P(Z1\|Y2, X2) = .4$$

$$P(Z2\|Y2, X2) = .6$$

Multiplying out the probabilities, we can clearly see that the decision is still based solely on **Y**:i.e. if we see **Y**=Yi we will decode **X**=Xi. By the above argument on the probability of being correct, we see that the performance does not change. Hence, the conditions stated originally in the book as necessary and sufficient are only sufficient, and we may be able to remove unnecessary observation data anyway.

At this point some of you may notice that we could have just said P(Y1‖X1)=P(Y2‖X2)=1 and P(Y1‖X2)=P(Y2‖X1)=0 (with P(Z1‖Y1,X1)=1 and P(Z2‖Y2,X2)=1, noting (Y1,X2) and (Y2,X1) never occur) To make the example seem a little less academic, we chose numbers that you had to actually to think about.

Hope this helps anyone who may have had questions about this *Irrelevant* Topic.

Atul Salvekar

salvekar@isl.stanford.edu