Random variables in Communication network

Pranav Manu
ECD
2020112019
pranav.m@research.iiit.ac.in

1. Dependent Random variables in Communication

We use Random vectors with independent components can be used to represent information in Communication network. However, in real world, it is not so as each component may be correlated to the other in some way and if we know this correlation we can perform better compression.To understand the correlation, we must know the models of dependence. Another application of dependent random variable is when a sequence of bits is transmitted through a noisy channel, then the sequence received at receiver, in the form of a random vector, depends on the transmitted random vector. If they were independent, then we could never be able to decode the received back to the original message with low probability.

1.1. Forward and inverse probability

Probability calculations fall into two categories: *forward probability* and *inverse probability*. **Forward probability** involves some generative model that describes a process that results in some data, the task then becomes to compute the probability distribution or expectation of some quantity that depends on the data. Entropy calculation falls into this category.

Inverse probability problems also involve some generative model of a process (In this discussion, this process is communication), but here instead of calculating the probability distribution of some quantity as a result of the process, we calculate the probability of one or more unobserved variables in the process, given the observed variables. This concept will be used ahead in the modelling of a decoder to estimate the transmitted sequence, given a received sequence in communication. This involves the use of Baye's Theorem, and can be used in predictions. It can also be used in data compression.

1.2. Types of Entropy

The following Entropy will be used ahead. Entropy can be said as the expected information content carried by a Random variable.

Entropy:

$$H(X) = \sum_{x \in Supp(P_X)} P_X(x).log\frac{1}{P_X(x)}$$

Joint Entropy:

$$H(X,Y) = \sum_{x,y \in Supp(P_{X,Y})} P_{X,Y}(x,y).log\frac{1}{P_{X,Y}(x,y)}$$

Here the entropy represents the expected information carried by X and Y simultaneously.

Conditional Entropy:

$$\begin{aligned} H(X|Y) &= \sum_{y \in Supp(P_Y)} P_Y(y).H(X|Y=y) \\ H(X|Y=y) &= \sum_{x \in Supp(P_{X|Y})} P_{X|Y}(x|y).H(x|y) \end{aligned}$$

This represents the expected information carried by X, given that we know the information in Y. It can also be described as the uncertainty that remains in X after we know Y, as information is the increases as uncertainty increases. We can see that the conditional entropy depends on the conditional probability of X given Y. For independent X and Y:

$$H(X|Y) = H(X)$$

This can be explained by the fact that since they are independent, knowing Y we still cannot say anything about X.

Information Content and chain rule: Information content of X is represented as:

$$\frac{1}{p_X(X = x_i)}$$

And chain rules of probability states that for two random variables:

$$p(x,y) = p(x).p(y|x)$$

Using this in the expression for entropy, we can derive the chain rule for Entropy:

$$H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$$

Mutual induction:

$$I(X;Y) \equiv H(X) - H(X|Y) = H(Y) - H(Y|X)$$

This is used to represent the reduction in information content after we have received Y, or vice versa. This concept is used to define the **channel capacity** or the maximum amount of information that can be sent through a channel. This can also express the average information content expressed by one about the other.

1.3. Gibbs Inequality

Also known as relative entropy or Kullback-Leibler divergence between two probability distribution $P(x)$ and $Q(x)$ defined over the same alphabet A_X, it is a very important concept in information theory. It can loosely be used to define the "distance" between two probability distributions. However it is not strictly distance. It is not symmetric, that is, relative entropy between P and Q is not the same as the relative entropy between Q and P.

$$D(P||Q) = \sum_x P(x).log\frac{P(x)}{Q(x)}$$

2. Communication over a noisy Channel

Channels, or the medium through which information is transmitted, are usually noisy. The need arises develop methods for error-free communication.For this purpose, we will elaborate on noisy error channel and channel coding. Channel coding is used to make the noisy channel behave close to a noiseless channel. The Channel Code is made such that noisy signal received can be decoded. As we have stated before, the measure of the information transmitted can be expressed by the mutual information between the transmitted and received signal.

2.1. Noisy Channels

Conditional probability between the transmitted signal and the received signal can be used to characterize the noisy channel.

Discrete Noiseless Channel(Q): It is characterized by an input alphabet A_X an output alphabet A_Y (Alphabet is the set of values that the message can be mapped to) , and a set of conditional probability distributions $P(y|x)$, one for each $x \in A_X$.

$$Q_{j|i} = P(y = b_j|x = a_i)$$

We can make this into a matrix such that each column is a probability vector, and then obtain the P_Y probability vector by multiplying Q matrix with P_X probability vector. Some models of noisy channel are: **Binary Symmetric Channel:** A_X=0,1. A_Y=0,1

$$P(y = 0|x = 0) = 1 - p;\ P(y = 1|x = 0) = p$$
$$P(y = 0|x = 1) = p;\ P(y = 1|x = 1) = 1 - p$$

Here p is said to be the probability of the transmitted bit to be flipped.

Binary Erasure Channel: A_X=0,1. $A_Y = 0, \epsilon, 1$

$$P(y = 0|x = 0) = 1 - p;\ P(y = 1|x = 0) = 0$$
$$P(y = \epsilon|x = 1) = p;\ P(y = \epsilon|x = 1) = p$$
$$P(y = 1|x = 0) = 0;\ P(y = 1|x = 1) = 1 - p$$

Here p is said to be the probability of the transmitted bit getting erased. ϵ represent the erasure of the bit. The bit is not flipped here, but erased with some probability. The probability of getting erased or flipped in the above noisy channels is conditionally dependent on the transmitted bit.

2.2. Estimating the input given the output

To estimate the transmitted symbol x from the received signal y, we can use the **Bayes' theorem** given y:

$$\begin{aligned} P(x|y) &= \frac{P(y|x).P(x)}{P(y)} \\ &= \frac{P(y|x).P(x)}{\sum_{x_i \in X} P(y|x_i)P(x_i)} \end{aligned}$$

A decent estimate here is: $\underset{x}{argmax}(P(x|y))$

2.3. Information conveyed by a Channel

To measure the amount of information the output conveys about particular input X, we use mutual Information:

$$I(X;Y) \equiv H(X) - H(Y|X)$$

Note that:$I(X;Y) \leq min(H(Y), H(X))$. Intuitively it means that no more than the conveyed information can be received. **Maximising mutual information**

Maximising mutual information conveyed by the channel would mean to maximise the amount of information transferred in communication. We define the channel capacity of a channel Q to be the maximum mutual information over P_X

$$C(Q) = \underset{P_X}{max} I(X;Y)$$

We have control over P_X and can maximising it by optimising the input coding. P_X at channel capacity is called *optimal input distribution*. The rate can be increased only up to the channel capacity if we want to have arbitrarily small probability of error. This is the converse of the **Shannon's Channel Coding theorem.**

2.4. Optimal Decoder for noisy channel coding

Since the transmitted sequence or codeword might have some error due to it being transmitted though a noisy channel. As a result, the received codeword is not the same as the transmitted one. Therefore, a need arises to infer the transmitted message from the received codeword, to enable communication. This is done by the use of a decoder. For a channel, an optimal decoder is one which minimises the probability of error. It decodes an output y as the input x that has maximum posterior probability P(x—y)

$$\begin{aligned} P(x|y) &= \frac{P(y|x).P(x)}{\sum_{x_i \in X} P(y|x_i)P(x_i)} \\ \hat{x}_{optimal} &= argmax P(x|y) \end{aligned}$$

If the prior distribution on x is uniform, then the optimal decoder is called *maximum likelihood decoder* i.e., the estimation is such that the output is such that P(y—x) has the maximum likelihood.

2.5. Gaussian Random variable to model Noise

Noise can also be modelled using Gaussian random variable. Noise can be defined as the sum of infinitely many independent random variable, as there may be several disturbances in the channel. According to *Central Limit theorem*, sum of infinitely many independent random variable results is a Gaussian random variable. The resultant sum can be said as the random variable representing the noise in the channel.