ENEE 627
SPRING 2011
INFORMATION THEORY

THE DISCRETE MEMORYLESS CHANNEL

Throughout, let $\mathcal{X}$ and $\mathcal{Y}$ denote two finite sets, called the input and output alphabets, respectively. For each $n = 1, 2, \ldots$ we shall write elements of $\mathcal{X}^n$ and $\mathcal{Y}^n$ as

$$\boldsymbol{x}^n = (x_1, \ldots, x_n)$$

and

$$\boldsymbol{y}^n = (y_1, \ldots, y_n)$$

with $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ elements of $\mathcal{X}$ and $\mathcal{Y}$, respectively.

Channels

Multiple input symbols are transmitted over a channel, or equivalently the channel is used repeatedly. Due to the vagaries of the communication process these input symbols may be modified or even garbled beyond the point of unrecognition. We modelled these possibilities by specifiying the channel probabilities

$$p_n(\boldsymbol{y}^n|\boldsymbol{x}^n), \qquad \begin{array}{c} \boldsymbol{x}^n \in \mathcal{X}^n \\ \boldsymbol{y}^n \in \mathcal{Y}^n \end{array}$$

for each $n = 1, 2, \ldots$. Thus, if the symbol string $\boldsymbol{x}^n$ is transmitted, then the string $\boldsymbol{y}^n$ of output symbols is received with probability $p_n(\boldsymbol{y}^n|\boldsymbol{x}^n)$.

The *memoryless* assumption stipulates that

$$p_n(\boldsymbol{y}^n|\boldsymbol{x}^n) = \prod_{k=1}^{n} p(y_k|x_k), \qquad \begin{array}{c} \boldsymbol{x}^n \in \mathcal{X}^n \\ \boldsymbol{y}^n \in \mathcal{Y}^n \end{array} \tag{1}$$

for each $n = 1, 2, \ldots$ where

$$p(y|x) = \begin{array}{c} \text{Probability that the symbol } y \ (\in \mathcal{Y}) \text{ is received by the receiver} \\ \text{given that the symbol } x \ (\in \mathcal{X}) \text{ was sent} \end{array}$$

Obviously we have $p_1(y|x) = p(y|x)$.

Sometimes it is convenient to organize these probabilities into the $|\mathcal{X}| \times |\mathcal{Y}|$ matrix $\boldsymbol{P}$, known as the *channel matrix* and given by

$$\boldsymbol{P} \equiv (p(y|x),\ x \in \mathcal{X},\ y \in \mathcal{Y}) .$$

Any channel which operates according to (1) is known as a *discrete memoryless channel* (DMC) with channel matrix $\boldsymbol{P}$.

Implementing DMCs

Assume some probability triple $(\Omega, \mathcal{F}, \mathbb{P})$. Let U and X denote two rvs defined on it. Assume U to be uniformly distributed on $[0,1]$, and X to be an $\mathcal{X}$-valued rv with pmf $\boldsymbol{p} = (p(x),\ x \in \mathcal{X})$.

For convenience, label the elements of $\mathcal{Y}$ so that $\mathcal{Y} = \{1, 2, \ldots, |\mathcal{Y}|\}$. Starting with the channel matrix $\boldsymbol{P}$, define the cumulative probability distribution associated with the pmf $(p(y|x),\ y \in \mathcal{Y})$, namely

$$P^\star(y) = \sum_{\eta=1}^{y} p(\eta|x), \quad y = 1, \ldots, |\mathcal{Y}|$$

with the convention $P^\star(0|x) = 0$.

For each x in $\mathcal{X}$, define the $\mathcal{Y}$-valued rv $Y(x)$ given by

$$Y(x) \equiv \sum_{y=1}^{|\mathcal{Y}|} y \cdot \mathbf{1}\left[P^\star(y-1|x) < U \leq P^\star(y|x)\right] .$$

Note that

$$\begin{aligned}\mathbb{P}\left[Y(x) = y\right] &= \mathbb{P}\left[P^\star(y-1|x) < U \leq P^\star(y|x)\right] \\ &= P^\star(y|x) - P^\star(y-1|x) \\ &= p(y|x), \quad y = 1, \ldots, |\mathcal{Y}|. \end{aligned} \tag{2}$$

Furthermore, if X were taken to be *independent* of U, with $Y \equiv Y(X)$, we would conclude that

$$\mathbb{P}\left[Y = y|X = x\right] = p(y|x), \quad \begin{array}{c} x \in \mathcal{X} \\ y \in \mathcal{Y} \end{array} . \tag{3}$$

Define the mapping

$$\Phi : [0,1] \times \mathcal{X} \to \mathcal{Y}$$

by

$$\Phi(u;x) \equiv \sum_{y=1}^{|\mathcal{Y}|} y \cdot \mathbf{1}\left[P^\star(y-1|x) < u \leq P^\star(y|x)\right], \qquad \begin{array}{c} 0 \leq u \leq 1 \\ x \in \mathcal{X} \end{array}$$

Thus,

$$Y = \Phi(U;X)$$

and the earlier calculations show that $\Phi(U;X)$ is the output to the transmission of the single symbol X over a channel with channel matrix $\boldsymbol{P}$.

Now let $\{U_k,\ k = 1,2,\ldots\}$ and $\{X_k,\ k = 1,2,\ldots\}$ be rvs defined on the probability triple $(\Omega,\mathcal{F},\mathbb{P})$. Assume (i) the rvs $\{U_k,\ k = 1,2,\ldots\}$ are i.i.d. rvs which are uniformly distributed on the unit interval $[0,1]$; (ii) the rvs $\{X_k,\ k = 1,2,\ldots\}$ are $\mathcal{X}$-valued rvs; and (iii) the collections $\{U_k,\ k = 1,2,\ldots\}$ and $\{X_k,\ k = 1,2,\ldots\}$ are mutually independent.

Set

$$Y_k \equiv \Phi(U_k;X_k), \quad k = 1,\ldots,n.$$

Under the assumptions above it is a simple matter to check that

$$\mathbb{P}\left[\boldsymbol{Y}^n = \boldsymbol{y}^n | \boldsymbol{X}^n = \boldsymbol{x}^n\right] = \prod_{k=1}^{n} p(y_k|x_k), \qquad \begin{array}{c} \boldsymbol{x}^n \in \mathcal{X}^n \\ \boldsymbol{y}^n \in \mathcal{Y}^n \end{array}$$

for each $n = 1,2,\ldots$. As a result, the rvs $\{\Phi(U_k;X_k),\ k = 1,2,\ldots\}$ can be interpreted as the output of the DMC with channel matrix $\boldsymbol{P}$ in response to the input sequence $\{X_k,\ k = 1,2,\ldots\}$.

Random codes

Given the positive integers n and M, let $\mathcal{C}(M;n)$ denote the collection of all codebooks used to encode M distinct messages with strings of n symbols from $\mathcal{X}$. We can identify $\mathcal{C}(M;n)$ with $\mathcal{X}^{nM}$, so that

$$|\mathcal{C}(M;n)| = |\mathcal{X}|^{nM}.$$

A *random code* is simply a rv $\Omega \to \mathcal{X}^{nM}$. A possible way to generate such a random code is as follows: With $\boldsymbol{p}$ denoting a pmf on $\mathcal{X}$, let

$$\{X_k(w),\ k = 1,\ldots,n;\ w = 1,\ldots,M\}$$

denote a collection of i.i.d. $\mathcal{X}$-valued rvs, each distributed according to $\boldsymbol{p}$. For each $w = 1, \ldots, M$, the random codeword associated with the message w is the random (row) vector

$$\boldsymbol{X}^n(w) = (X_1(w), \ldots, X_n(w)).$$

This allows us to write the random code as the random matrix

$$\mathbb{X}^n = \begin{bmatrix} \boldsymbol{X}^n(1) \\ \vdots \\ \boldsymbol{X}^n(M) \end{bmatrix}.$$

This random matrix has M rows and n columns. Sometimes it is convenient to write the output of the DMC as

$$\boldsymbol{Y}^n = \boldsymbol{Y}^n(\boldsymbol{X}^n(W)).$$