Mathematical methods in communication July 10st, 2009

Appendix

Lecturer: Haim Permuter *Scribe: Noa Zuaretz*

I. Notation

A. General

- X - random variable
- $\mathcal{X}$ - alphabet
- $|\mathcal{X}|$ - cardinality of the alphabet. Unless it is said otherwise, we assume that the alphabet is finite, i.e., $|\mathcal{X}| < \infty$.
- x - an observation or a specific value. Clearly, $x \in \mathcal{X}$.
- $P_X(x)$ - the probability that the random variable X gets the value x, i.e., $P_X(x) = \Pr\{X = x\}$.
- P_X - denotes the whole vector of probabilities. One may also use the notation $P_X(\cdot)$.
- $P(x)$ - this is a short notation for $P_X(x)$.
- x^n - is the vector $(x_1, x_2, ..., x_n)$ for $n \geq 1$. If $n = 0$ then the vector is empty.
- x_i^j - is the vector $(x_i, x_{i+1}, ..., x_j)$, for $j > i$. If $j = i$, then the vector has only one element x_i and if $j < i$, the vector is empty.
- $\perp\!\!\!\perp$ - statistically independent symbol.
- $\mathbb{E}[X]$ - expectation, i.e.,

$$\mathbb{E}[X] \triangleq \sum_{x \in \mathcal{X}} xP(x). \tag{1}$$

  Similarly $\mathbb{E}[g(X)]$ is

$$\mathbb{E}[g(X)] \triangleq \sum_{x \in \mathcal{X}} g(x)P(x) \tag{2}$$

B. Information measures

- $H(X)$ - entropy, i.e.,

$$H(X) \triangleq \mathbb{E}[-\log_2 P_X(X)] = -\sum_{x \in \mathcal{X}} P_X(x) \log_2 P_X(x).$$

- $H(X, Y)$ - joint entropy, i.e.,

$$H(X, Y) \triangleq \mathbb{E}[-\log P(X, Y)] = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(x, y) \log P(x, y)]$$

- $H(X|Y)$ - conditional entropy, i.e.,

$$H(X|Y) \triangleq \mathbb{E}[-\log P(X|Y)] = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P(x,y)\log P(x|y)].$$

- $I(X;Y)$ - mutual information, i.e.,

$$I(X;Y) \triangleq \sum_{x,y} P(x,y)\log\frac{P(x,y)}{P(x)P(y)}. \tag{3}$$

- $D(P||Q)$ - relative entropy or KullbackLeibler divergence, i.e.,

$$D(P||Q) \triangleq \sum_{x} P(x)\log\frac{P(x)}{Q(x)} = -\mathbb{E}_P\log\frac{Q(x)}{P(x)} \tag{4}$$

C. Convexity

- $f(x)$ Convex Function - if

$$f(\lambda x_1 + \bar{\lambda}x_2) \leq \lambda f(x_1) + \bar{\lambda}f(x_2) \tag{5}$$

  for all x_1, x_2 in its domain $\forall\lambda \in [0,1]$, where $\bar{\lambda} = 1-\lambda$. A function is strictly convex if

$$f(\lambda x_1 + \bar{\lambda}x_2) < \lambda f(x_1) + \bar{\lambda}f(x_2). \tag{6}$$

- $f(x)$ Concave Function - if $-f(x)$ is (strictly) convex.

D. Source Coding

- $C(x)$ or C_x- denotes the codeword corresponding to x.
- D - set of finete length string of symbols from a D-ary alphabet.
- $l(x)$ - length of $C(x)$ associated with value $x \in \mathcal{X}$.
- $\mathbb{L} = E[l(x)]$ - expected length of the $C(x)$.
- R - rate transmitted through a channel.

E. Channel Capacity and Gaussian Channel

- C - information channel capacity.

$$C \triangleq \max_{P_X} I(X;Y) \tag{7}$$

- Gaussian Channel - discrete channel whit output Y_i at time i, where Y_i is the sum of the input X_i and the noise Z_i. The noise Z_i is drawn i.i.d from a Gaussian distribution with variance σ_z^2. Thus,

$$Y_i = X_i + Z_i, \quad Z_i \sim \mathcal{N}(0,\sigma_z^2) \tag{8}$$

  Where the noise Z_i is assumed to be independent of the signal X_i.

- P - average power constraint, limitation on the input is an energy or power constraint. For any codeword $X^n = (x_1, x_2, ..., x_n)$ transmitted over the channel, we require that

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2 \leq P \tag{9}$$

F. Differential Entropy

- $h(x)$ - differential entropy.

$$h(X) \triangleq -\int f_X(x) \log_2(f_X(x))\, dx \triangleq \mathbb{E}[-\log_2 f_X] \tag{10}$$

- $h\left(X|Y\right)$ - conditional entropy.

$$h\left(X|Y\right) \triangleq -\int f_{X,Y}\left(x,y\right) \log_2 f_{X|Y}\left(x|y\right)\, dx\, dy = \mathbb{E}\left[-\log_2 f_{X|Y}\left(X|Y\right)\right] \tag{11}$$

- $I(X;Y)$ - mutual information.

$$I(X;Y) \triangleq D\left[f_{X,Y} \| f_X f_Y\right] = h\left(X\right) - h\left(X|Y\right) \tag{12}$$

- $D\left(f_X \| g_X\right)$ - divergence.

$$D\left(f_X \| g_X\right) \triangleq \int_{x \in \mathcal{S}} f_X(x) \log_2 \frac{f_X(x)}{g_X(x)}\, dx \tag{13}$$