Lecture 20 Continuous Channel

Textbook 8.3,9.1-9.5

November 18th, 2022

Consider a random variable X with density $f(x)$. Suppose that we divide the range of X into bins of length Δ. Let us assume that the density is continuous with the bins. Then, by the mean value theorem, there exists a value x_i within each bin such that

$$f(x_i)\Delta = \int_{i\Delta}^{(i+1)\Delta} f(x)dx.$$

Consider the quantized random variable X^{Δ} , which is defined by

$$X^{\Delta} = x_i \quad \text{if } i\Delta \leq X < (i+1)\Delta.$$

Then the probability that $X^{\Delta} = x_i$ is

$$p_i = \int_{i\Delta}^{(i+1)\Delta} f(x)dx = f(x_i)\Delta.$$

The entropy of the quantized version is

$$\begin{aligned}
H(x^{\Delta}) &= -\sum_{i=-\infty}^{\infty} p_i \log p_i \\
&= -\sum_{i=-\infty}^{\infty} f(x_i)\Delta \log(f(x_i)\Delta) \\
&= -\sum \Delta f(x_i) \log f(x_i) - \sum f(x_i)\Delta \log \Delta \\
&= -\sum \Delta f(x_i) \log f(x_i) - \log \Delta
\end{aligned}$$

Theorem

If the density $f(x)$ of the random variable X is Riemann integrable, then

$$H(X^{\Delta}) + \log \Delta \to h(f) = h(X), \quad \textit{when} \Delta \to 0.$$

Example

1. *If* X *has uniform distribution on* $[0,1]$ *and we let* $\Delta = 2^{-n}$*, then* $h = 0$*,* $H(X^{\Delta}) = n$*, and* n *bits suffice to describe* X *to* n *bit accuracy.*
2. *If* X *is uniformly distributed on* $[0, \frac{1}{8}]$*, the first* 3 *bits to the right of the decimal point must be* 0*. To describe* X *to* n*-bit accuracy requires only* $n-3$ *bits, which agrees with* $h(X) = -3$*.*
3. *If* $X \sim \mathcal{N}(0, \sigma^2)$ *with* $\sigma^2 = 100$*, describing* X *to* n *bit accuracy would requires on the average* $n + \frac{1}{2}\log(2\pi e \sigma^2) = n + 5.37$ *bits.*

The Guassian channel is a time-discrete channel with 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 N. Thus,

$$Y_i = X_i + Z_i, \quad Z_i \sim \mathcal{N}(0, N).$$

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

The most common limitation on the input is an energy or power constraint. We assume an average power constraint. For any codeword $(x_1, x_2, \ldots, x_n)$ transmitted over the channel, we require that

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

Definition

The information capacity of the Gaussian channel with power constraint P is

$$C = \max_{f(x):EX^2 \leq P} I(X;Y).$$

We can calculate the information capacity as follows: Expanding $I(X;Y)$, we have

$$\begin{aligned} I(X;Y) &= h(Y) - h(Y|X) \\ &= h(Y) - h(X+Z|X) \\ &= h(Y) - h(Z|X) \\ &= h(Y) - h(Z). \end{aligned}$$

Now, $h(Z) = \frac{1}{2}\log 2\pi e N$. Also, since X and Z are independent and $EZ = 0$, we have that

$$EY^2 = E(X+Z)^2 = EX^2 + 2EXEZ + EZ^2 = P + N.$$

Given $EY^2 = P + N$, the entropy of Y is bounded by $\frac{1}{2}\log 2\pi e(P+N)$.

Applying this result to bound the mutual information, we obtain

$$
\begin{aligned}
I(X;Y) &= h(Y) - h(Z) \\
&\leq \frac{1}{2}\log 2\pi e \\
&= \frac{1}{2}\log(1+\frac{P}{N}).
\end{aligned}
$$

Hence, the information capacity of the Gaussian channel is

$$C = \max_{EX^2 \leq P} I(X;Y) = \frac{1}{2}\log(1+\frac{P}{N}),$$

and the maximum is attained when $X \sim \mathcal{N}(0,P)$.

An (M, n) code for the Gaussian channel with power constraint P consists of the following:

1. An index set $\{1, 2, 3, \cdots, M\}$.
2. An encoding function $x : \{1, 2, \cdots, M\} \to \mathcal{X}^n$ yielding codewords $x^n(1), x^n(2), \cdots, x^n(M)$, satisfying

$$\sum_{i=1}^{n} x_i^2(\omega) \leq nP, \ \omega = 1, 2, \cdots, M.$$

3. A decoding function

$$g : \mathcal{Y}^n \to \{1, 2, \cdots, M\}.$$

The rate and probability of error of the code are defined as in the discrete case. The arithmetic average of the probability of the error is defined by by $P_e^{(n)} = \frac{1}{2^{nR}} \sum \lambda_i$.

Definition

A rate R is said to be *achievable* for a Gaussian channel with a power constraint P if there exists a sequence of $(2^{nR}, n)$ codes with codewords satisfying the power constraint such that the maximal probability of error $\lambda^{(n)}$ tends to zero. The capacity of the channel is the supremum of the achievable rates.

Theorem

The capacity of a Gaussian channel with power constraint P and noise variance N is

$$C = \frac{1}{2}\log(1 + \frac{P}{N}) \quad \textit{bits per transmission.}$$

Theorem

Suppose that a function $f(t)$ is bandlimited to W, namely, the spectrum of the function is 0 for all frequencies grater than W. Then the function is completely determined by samples of the function spaces $\frac{1}{2W}$ second apart.