EE 126 Probability and Random Processes
February 20, 2018 Binary Erasure Channel I

1 Lecture

CLT

We have a sum of i.i.d. random variables $S_n = \sum_{i=1}^{n} X_i$, and we want to know the distribution of S_n. If we don't normalize, the variance will blow up. Therefore we should normalize:

$$Z_n = \frac{S_n - n\mu}{\sqrt{n\sigma^2}}$$

Z_n will then have zero mean and variance one. The CLT characterizes the distribution of Z_n, and says that

$$\lim_{n\to\infty} P(Z_n \leq x) = \Phi(x) \quad \text{for every } x$$

where $\Phi(x)$ is the CDF of the standard normal distribution.

The implications are as follows:

- The distribution of S_n (and Z_n, which is just a normalized version of S_n) will "wipe out" all the information about the X_is except for μ and σ^2. The mean and the variance are the only significant pieces of information.
- If there is a large number of independent small factors, the aggregate of these factors will be normally distributed. Therefore, this kind of setup is used to model noise.

Proof. First, an aside: if $Y \sim \mathcal{N}(0,1)$, its MGF $M_Y(s) = \mathbb{E}[e^{sY}] = e^{s^2/2}$.

We have that $X_1, ..., X_n$ are i.i.d. with mean 0 and variance 1 (without loss of generality, because we can always normalize). Let $M_X(s) = \mathbb{E}[e^{sX}]$ be the MGF of each X_i, and let Z be $(X_1 + ... + X_n)/\sqrt{n}$ (we divide by $\sqrt{n}$ in order to ensure that the variance is 1). Hence $\mathbb{E}[Z] = 0$ and $var(Z) = 1$.

Let's find the MGF of Z.

$$\begin{aligned} M_Z(s) = \mathbb{E}[e^{sZ}] &= \mathbb{E}[e^{s\frac{1}{\sqrt{n}}(X_1+...+X_n)}] \\ &= \mathbb{E}[e^{(sX_1)/\sqrt{n}}] \cdots \mathbb{E}[e^{(sX_n)/\sqrt{n}}] \\ &= \mathbb{E}[e^{(sX)/\sqrt{n}}]^n \\ &= [M_X(s/\sqrt{n})]^n \end{aligned}$$

Recall: the Taylor series of $f(x)$ is

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)(x-a)^2}{2!} + \frac{f'''(a)(x-a)^3}{3!} + ...$$

so $M_X(s) = M_X(0) + M_X'(0)s + M_X''(0)\frac{s^2}{2} + M_X'''(0)\frac{s^3}{6} + ... = 1 + \frac{s^2}{2} + \frac{\mathbb{E}[X^3]s^3}{6} + ...$

This gives us

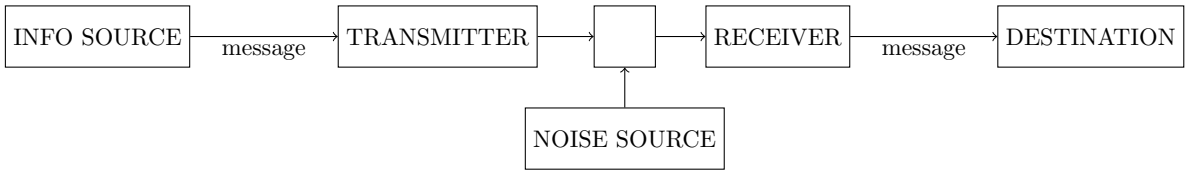
$$M_Z(s) = \left[M_X\left(\frac{s}{\sqrt{n}}\right)\right]^n = \left(1 + \frac{s^2}{2n} + \frac{\mathbb{E}[X^3]s^3}{6n^{3/2}} + ...\right)^n$$

If we send n to ∞, we see

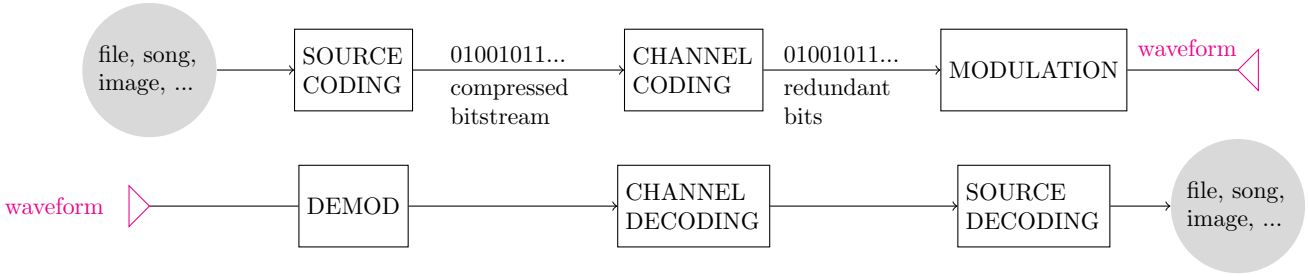
$$\lim_{n\to\infty} M_Z(s) = \lim_{n\to\infty} \left(1 + \frac{s^2}{2n} + O\left(\frac{1}{n}\right)\right)^n = e^{s^2/2}$$

Since this is the same as the MGF of the standard normal, Z must be $\sim \mathcal{N}(0,1)$. Q.E.D.!

Information Theory



Block diagram of communication system.



A more detailed diagram of a communication channel.

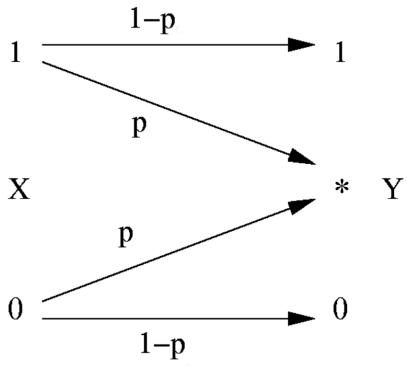
A channel has an input and an output; in the context of information theory, these both correspond to a message. (We're trying to send information over a *communication channel*.) Shannon showed that even in the presence of noise we can send things reliably over a channel. At the time, people didn't think we could send anything reliably over a channel if there was noise.

Example. We want to transmit an English novel error-free over a wireless channel. How quickly can we do this? The novel is $500,000$ characters, and the wireless channel has 10 MHz bandwidth with a SNR of 30 dB.

Shannon gives us that each character can be described on average with 2.6 bits (assuming an i.i.d. model). Our wireless channel has a capacity of 100 Mbps (we arrive at this result via a certain formula). Therefore, the novel can be sent in $(2.6 \cdot 500,000)/10^8 \approx 13$ ms.

Binary Erasure Channel

Let's study the capacity of the BEC (binary erasure channel). In a BEC, either the data comes through or it gets clobbered (with probability p). It turns out the capacity of the BEC is $C = 1 - p$ bits per channel use.



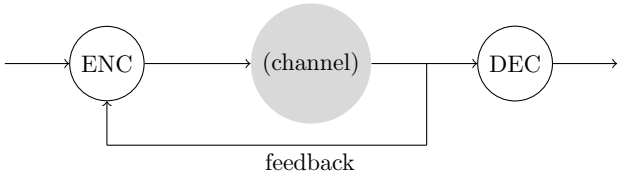
The binary erasure channel. Source: EE 126 HW2.

We call X the input alphabet and Y the output alphabet, where $X = \{0, 1\}$ and $Y = \{0, 1, *\}$. We would like an **encoder** to take a batch of L bits and encode them into a message of n bits: $f_n : X^L \mapsto X^n$ $(n > L)$. The **decoder** should be a function that maps the encoded message to the original input: $g_n : y^n \mapsto X^L$.

Call the input $X^{(n)} = (X_1, ..., X_n)$ the input bits to the channel, and $Y^{(n)} = (Y_1, ..., Y_n)$ the output symbols from the channel. We would like to determine the maximum probability of error $P_e^{(n)} = \max_{m \in X^L} P(g_n(Y^{(n)} \neq m \mid X^{(n)} = f_n(m))$, and then minimize this.

Shannon says that it's possible to send reliably over the BEC at a rate of $(1-p)$ bits per channel usage. For example, when $p = 0.5$, we can send at a rate of 0.5 bits.

The claim is that $C_{BEC} \leq (1 - p)$ bits / channel use for all possible schemes. Suppose we have feedback:



In this case, we have a genie telling us whether we've sent the correct bit, and on average we take $1/(1 - p)$ bits to send the correct one. Therefore, the rate is $1 - p$. And if we can't do better than $1 - p$ with a genie, we "sure as hell" can't do better than that without a genie (real-world situation).