Source-Channel Separation Theorem: Forward Direction

So far in this course, we have treated the 'encoding' of information from two different perspectives:

- Codes for compressing information (in order to achieve **efficient** communication). **Shannon's source-coding theorem** tells us that the minimal codeword length $\ell_{min}(P_{V^n})$ for encoding a input block from the source V^n (where all V_i are i.i.d. according to some distribution P_V) is lower bounded by $H(V^n) = nH(V)$. In other words, the rate for such a code, which is $R = \frac{\log |\mathcal{V}^n|}{n}$, is lower bounded by $H(V)$.
- Codes for protecting information (in order to achieve **low-error** communication). **Shannon's channel-coding theorem** tells us that in order to achieve an arbitrarily low error on the communication over a specific channel, it is necessary that the rate R is strictly upper bounded by the channel capacity C.

Combining these two perspectives, we can ask ourselves what happens in the 'sweet spot' where both $H(V) \leq R$ (as given by the first perspective) and $R < C$ (as given by the second perspective). Do codes with such R always exist? The answer turns out to be yes:

Theorem: Source-channel separation theorem (forward direction)

Let $V_1, V_2, \ldots, V_n$ be i.i.d. random variables (the source) distributed according to some P_V. Let $(\mathcal{X}, P_{Y|X}, \mathcal{Y})$ be a channel with capacity C. If $H(V) < C$, then there exists a **source-channel code** with error probability $P[\hat{V}^n \neq V^n] \to 0$ as $n \to \infty$.

Proof

The main idea is to construct our code in two steps: first, we optimally compress the source. Then, we apply an error-correcting code to that compression. Let $0 < \varepsilon < C - H(V)$. We will exhibit a code that has error probability at most ε. (This immediately implies the existence of such a code for $\varepsilon \geq C - H(V)$ as well.) For the first part (compression), consider the typical set $A_{\varepsilon/2}^{(n)}$ for the source P_V. The typical set has size at most $2^{n(H(V)+\frac{\varepsilon}{2})}$, and, for large enough n, contains at least $1-\frac{\varepsilon}{2}$ of the probability mass of the source. Hence, we need $n(H(V)+\frac{\varepsilon}{2})$ bits to compress the source, with a loss (error) of at most $\frac{\varepsilon}{2}$.

For the second part (error-correction), we just learned that it is possible (for large enough n) to transmit with error less than $\frac{\varepsilon}{2}$, as long as $R<C$. Since we have compressed n source symbols into $n(H(V)+\frac{\varepsilon}{2})$ bits, the rate we want to achieve is $R=H(V)+\frac{\varepsilon}{2}$. By assumption that $\varepsilon<C-H(V)$, it follows that $R<C$, such a code indeed exists. Combining the two codes, each with error at most $\frac{\varepsilon}{2}$, and applying the union bound, we get

$$\begin{aligned}
P[\hat{V}^n \neq V^n] &\leq P[V^n \notin A_{\varepsilon/2}^{(n)}] + P[\hat{V}^n \neq V^n | V^n \in A_{\varepsilon/2}^{(n)}] \\
&\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\
&= \varepsilon.
\end{aligned}$$

It is interesting to note that the i.i.d. assumption on the source can be relaxed, and the theorem actually holds for any finite-alphabet stochastic process satisfying the AEP and entropy rate $H(\{V_i\})<C$.