Statement of Shannon's Noisy Coding Theorem for the BSC

Consider a BSC with probability of correct transmission of $P > 1/2$, and hence probability of error of $Q = 1 - P < 1/2$. This channel has capacity $C = 1 - H(P) = 1 - H(Q)$.

For any desired closeness to capacity, $\epsilon > 0$, and for any desired limit on error probability, $\delta > 0$, there is a code of some length n that has rate, R, of at least $C - \epsilon$, and for which the probability of error using nearest-neighbor decoding, P_E, is less than δ.

I'll now sketch a proof of this, roughly following the sketch given by Jones & Jones in Section 5.4. Details are in Appendix C of Jones & Jones.

Strategy for Proving the Theorem

Rather than showing how to construct a specific code for any values of Q, ϵ, and δ, we will consider choosing a code of an appropriate length, n, and rate, $R = \log_2(M)/n$, *at random*, from among all subsets of F_2^n of size M.

We consider the following scenario:

1. We randomly pick a code, $\mathcal{C}$, which we give to both the sender and the receiver.
2. The sender randomly picks a codeword $\mathbf{u} \in \mathcal{C}$, and transmits it through the channel.
3. The channel randomly generates an error pattern, $\mathbf{e}$, and delivers $\mathbf{v} = \mathbf{u} + \mathbf{e}$ to the receiver.
4. The receiver decodes $\mathbf{v}$ to a codeword, $\mathbf{u}^*$, that is nearest to $\mathbf{v}$ in Hamming distance.

If the probability that this process leads to $\mathbf{u}^* \neq \mathbf{u}$ is less than δ, then there must be some specific code for which $P_E < \delta$.

How to Choose n and M

Given Q, ϵ, and δ, we need to choose the length of the codewords, n, and the number of codewords, M. How do we do this so that the proof will work?

1. We choose a value $\eta > 0$ so that $Q+\eta < 1/2$ and $1 - H(Q+\eta) \geq C - \epsilon/3$. Our aim is to almost always correct up to a fraction $Q+\eta$ of errors — slightly more than the average.
2. We choose n to be big enough that the Law of Large Numbers guarantees that the probability of getting more than $n(Q+\eta)$ errors is less than $\delta/2$.
3. We also make sure $n > -(3/\epsilon)\log_2(\delta/2)$.
4. We choose the number of codewords, M, so that the rate, $R = \log_2(M)/n$, satisfies $C - \epsilon \leq R \leq C - (2/3)\epsilon$. If necessary, we make n even bigger than needed above so that this is possible.

Rearranging the Order of Choices

It will be convenient to rearrange the order in which random choices are made, as follows:

1. We randomly pick *one* codeword, $\mathbf{u}$, which is the one the sender transmits.
2. The channel randomly generates an error pattern, $\mathbf{e}$, that is added to $\mathbf{u}$ to give the received data, $\mathbf{v}$. Let the number of transmission errors, $d(\mathbf{u}, \mathbf{v})$, be e.
3. We now randomly pick the other $M-1$ codewords. If the Hamming distance from $\mathbf{v}$ of all these codewords is greater than e, nearest-neighbor decoding will make the correct choice.

We chose η so the probability that $e > n(Q+\eta)$ is less than $\delta/2$. We need to show that **if** $e \leq n(Q+\eta)$, the probability is less than $\delta/2$ that **any** of the $M-1$ codewords chosen in step (3) has distance from $\mathbf{v}$ of e or less.

Probability of A Codeword Being Close to the Received Vector

Consider the probability that a randomly chosen codeword, $\mathbf{u}'$, will have Hamming distance from $\mathbf{v}$ of no more than $n(Q+\eta)$, when the Hamming distance from $\mathbf{v}$ to $\mathbf{u}$ is also no more than this.

This probability satisfies

$$\Pr\left(d(\mathbf{u}',\mathbf{v}) \le n(Q+\eta)\right) \;<\; \frac{1}{2^n} \sum_{i=0}^{\lfloor n(Q+\eta)\rfloor} \binom{n}{i}$$

Here, 2^n is the number of possible codewords. The sum counts the number of these at Hamming distances from 0 up to the largest integer no bigger than $n(Q+\eta)$. From each of these, we should subtract one, because we're considering a codeword *other* than the one actually transmitted. That decreases the probability, so we write $<$ rather than $=$.

Bounding this Probability

Exercise 5.7 in Jones & Jones shows that

$$\sum_{i=0}^{\lambda n} \binom{n}{i} \;\le\; 2^{nH(\lambda)}$$

where H is the binary entropy function, $H(\lambda) = -\lambda \log_2(\lambda) - (1-\lambda)\log_2(1-\lambda)$.

We can use this to bound the probability of another codeword besides $\mathbf{u}$ being too near $\mathbf{v}$:

$$\Pr\left(d(\mathbf{u}',\mathbf{v}) \le n(Q+\eta)\right) \;<\; \frac{1}{2^n} 2^{nH(Q+\eta)}$$

Now We Consider All $M-1$ Other Codewords

The probability that **any** of the $M-1$ codewords other than $\mathbf{u}$, the one actually transmitted, will be as near to $\mathbf{v}$ as $\mathbf{u}$ is no more than $M-1$ times the probability that a single codeword other than $\mathbf{u}$ will be that near.

So the probability of any other codeword being too near $\mathbf{v}$ is bounded as follows

$$\begin{aligned}
&\Pr(\text{some } \mathbf{u}' \ne \mathbf{u} \text{ is too near } \mathbf{v}) \\
&\quad< (M-1)\,\frac{1}{2^n}\,2^{nH(Q+\eta)} \\
&\quad< \frac{M}{2^n}\,2^{nH(Q+\eta)} \\
&\quad= \frac{2^{nR}}{2^n}\,2^{nH(Q+\eta)} \\
&\quad= 2^{n(R-(1-H(Q+\eta)))}
\end{aligned}$$

Here, we use the fact that $R = \log_2(M)/n$ to replace M by 2^{nR}.

Finishing the Proof

Now, recall that we chose η so that

$$1 - H(Q+\eta) \;\ge\; C - \epsilon/3$$

So our upper bound on the probability of a codeword other than the right one being too near $\mathbf{v}$ can be changed as follows:

$$\begin{aligned}
&\Pr(\text{some } \mathbf{u}' \ne \mathbf{u} \text{ is too near } \mathbf{v}) \\
&\quad< 2^{n(R-(1-H(Q+\eta)))} \\
&\quad\le 2^{n(R-(C-\epsilon/3))}
\end{aligned}$$

We also chose R so that $R \le C - (2/3)\epsilon$, which implies that $R - (C - \epsilon/3) \le -\epsilon/3$. Recalling that $n > -(3/\epsilon)\log_2(\delta/2)$, we get:

$$\begin{aligned}
\Pr(\text{some } \mathbf{u}' \ne \mathbf{u} \text{ is too near } \mathbf{v}) &< 2^{-n\epsilon/3} \\
&< 2^{\log_2(\delta/2)} \\
&= \delta/2
\end{aligned}$$

We've bounded the probabilities of the two ways an error can occur by $\delta/2$, so the overall error probability must be less than δ.