7.3.2 The deletion channel

In many applications the erasures or errors suffered by the messages when sent through a channel are random, and not adversarial. There is a beautiful theory understanding the amount of information that can be sent by different types of noisy channels, we refer the reader to [CT] and references therein for more information.

A particularly challenging channel to understand is the deletion channel. The following open problem will envolve a particular version of it. Say we have to send a binary string "10010" through a deletion channel and the first and second bits get deleted, then the message receive would be "010" and the receiver would not know which bits were deleted. This is in contrast with the erasure channel where bits are erased but the receiver knows which bits are missing (in the case above the message received would be "??010"). We refer the reader to this survey on many of the interesting questions (and results) regarding the Deletion channel [Mit09].

A particularly interesting instance of the problem is the Trace Reconstruction problem, where the same message is sent multiple times and the goal of the receiver is to find exactly the original message sent from the many observed corrupted version of it. We will be interested in the following quantity: Draw a random binary string with n bits, suppose the channel has a deletion probability of $\frac{1}{2}$ for each bit (independently), define $\mathcal{D}\left(n; \frac{1}{2}\right)$ has the number of times the receiver needs to receive the message (with independent corruptions) so that she can decode the message exactly, with high probability. It is easy to see that $\mathcal{D}\left(n; \frac{1}{2}\right) \leq 2^n$, since roughly once in every 2^n times the whole message will go through the channel unharmed. It is possible to show (see [HMPW]) that $\mathcal{D}\left(n; \frac{1}{2}\right) \leq 2^{\sqrt{n}}$ but it is not known whether this bound is tight.

Open Problem 7.4   *1. What are the asymptotics of $\mathcal{D}\left(n; \frac{1}{2}\right)$?*

2. An interesting aspect of the Deletion Channel is that different messages may have different difficulties of decoding. This motivates the following question: What are the two (distinct) binary sequences $x^{(2)}$ and $x^{(2)}$ that are more difficult to distinguish (let's say that the receiver knows that either $x^{(1)}$ or $x^{(2)}$ was sent but not which)?

References

[CT]   T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience.

[HMPW]   T. Holenstein, T. Mitzenmacher, R. Panigrahy, and U. Wieder. Trace reconstruction with constant deletion probability and related results. *In Proceedings of the Nineteenth Annual ACM-SIAM*.

[Mit09]   M. Mitzenmacher. A survey of results for deletion channels and related synchronization channels. *Probability Surveys*, 2009.

18.S096 Topics in Mathematics of Data Science
Fall 2015