EE2340/EE5847: Information Sciences/Information Theory **2020**

Handout 6: Applications

Instructor: Shashank Vatedka

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. Please email the course instructor in case of any errors.*

We now look at some applications of the inequalities that we have studied so far.

6.1 Minimum rate of a fixed-length compression scheme

The source coding theorem says that the minimum rate of any fixed-length compression scheme for a discrete memoryless source with distribution p_X is equal to $H(X)$. This statement says two things:

1. Existence of an entropy-achieving compression scheme (achievability): There exists a compression scheme such that as $n \to \infty$, the rate $R \to H(X)$, whereas the probability of error $\Pr[\hat{X}^n \neq X^n] \to 0$.
2. No compression scheme can beat entropy (converse): For every compression scheme that satisfies $\lim_{n\to\infty} \Pr[\hat{X}^n \neq X^n] = 0$, the asymptotic rate cannot be below $H(X)$.

The source coding theorem requires a proof for both parts. We will now give a proof of the converse (part 2).

Theorem 6.1. *Consider any fixed-length compression scheme for a discrete memoryless source $X^n \sim$ i.i.d.(p_X). Suppose that the scheme has deterministic encoder f, deterministic decoder g and rate R. If the probability of error $P_e = \Pr[g(f(X^n)) \neq X^n]$ satisfies $\lim_{n\to\infty} P_e = 0$, then*

$$\lim_{n\to\infty} R \geqslant H(X).$$

Proof. We first show that if the probability of error is small, then $H(\hat{X}^n) \approx H(X^n)$.

$$\begin{aligned}
H(\hat{X}^n) &= H(X^n, \hat{X}^n) - H(X^n|\hat{X}^n) && (6.1)\\
&= H(X^n) + H(\hat{X}^n|X^n) - H(X^n|\hat{X}^n) && (6.2)\\
&= H(X^n) - H(X^n|\hat{X}^n) && (6.3)\\
&\geqslant H(X^n) - H_2(P_e) - P_e \log_2 |\mathcal{X}|^n && (6.4)\\
&= nH(X) - H_2(P_e) - P_e \log_2 |\mathcal{X}|^n && (6.5)\\
&= nH(X)\left(1 - \frac{H_2(P_e)}{nH(X)} - P_e \frac{\log_2 |\mathcal{X}|}{H(X)}\right) && (6.6)
\end{aligned}$$

where (6.1) and (6.2) follow from the chain rule of entropy, (6.3) since $\hat{X}^n$ is a deterministic function of X^n and hence $H(\hat{X}^n|X^n) = 0$. Inequality (6.4) is obtained from Fano's inequality.

Let $C^{nR} = f(X^n)$ denote the codeword.

$$H(\hat{X}^n) = H(g(C^{nR}))$$

$$
\begin{aligned}
&= H(g(C^{nR}), C^{nR}) - H(C^{nR}|g(C^{nR})) \\
&\leqslant H(g(C^{nR}), C^{nR}) \\
&= H(C^{nR}) + H(g(C^{nR})|C^{nR}) && (6.7) \\
&= H(C^{nR}) && (6.8) \\
&\leqslant \sum_{i=1}^{nR} H(C_i) && (6.9) \\
&\leqslant nR && (6.10)
\end{aligned}
$$

where (6.7) and (6.9) follow from the chain rule, and 6.10 from the fact that $H(C_i) \leqslant \log_2 2 = 1$. Combining (6.6) and (6.10), we get

$$\lim_{n\to\infty} R \geqslant \lim_{n\to\infty} H(X)\left(1 - \frac{H_2(P_e)}{nH(X)} - P_e\frac{\log_2|\mathcal{X}|}{H(X)}\right) = H(X).$$

□

6.2 Maximum rate of communication over a noisy channel

For a given channel $p_{Y|X}$, let us define the capacity to be the quantity $C = \max_{p_X} I(X;Y)$.

Just as in the source coding problem, the channel coding theorem consists of two parts:

1. Existence of a capacity-achieving coding scheme (achievability): There exists a channel code such that as $n \to \infty$, the rate $R \to C$, whereas the probability decoding the message incorrectly $\Pr[\hat{M} \neq M] \to 0$.
2. No channel code can beat capacity (converse): For every compression scheme that satisfies $\lim_{n\to\infty} \Pr[\hat{M} \neq M] = 0$, the asymptotic rate cannot be greater than C.

Let us prove the converse.

Theorem 6.2. *Consider any channel code for a discrete memoryless channel $p_{Y|X}$. Suppose that the scheme has deterministic encoder f, deterministic decoder g and rate R. If the probability of error $P_e = \Pr[g(Y^n) \neq M]$ satisfies $\lim_{n\to\infty} P_e = 0$, then*

$$\lim_{n\to\infty} R \leqslant C \stackrel{def}{=} \max_{p_X} I(X;Y).$$

To prove this theorem, we will need the following lemma:

Lemma 6.3. *For any n and arbitrarily jointly distributed X^n, let Y^n be obtained by passing X^n through the DMC $p_{Y|X}$. Then,*

$$I(X^n;Y^n) \leqslant nC$$

Proof. Let us write the mutual information in terms of entropies

$$I(X^n;Y^n) = H(Y^n) - H(Y^n|X^n)$$

Using the chain rule of entropy,

$$I(X^n;Y^n) = \sum_{i=1}^{n} H(Y_i|Y_1,\ldots,Y_{i-1}) - \sum_{i=1}^{n} H(Y_i|X^n,Y_1,\ldots,Y_{i-1})$$

$$\leqslant \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X^n, Y_1, \ldots, Y_{i-1})$$

since conditioning reduces entropy. However, $H(Y_i|X^n, Y_1, \ldots, Y_{i-1}) = H(Y_i|X_i)$, since conditioned on the input to the channel X_i, the output Y_i is conditionally independent of everything else (since Y^n is obtained by passing through a DMC). Therefore,

$$I(X^n; Y^n) \leqslant \sum_{i=1}^{n} \Big(H(Y_i) - H(Y_i|X_i) \Big) = \sum_{i=1}^{n} I(X_i; Y_i).$$

For each i, $I(X_i; Y_i) \leqslant C$ (by definition of C). Hence,

$$I(X^n; Y^n) \leqslant nC$$

proving the lemma. □

6.2.1 Proof of Theorem 6.2

Recall that the message consists of $k = nR$ uniformly distributed random bits. Therefore,

$$nR = H(M) = I(M; \hat{M}) + H(M|\hat{M})$$

by definition of mutual information. Using Fano's inequality, $H(M|\hat{M}) \leqslant H_2(P_e) + P_e \log |\{0,1\}^{nR}| = H(P_e) + nRP_e$. Using this in the above,

$$nR \leqslant I(M; \hat{M}) + H(P_e) + nRP_e$$

Note that $M - X^n - Y^n - \hat{M}$ forms a Markov chain. By the data processing inequality,

$$nR \leqslant I(X^n; Y^n) + H(P_e) + nRP_e.$$

We now invoke Lemma 6.3.

$$nR \leqslant nC + H(P_e) + nRP_e$$

Dividing both sides by n and letting $n \to \infty$,

$$\lim_{n\to\infty} R \leqslant C + R \times \lim_{n\to\infty} P_e = C$$

since by assumption, $\lim_{n\to\infty} P_e = 0$. This completes the proof.

6.3 Maximizing entropy distributions

6.3.1 Gaussian maximizes differential entropy among random variables with the same variance

Fix a $\sigma > 0$. Among all probability density functions on $\mathbb{R}$ with zero mean and variance σ^2, which one maximizes differential entropy?

Answer: The Gaussian distribution $\mathcal{N}(0, \sigma^2)$.

To state the problem more precisely, let $\mathcal{F}$ be the set of all density functions f on $\mathbb{R}$ that must satisfy:

1. $f(x) \geqslant 0$ for all $x \in \mathbb{R}$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. $\int_{-\infty}^{\infty} xf(x) = 0$, and
4. $\int_{-\infty}^{\infty} x^2 f(x) = \sigma^2$.

Our goal is to compute

$$f^* = \arg\max_{f \in \mathcal{F}} \int_{-\infty}^{\infty} f(x) \log_2 \frac{1}{f(x)} dx.$$

We will show that f^* is the Gaussian. There are two approaches: One, use calculus to solve the above optimization problem. The second approach is to use information theoretic inequalities. Specifically, we will use the fact that for any two pdfs f, g, the KL divergence $D(f\|g) \geqslant 0$.

To show that the Gaussian maximizes entropy, it suffices to show that if $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/(2\sigma^2)}$, then for any $g \in \mathcal{F}$, we have $h(g) \leqslant h(f)$. Let us show this.

$$\begin{aligned} h(g) &= -\int_{-\infty}^{\infty} g(x) \log_2 g(x) dx \\ &= -\int_{-\infty}^{\infty} g(x) \log_2 \frac{g(x) f(x)}{f(x)} dx \\ &= -D(g\|f) - \int_{-\infty}^{\infty} g(x) \log_2 f(x) dx \\ &\leqslant -\int_{-\infty}^{\infty} g(x) \log_2 f(x) dx \end{aligned} \tag{6.11}$$

where the last step follows from $D(g\|f) \geqslant 0$. Substituting for f, we obtain

$$h(g) \leqslant -\int_{-\infty}^{\infty} g(x) \log_2 \frac{1}{\sqrt{2\pi\sigma^2}} - \int_{-\infty}^{\infty} g(x) \log_2 e^{-x^2/(2\sigma^2)} \tag{6.12}$$

$$= -\int_{-\infty}^{\infty} g(x) \log_2 \frac{1}{\sqrt{2\pi\sigma^2}} - \int_{-\infty}^{\infty} g(x) \frac{-x^2}{2\sigma^2} \log_2 e \tag{6.13}$$

Since both f, g are in $\mathcal{F}$, it must be the case that

$$\int_{-\infty}^{\infty} g(x) dx = \int_{-\infty}^{\infty} f(x) dx = 1,$$

and

$$\int_{-\infty}^{\infty} x^2 g(x) dx = \int_{-\infty}^{\infty} x^2 f(x) dx = \sigma^2.$$

Using this in 6.13, we get

$$\begin{aligned} h(g) &\leqslant -\int_{-\infty}^{\infty} f(x) \log_2 \frac{1}{\sqrt{2\pi\sigma^2}} - \int_{-\infty}^{\infty} f(x) \frac{-x^2}{2\sigma^2} \log_2 e \\ &= -\int_{-\infty}^{\infty} f(x) \log_2 \frac{1}{\sqrt{2\pi\sigma^2}} - \int_{-\infty}^{\infty} f(x) \log_2 e^{-x^2/(2\sigma^2)} \\ &= -\int_{-\infty}^{\infty} f(x) \log_2 f(x) dx \\ &= h(f). \end{aligned}$$

This completes the proof.