EE 231A | Solution Set 5
Information Theory | Wednesday, May 6, 2020
Instructor: Rick Wesel

94 pts
Reading: Chapters 8 & 9 of *Elements of Information Theory*

Lecture 9: Fano's Inequality and the Channel Coding Converse

1. (16 pts) *Fano's inequality without conditioning.* Let $\Pr(X = i) = p_i, i = 1, 2, \ldots, m$ and let $p_1 \geq p_2 \geq p_3 \geq \cdots \geq p_m$. The minimal probability of error predictor of X is $\hat{X} = 1$, with resulting probability of error $P_e = 1 - p_1$.

   (a) (8 pts) Choose $p_2, \ldots, p_m$ so as to maximize $H(X)$ subject to the constraint $1 - p_1 = P_e$ to find an upper bound on $H(X)$ that is a function of the constrain parameter P_e. This is Fano's inequality as expressed in (2.130) without conditioning on $\hat{X}$.
   Solution: The entropy,

$$\begin{aligned} H(X) &= -p_1 \log p_1 - \sum_{i=2}^{m} p_i \log p_i & (1) \\ &= -p_1 \log p_1 - \sum_{i=2}^{m} P_e \frac{p_i}{P_e} \log \frac{p_i}{P_e} - P_e \log P_e & (2) \\ &= H(P_e) + P_e H\left(\frac{p_2}{P_e}, \frac{p_3}{P_e}, \ldots, \frac{p_m}{P_e}\right) \text{ This is the Grouping Axiom.} & (3) \\ &\leq H(P_e) + P_e \log(m-1), & (4) \end{aligned}$$

   since the maximum of $H\left(\frac{p_2}{P_e}, \frac{p_3}{P_e}, \ldots, \frac{p_m}{P_e}\right)$ is attained by an uniform distribution. Hence any X that can be predicted with a probability of error P_e must satisfy

$$H(X) \leq H(P_e) + P_e \log(m-1), \qquad (5)$$

   which is the unconditional form of Fano's inequality.

   (b) (8 pts) Now upper bound $H(P_e)$ to provide a lower bound on P_e that corresponds to (2.132) but without the conditioning on Y.

   Solution: We can weaken this inequality to obtain an explicit lower bound for P_e,

$$P_e \geq \frac{H(X) - 1}{\log(m-1)}. \qquad (6)$$

Lecture 10: Differential Entropy

2. (12 pts) *Differential Entropy.*

   (a) Exponential distribution.

$$h(f) = -\int_0^\infty \lambda e^{-\lambda x}[\ln \lambda - \lambda x]dx \tag{7}$$
$$= -\ln \lambda + 1 \text{ nats.} \tag{8}$$
$$= \log \frac{e}{\lambda} \text{ bits.} \tag{9}$$

   (b) Laplace density.

$$h(f) = -\int_{-\infty}^\infty \frac{1}{2}\lambda e^{-\lambda|x|}[\ln \frac{1}{2} + \ln \lambda - \lambda|x|]\, dx \tag{10}$$
$$= -\ln \frac{1}{2} - \ln \lambda + 1 \tag{11}$$
$$= \ln \frac{2e}{\lambda} \text{ nats.} \tag{12}$$
$$= \log \frac{2e}{\lambda} \text{ bits.} \tag{13}$$

   (c) Sum of two normal distributions.
   The sum of two normal random variables is also normal, so applying the result derived the class for the normal distribution, since $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$,

$$h(f) = \frac{1}{2} \log 2\pi e(\sigma_1^2 + \sigma_2^2) \text{ bits.} \tag{14}$$

3. (12 pts) *Exponential Channel*
   Consider a channel in which the input X is one of two discrete values $X \in \{0, 1\}$. The output Y takes on one of two different distributions depending on the value of X. Specifically,

$$f_{Y|X}(y|x) = \begin{cases} \frac{3e^{-y}}{2} \text{ for } 0 \leq y \leq \ln 3 \text{ and } 0 \text{ otherwise} & \text{if } x = 0 \\ \frac{3e^{-y}}{2} \text{ for } \ln \frac{3}{2} \leq y \leq \infty \text{ and } 0 \text{ otherwise} & \text{if } x = 1 \end{cases} \tag{15}$$

   (a) (3 pts) Give the expression for $f_Y(y)$ and show that your expression integrates to 1. *Hints:* Your description of $f_Y(y)$ should have three distinct regions of nonzero density and $f_Y(y) = \sum_x f_{Y|X}(y|x)P_X(x)$.

   Solution:

$$f_Y(y) = \begin{cases} \frac{3}{4}e^{-y} & \text{for } 0 \leq y \leq \ln \frac{3}{2} \\ \frac{3}{2}e^{-y} & \text{for } \ln \frac{3}{2} \leq y \leq \ln 3 \\ \frac{3}{4}e^{-y} & \text{for } \ln 3 \leq y \leq \infty \\ \text{and } 0 & \text{otherwise.} \end{cases} \tag{16}$$

$$\int_0^\infty f_Y(y)dy = \int_0^{\ln\frac{3}{2}} \frac{3}{4}e^{-y}dy + \int_{\ln\frac{3}{2}}^{\ln 3} \frac{3}{2}e^{-y}dy + \int_{\ln 3}^{\infty} \frac{3}{4}e^{-y}dy \tag{17}$$

$$= \frac{1}{4} + \frac{1}{2} + \frac{1}{4} \tag{18}$$

$$= 1 \tag{19}$$

(b) (1 pt) Compute $H(X)$ for X equally likely to be 0 or 1.

Solution: $H(X) = 1$

(c) (3 pts) Compute $H(X|Y = y)$ for the three cases $0 \le y \le \ln\frac{3}{2}$, $\ln\frac{3}{2} \le y \le \ln 3$, $\ln 3 \le y \le \infty$. **Solution:** First consider $P(X = 1)$ for the three cases.

$$P(X = 1) = \begin{cases} 0 & \text{for } 0 \le y \le \ln\frac{3}{2} \\ \frac{1}{2} & \text{for } \ln\frac{3}{2} \le y \le \ln 3 \\ 1 & \text{for } \ln 3 \le y \le \infty \end{cases} \tag{20}$$

Thus we have:

$$H(X|Y = y) = \begin{cases} 0 & \text{for } 0 \le y \le \ln\frac{3}{2} \\ 1 & \text{for } \ln\frac{3}{2} \le y \le \ln 3 \\ 0 & \text{for } \ln 3 \le y \le \infty \end{cases} \tag{21}$$

(d) (3 pts) Compute $H(X|Y)$

Solution:

$$\begin{aligned}
H(X|Y) &= \int_{y=0}^{\infty} f_Y(y) H(X|Y=y) dy \\
&= \int_0^{\ln\frac{3}{2}} \frac{3}{4}e^{-y} \times 0 dy + \int_{\ln\frac{3}{2}}^{\ln 3} \frac{3}{2}e^{-y} \times 1 dy + \int_{\ln 3}^{\infty} \frac{3}{4}e^{-y} \times 0 dy \\
&= 0 + \frac{1}{2} + 0 \\
&= \frac{1}{2}
\end{aligned}$$

(e) (2 pts) For X equally likely to be 0 or 1, compute $I(X;Y)$.

Solution:

$$\begin{aligned}
I(X;Y) &= H(X) - H(X|Y) \\
&= 1 - \frac{1}{2} \\
&= \frac{1}{2}
\end{aligned}$$

(f) (5 pts) Compute $h(Y) = -\int_0^\infty f_Y(y)\ln(f_Y(y))\,dy$. *Note:* Use the natural logarithm ln instead of $\log_2$ to simplify the calculation.

Solution:

$$\begin{aligned}
h(Y) &= -\int_0^{\infty} f_Y(y)\ln\left(f_Y(y)\right) dy \\
&= -\int_0^{\ln\frac{3}{2}} \frac{3}{4}e^{-y}\ln\left(\frac{3}{4}e^{-y}\right) dy - \int_{\ln\frac{3}{2}}^{\ln 3} \frac{3}{2}e^{-y}\ln\left(\frac{3}{2}e^{-y}\right) dy - \int_{\ln 3}^{\infty} \frac{3}{4}e^{-y}\ln\left(\frac{3}{4}e^{-y}\right) dy \\
&= \int_0^{\ln\frac{3}{2}} \frac{3}{4}e^{-y}\left(\ln\frac{4}{3} + y\right) dy + \int_{\ln\frac{3}{2}}^{\ln 3} \frac{3}{2}e^{-y}\left(\ln\frac{2}{3} + y\right) dy + \int_{\ln 3}^{\infty} \frac{3}{4}e^{-y}\left(\ln\frac{4}{3} + y\right) dy \\
&= \frac{3}{4}\int_0^{\infty} e^{-y} y dy + \frac{3}{4}\int_{\ln\frac{3}{2}}^{\ln 3} e^{-y} y dy + \frac{3}{4}\ln\frac{4}{3}\int_0^{\infty} e^{-y} dy + \left(\frac{3}{2}\ln\frac{2}{3} - \frac{3}{4}\ln\frac{4}{3}\right)\int_{\ln\frac{3}{2}}^{\ln 3} e^{-y} dy \\
&= \frac{3}{4} + \frac{3}{4}\left(e^{-y}(-y-1)\right)\Big|_{\ln\frac{3}{2}}^{\ln 3} + \frac{3}{4}\ln\frac{4}{3} - \frac{3}{4}\ln 3\left(\frac{2}{3} - \frac{1}{3}\right) \\
&= \frac{3}{4} - \frac{1}{4}\left(\ln 3 + 1 - 2\ln\frac{3}{2} - 2\right) + \frac{3}{4}\ln\frac{4}{3} - \frac{1}{4}\ln 3 \quad = \\
&= \ln 2 - \frac{3}{4}\ln 3 + 1
\end{aligned}$$

(g) (5 pts) Compute $h(Y|X)$.

Solution:

$$
\begin{aligned}
h(Y|X) &= \sum_x P_X(x)h(Y|X=x) \\
&= -\frac{1}{2}\int_0^{\ln 3}\frac{3e^{-y}}{2}\ln\left(\frac{3e^{-y}}{2}\right)dy - \frac{1}{2}\int_{\ln\frac{3}{2}}^{\infty}\frac{3e^{-y}}{2}\ln\left(\frac{3e^{-y}}{2}\right)dy \\
&= -\frac{1}{2}\int_0^{\infty}\frac{3e^{-y}}{2}\ln\left(\frac{3e^{-y}}{2}\right)dy - \frac{1}{2}\int_{\ln\frac{3}{2}}^{\ln 3}\frac{3e^{-y}}{2}\ln\left(\frac{3e^{-y}}{2}\right)dy \\
&= -\frac{3}{4}\int_0^{\infty}e^{-y}\left(\ln 3 - \ln 2 - y\right)dy + -\frac{3}{4}\int_{\ln\frac{3}{2}}^{\ln 3}e^{-y}\left(\ln 3 - \ln 2 - y\right)dy \\
&= -\frac{3}{4}\left(\ln 3 - \ln 2 - 1\right) - \frac{1}{4}\left(\ln 3 - \ln 2\right) - \frac{3}{4}e^{-y}\left(y+1\right)\Big|_{\ln\frac{3}{2}}^{\ln 3} \\
&= -\frac{3}{4}\left(\ln 3 - \ln 2 - 1\right) - \frac{1}{4}\left(\ln 3 - \ln 2\right) - \frac{1}{4}\left(\ln 3 + 1 - 2\ln 3 + 2\ln 2 - 2\right) \\
&= -\left(\frac{3}{4} + \frac{1}{4} + \frac{1}{4} - \frac{1}{2}\right)\ln 3 - \left(-\frac{3}{4} - \frac{1}{4} + \frac{1}{2}\right)\ln 2 + \frac{3}{4} + \frac{1}{4} \\
&= \frac{1}{2}\ln 2 - \frac{3}{4}\ln 3 + 1
\end{aligned}
$$

(h) (2 pts) For X equally likely to be 0 or 1, compute $I(X;Y)$ using $h(Y)$ abd $h(Y|X)$

Solution:

$$
\begin{aligned}
I(X;Y) &= h(Y) - h(Y|X) \\
&= \ln 2 - \frac{3}{4}\ln 3 + 1 - \left(\frac{1}{2}\ln 2 - \frac{3}{4}\ln 3 + 1\right) \\
&= \frac{1}{2}\ln 2 \text{ nats}
\end{aligned}
$$

$$= \frac{1}{2} \text{ bits}$$

4. (10 pts) *Conditional entropy of a product.*
This is a question that explores a difference between the entropy H and the differential entropy h.

(a) (2 pts) For a discrete random variable Y, express $H(aY)$ in terms of $H(Y)$. Assume that $a \neq 0$. Give a simple argument to support your result.
When $a \neq 0$, $f(Y) = aY$ is a one-to-one and onto mapping.
Hence, $H(aY) = H(Y)$.

(b) (3 pts) Find a simplified expression for $H(XY|X)$ involving $H(Y|X)$. Show your derivation. As above, assume that $P(X=0)=0$.

$$
\begin{aligned}
H(XY|X) &= \sum_x P(X=x)H(xY|X=x) && (22)\\
&= \sum_x P(X=x)H(Y|X=x) \qquad \text{using part (a)} && (23)\\
&= H(Y|X) && (24)
\end{aligned}
$$

If you permitted $X = 0$ you get a slightly more complicated result:

$$
\begin{aligned}
H(XY|X) &= \sum_x P(X=x)H(xY|X=x) && (25)\\
&= \sum_{x\neq 0} P(X=x)H(Y|X=x) + P(X=0)\times 0 \qquad \text{using part (a)} && (26)\\
&= \sum_x P(X)H(Y|X=x) - P(X=0)H(Y|X=0) && (27)\\
&= H(Y|X) - P(X=0)H(Y|X=0) && (28)
\end{aligned}
$$

(c) (3 pts) Now consider a continuous random variable Y with pdf $f(Y)$. Find a simplified expression for $h(XY|X)$ involving $h(Y|X)$. Show your derivation. *Hint:* You may use without proof the result $h(aY) = h(Y) + \log|a|$, which we derived in lecture.

Answer: For discrete X:

$$
\begin{aligned}
h(XY|X) &= \sum_x P(X=x)h(xY|X=x) && (29)\\
&= \sum_x P(X=x)\Big(h(Y|X=x) + \log|x|\Big) && (30)\\
&= h(Y|X) + E\log|X| && (31)
\end{aligned}
$$

For continuous X:

$$
\begin{aligned}
h(XY|X) &= \int_x f(X=x)h(xY|X=x)dx && (32)\\
&= \int_x f(X=x)\Big(h(Y|X=x) + \log|x|\Big)dx && (33)\\
&= h(Y|X) + E\log|X| && (34)
\end{aligned}
$$

And finally for any distribution on X:

$$
\begin{aligned}
h(XY|X) &= E_x h(xY|X=x) && (35)\\
&= E_x\Big(h(Y|X=x) + \log|x|\Big) && (36)\\
&= h(Y|X) + E\log|X| && (37)
\end{aligned}
$$

5. (9 pts) *Data Processing and Entropy.*

(a)

$$\begin{aligned} H(X) &= H(X) + H(g(X)|X) && (38)\\ &= H(g(X), X) && (39)\\ &= H(g(X)) + H(X|g(X)) && (40)\\ &\geq H(g(X)) && (41) \end{aligned}$$

(b) Let X be uniform on $(0, 1]$ and $g(x) = 2X$.

$$\begin{aligned} h(X) &= \log 1 = 0 && (42)\\ h(g(X)) &= \log 2 > 0 && (43) \end{aligned}$$

(c)

$$\begin{aligned} h(X) &= -\int_{-\infty}^{\infty} f(x)\log f(x)dx && (44)\\ &= -\sum_{i=-\infty}^{\infty}\int_{i-1/2}^{i+1/2} f(x)\log f(x)dx && (45)\\ &= -\sum_{i=-\infty}^{\infty}\int_{-1/2}^{1/2} f(x+i)\log f(x+i)dx && (46)\\ &= -\int_{-1/2}^{1/2}\sum_{i=-\infty}^{\infty} f(x+i)\log f(x+i)dx && (47)\\ &\geq -\int_{-1/2}^{1/2}\left(\sum_{i=-\infty}^{\infty} f(x+i)\right)\log\left(\sum_{i=-\infty}^{\infty} f(x+i)\right)dx && (48)\\ &= h(g(X)) && (49) \end{aligned}$$

6. (10 pts) *More Modulo Mischief*

This problem is a continuation of problem 4

(a) (2 pts) For positive a and b show that

$$a\log a + b\log b \leq (a+b)\log(a+b). \quad (50)$$

Hint: This has nothing to do with information theory, per se. **Solution**

$$\begin{aligned} a\log a + b\log b &\leq a\log(a+b) + b\log(a+b) && (51)\\ &= (a+b)\log(a+b) && (52) \end{aligned}$$

(b) (2 pts) Use the generalization of part (a) to prove the hint of problem 4 on problem set 5 as follows: Suppose that $f(x)$ is a probability density function. Prove that

$$\sum_{i=-\infty}^{\infty} f(x+i)\log(f(x+i)) \leq \left(\sum_{i=-\infty}^{\infty} f(x+i)\right)\log\left(\sum_{i=-\infty}^{\infty} f(x+i)\right). \tag{53}$$

Solution

$$\sum_{i=-\infty}^{\infty} f(x+i)\log(f(x+i)) \leq \sum_{i=-\infty}^{\infty} f(x+i)\log\left(\sum_{i=-\infty}^{\infty} f(x+i)\right) \tag{54}$$

$$= \left(\sum_{i=-\infty}^{\infty} f(x+i)\right)\log\left(\sum_{i=-\infty}^{\infty} f(x+i)\right) \tag{55}$$

(c) (2 pts) For positive a and b show that

$$a\log a + b\log b = (a+b)\log(a+b) + a\log\left(\frac{a}{a+b}\right) + b\log\left(\frac{b}{a+b}\right). \tag{56}$$

Solution

$$a\log a + b\log b = a\log a + b\log b \tag{57}$$

$$+ a\log(a+b) - a\log(a+b) + b\log(a+b) - b\log(a+b) \tag{58}$$

$$= \big(a\log(a+b) + b\log(a+b)\big) \tag{59}$$

$$+ \big(a\log a - a\log(a+b)\big) \tag{60}$$

$$+ \big(b\log(b) - b\log(a+b)\big) \tag{61}$$

$$= (a+b)\log(a+b) + a\log\left(\frac{a}{a+b}\right) + b\log\left(\frac{b}{a+b}\right) \tag{62}$$

(d) (4 pts) Now recall the modulo operation $g(\cdot)$ of problem 4 on problem set 5, which maps the real line to the interval (-.5, .5] as follows:

$$g(x) = x + n(x), \tag{63}$$

where $n(x)$ is the unique (possibly negative) integer such that $g(x) \in (-.5, .5]$.

For a continuous random variable X with pdf $f(x)$ define two random variables $Y = g(X)$ and $Z = n(X)$, with $g(x)$ and $n(x)$ as defined above. Note that Y is continuous (with a pdf) and Z is discrete (with a pmf). Show that

$$h(X) = h(Y) + H(Z|Y) \tag{64}$$

Hint: You may use the fact that the conditional pmf for Z given Y is as follows:

$$P(Z = z|Y = y) = \frac{f(y-z)}{\sum_{i=-\infty}^{\infty} f(y-i)} \tag{65}$$

More room for proving $h(X) = h(Y) + H(Z|Y)$.

$$
\begin{aligned}
h(X) &= -\int_{-\infty}^{\infty} f_X(x)\log f_X(x)dx && (66)\\
&= -\sum_{i=-\infty}^{\infty}\int_{i-1/2}^{i+1/2} f_X(x)\log f_X(x)dx && (67)\\
&= -\sum_{i=-\infty}^{\infty}\int_{-1/2}^{1/2} f_X(y-i)\log f_X(y-i)dy && (68)\\
&= -\int_{-1/2}^{1/2}\sum_{i=-\infty}^{\infty} f_X(y-i)\log f_X(y-i)dy && (69)\\
&= -\int_{-1/2}^{1/2}\sum_{i=-\infty}^{\infty} f_X(y-i)\log f_X(y-i)dy && (70)\\
&\quad -\int_{-1/2}^{1/2}\sum_{i=-\infty}^{\infty} f_X(y-i)\log\underbrace{\left(\sum_{i=-\infty}^{\infty} f_X(y-i)\right)}_{f_Y(y)}dy && (71)\\
&\quad +\int_{-1/2}^{1/2}\sum_{i=-\infty}^{\infty} f_X(y-i)\log\left(\sum_{i=-\infty}^{\infty} f_X(y-i)\right)dy && (72)\\
&= \underbrace{-\int_{-1/2}^{1/2}\left(\sum_{i=-\infty}^{\infty} f_X(y-i)\right)\log\left(\sum_{i=-\infty}^{\infty} f_X(y-i)\right)dy}_{h(Y)} && (73)\\
&\quad +\int_{-1/2}^{1/2}\left(-\sum_{i=-\infty}^{\infty} f_X(y-i)\log\left(\frac{f_X(y-i)}{\sum_{i=-\infty}^{\infty} f_X(y-i)}\right)\right)dy && (74)\\
&= h(Y) + \int_{-1/2}^{1/2} f_Y(y)\left(-\sum_{i=-\infty}^{\infty}\left(\frac{f_X(y-i)}{f_Y(y)}\right)\log\left(\frac{f_X(y-i)}{f_Y(y)}\right)\right)dy && (75)\\
&= h(Y) + \int_{-1/2}^{1/2} f_Y(y)H(Z|Y=y)dy && (76)\\
&= h(Y) + H(Z|Y) && (77)
\end{aligned}
$$

7. (13 pts) *Mutual information for a mixed distribution.(Cong Shen's distribution)* Consider the following channel:

- The input X is a binary random variable $X \in \{0, 1\}$. For all parts of this problem, assume that X is equally likely to be 0 or 1.
- The output Y is a neither completely discrete or completely continuous as described below.
- When the input X equals 0, the output Y is also 0 with probability 1.

- When the input X equals 1 the output Y is uniformly distributed on the closed interval $\left[\frac{1}{2}, \frac{3}{2}\right]$

(a) (1 pt) Find $H(X)$.
Solution: $H(X) = 1$.

(b) (2 pts) Find $H(X|Y)$.
Solution: $H(X|Y) = 0$.

(c) (6 pts) Ultimately, find the differential entropy $h(Y|X)$. Along the way, you will compute two differential entropies with specific conditioning.

i. (2 pts) $h(Y|X = 0)$.
Solution: $h(Y|X = 0) = -\infty$ since Y is deterministically zero in this case.

ii. (2 pts) $h(Y|X = 1)$.
Solution: $h(Y|X = 1) = 0$ since Y is a unit-width uniform PDF in this case.

iii. (2 pts) $h(Y|X)$
Solution:

$$
\begin{aligned}
h(Y|X) &= P(X = 0)h(Y|X = 0) + P(X = 1)h(Y|X = 1) && (78)\\
&= \frac{1}{2} \times -\infty + \frac{1}{2} \times 0 && (79)\\
&= -\infty && (80)
\end{aligned}
$$

(d) (2 pts) Find $h(Y)$.
Solution: $h(Y) = -\infty$ since it has a mass point at zero.

(e) (2 pts) Find $I(X; Y)$.
Solution:

$$
\begin{aligned}
I(X; Y) &= H(X) - H(X|Y) && (81)\\
&= 1 - 0 && (82)\\
&= 1 && (83)
\end{aligned}
$$

Note that in this case it is not useful to attempt $I(X; Y) = h(Y) - h(Y|X)$ since both terms on the right are $-\infty$.