Miana Smith
PIT 2022

PSET 3

4.1

$$H(p) = -\sum_{i=1}^{X} p_i \log(p_i)$$

Continuity

The sum of a continuous function is also continuous. We know that $f(p) = p\log(p)$ is a continuous function, so $H(p)$ is continuous.

If this does not feel one-hundo sufficient, we can quickly show that $f(p)$ is continuous with a little epsilon-delta. Recall: $f(x)$ is continuous if given an $\epsilon > 0$, there exists $\delta > 0$ s.t. if $|x - c| < \delta$, then $|f(x) - f(c)| < \epsilon$.

We will do this for $f(p) = \log p$ because p is continuous and the product of two continuous functions is also continuous. We have:

$$|\log(p) - \log(c)| < \epsilon$$
$$\log(c) - \epsilon < \log(p) < \log(c) + \epsilon$$

Raise everything by e:

$$ce^{-\epsilon} < p < ce^{\epsilon}$$

We subtract c from everything:

$$ce^{-\epsilon} - c < p - c < ce^{\epsilon} - c$$

This gives us bounds for δ.

Non-negativity

We know that $0 \le p_i \le 1$, which means that $\log(p_i) \le 0$. So, $-p_i\log(p_i)$ will always be positive.

Boundedness

Or, $H(p) \le C(X)$. $H(p) = C(X)$ when all values are equally likely.

First, we show that $H(p)$ is monotonic, specifically that it does not decrease.

We can do this inductively on by summing over intervals from $i = 1$ to $i = k$, where $k < X$ and comparing against the $k + 1$ sum, which must always be greater than or equal to the prior sum for $H(p)$ to be not decreasing. Good writing!

Let's consider the $k = 1$ case first:

$$-p_1 \log(p_1) \geq -p_1 \log(p_1) - p_2 \log(p_2)$$

This is true based on our non-negativity constraint. A non-negative value is always greater than or equal to itself minus a non-negative value.

Let's then consider the $k + 1$ case assuming case k to be true.
Then we want to show that $-\sum_{i=1}^{k+2} p_i \log p_i \geq -\sum_{i=1}^{k+1} p_i \log p_i$

$$-\sum_{i=1}^{k+2} p_i \log p_i = -\left(\sum_{i=1}^{k} p_i \log p_i\right) - p_{k+1} \log(p_{k+1}) - p_{k+2} \log p_{k+2}$$

$$= -\left(\sum_{i=1}^{k+1} p_i \log p_i\right) - p_{k+2} \log p_{k+2}$$

So it follows that the $k + 1$ case is greater than or equal to the prior step on the basis of non-negativity and the assumption of truth for case k. The gives us monotonicity, or that the equation does not decrease. Next, we will find the maximum of this, which will give us C.

We know that $\sum_{i=1}^{X} p_i = 1$, since this is normalized. The maximum $H(p)$ occurs when all the values are equally likely, i.e. when the distribution is flat: $p_i = 1/X$. So,

$$H_C(p) = -\sum_{i=1}^{X} \frac{1}{X} \log\left(\frac{1}{X}\right) = -\log\frac{1}{X}$$

This gives us our maximum value. I actually think that based on how the problem is stated this alone (without the potentially problematic inductive proof) is sufficient.

Another way for doing this is using Lagrange multipliers (see http://underactuated.csail.mit.edu/optimization.html#section2), which are a way of solving optimization problems subject to equality constraints.

The problem is given as (as per the link above):

$$Minimize_z l(\boldsymbol{z}) \ subject \ to \ \phi(\boldsymbol{z}) = 0$$

To do this, we define the Lagrangian function:

$$L(\boldsymbol{z}, \lambda) = l(\boldsymbol{z}) + \lambda^T \phi(\boldsymbol{z})$$

Where λ is a vector of the Lagrange multipliers.
For $\boldsymbol{z}^*$ to be optimal, $\frac{\partial L}{\partial \boldsymbol{z}} = 0$ and $\frac{\partial L}{\partial \lambda} = 0$.

We can now apply this to the problem at hand: Frankly, it is a big pain to type the summations, so appreciate this bonus.

$$L = -\sum_{i=1}^{X} p_i \log(p_i) - \lambda\left(\sum_{i=1}^{X} p_i - 1\right)$$

The partials are then:

$$\frac{\partial L}{\partial p_i} = 0 = -\log(p_i) - 1 - \lambda$$

And

$$\frac{\partial L}{\partial \lambda} = 0$$

(because of normalization).

So we have:

$$\log(p_i) = -1 - \lambda$$

Or:

$$p_i = e^{-1-\lambda}$$

Which via normalization means that $p_i = \frac{1}{X}$ as we previously reasoned, and $H_C(p) = -\log\frac{1}{X}$

Additive independence

$$H(p,q) = -\sum_{i=1}^{X}\sum_{j=1}^{Y} p_i q_j \log(p_i q_j)$$

$$= -\sum_{i=1}^{X}\sum_{j=1}^{Y} p_i q_j \log(p_i) - \sum_{i=1}^{X}\sum_{j=1}^{Y} p_i q_j \log(q_j)$$

Recall that the sum over p or q is normalized.

$$= -\sum_{i=1}^{X} p_i \log(p_i) - \sum_{j=1}^{Y} q_j \log(q_j)$$

$$= H(p) + H(q)$$

4.2

We are given $I(x,y) = H(x) + H(y) - H(x,y)$ and we want to demonstrate the relationship between that and:

$$\sum_{x,y} p(x,y) \log\frac{p(x,y)}{p(x)p(y)}$$

To do this, we shall do everyone’s favorite thing and split up the logarithm according to those logarithm rules.

$$\sum_{x,y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}$$
$$= \sum_{x,y} p(x,y)\log p(x,y) - \sum_{x,y} p(x,y)\log p(x) - \sum_{x,y} p(x,y)\log p(y)$$

The last two terms have the logs split out in their variable dependence. Because we have normalized probability, this means we can sum over either x or y in those terms to further simplify:

$$= \sum_{x,y} p(x,y)\log p(x,y) - \sum_{x} p(x)\log p(x) - \sum_{y} p(y)\log p(y)$$

These summations now look like the definition of entropy!

$$H(p) = -\sum_{i=1}^{X} p_i \log(p_i)$$

So:

$$= -H(x,y) + H(x) + H(y) = I(x,y)$$

4.3

Small probability ϵ of making a bit error.

a.

For an error to be made we need at least 2/3 to fail. The probability of 2/3 occurring per instance is $\epsilon^2(1-\epsilon)$ and there are 3 ways for this to happen, i.e. 110, 011, 101. The probability of 3/3 occurring is ϵ^3 and there is only 1 way for this to happen: 111.

So the overall probability is $3\epsilon^2(1-\epsilon)+\epsilon^3$. We are given that ϵ is quite small; this is likely a hint that it is time to drop some terms, so we're down to: $3\epsilon^2$.

b.

Now we do this three times and majority vote on the outcome. This is the same as above but our new probability is $3\epsilon^2$. Again, ignoring the cubes, we get:

$$3 * (3\epsilon^2)^2 = 27\epsilon^4$$

c.

We're sticking with $3\epsilon^2$ for $N = 1$.

Looking at this almost pattern, it looks like this will go something like:

$$3^N\epsilon^{2(N)}$$

Or $\propto e^{2N}$ for large N.

4.4

$$H = -\int_{-\infty}^{\infty} p(x)\log\big(p(x)\big)\,dx$$

Let’s use the natural logarithm because life does not need to be painful.

$$H = -\int_{-\infty}^{\infty} p(x)\ln\big(\mathrm{p(x)}\big)\,dx$$

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{(-x-x_0)^2/2\sigma^2}$$

$$H = -\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} e^{(-x-x_0)^2/2\sigma^2}\ln\left(e^{(-x-x_0)^2/2\sigma^2}\right)dx$$

$$= -\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} e^{\frac{(-x-x_0)^2}{2\sigma^2}}\left(\frac{-x-x_0}{2\sigma^2} - 2\ln(2\pi\sigma^2)\right)dx$$

From the first term, we see that this integral actually evaluates to the $\langle (x-x_0)^2\rangle$ since it defines the variance. And for the second term, the natural log will pull out as it has no x dependence. This leaves only the awkward exponential to evaluate. With some help from MATLAB, this all comes to:

$$= \frac{1}{2}\ln 2\pi\sigma^2 + \frac{1}{2}$$

4.5

Bandwidth: 3300Hz
SNR =20dB

a.

Capacity is given by:

$$C = \Delta f \log_2\left(1 + \frac{S}{N_0 \Delta f}\right)$$

Our latter term is given by:

$$20dB = 10\log_{10}\frac{S}{N} \to \frac{S}{N} = 10^2$$

$$C = 22 * 10^3\ bit/s$$

b.

$$C = 10^9$$

$$10^9 = 3300\log_2\left(1 + \frac{S}{N}\right)$$

$$2^{\frac{10^9}{3300}} - 1 = S/N$$

$$\approx 2^{10^6} \approx 10^{10^5}$$

$$\to 10^6 dB$$

4.6

We want to show $f(x_1, \dots x_n) = n^{-1} \sum_{i=1}^{n} x_i$ is an estimator for the mean that is unbiased and achieves the Cramér-Rao bound.

$(x_1, \dots x_n)$ is drawn from Gaussian distribution with variance σ^2 and unknown mean value x_0

Unbiased

We want $\langle f(x_1, \dots x_n) \rangle = x_0$

$$\langle f(x) \rangle = \int f(x) p(x) dx$$

Our probability in this case is given by:

$$p(x_1, \dots x_N) = \prod_{n=1}^{N} p(x_n)$$

$$\langle f(x) \rangle = \int \left(n^{-1} \sum_{i=1}^{n} x_i \right) \left(\prod_{n=1}^{N} p(x_n) \right) dx_1 \dots dx_N$$

Oh no!

Anyways, I'm already not writing this down rigorously, but: each x_i is its own thing, so it is equivalent to think of this integral as the sum of many integrals over each individual x_i, p_i pairing:

$$\langle f(x) \rangle = \sum_{i=1}^{n} \frac{1}{n} \int x_i p_i dx_i$$

This integral term is now the expected value for x_i.

$$\langle f(x) \rangle = \sum_{i=1}^{n} \frac{1}{n} \langle x_i \rangle = x_0$$

This is as we desire.

Cramér-Rao

(From the text), this states that $\sigma^2(f) \geq \frac{1}{J(\alpha)}$ So we need to find the Fisher information and the variance of the estimator.

I am tired now so I do not think I will do this today! <3