18.600 Recitation 12

Recitation Instructor: Vishesh Jain

Partial solutions available at math.mit.edu/∼visheshj

Thursday, Dec. 6th, 2018

Solution Problem 1. **(a)** Since X and Y are independent, $H(X,Y) = H(X)+H(Y) = 2\log_2(6)$.

(b) $H(Z) = (1/36)\log_2(36)+(2/36)\log_2(36/2)+\cdots+(6/36)\log_2(36/6)+(5/36)\log_2(36/5)+\cdots+(1/36)\log_2(36)$.

(c) $H_X(Z) = \log_2(6)$.

(d) $H(X,Z) = H(X) + H_X(Z) = 2\log_2(6)$. We could have deduced this directly from part (a) by noting that knowing (X,Z) is equivalent to knowing (X,Y).

Solution Problem 2. **(a)** By Shannon's Noiseless Coding Theorem, Bob can encode his message using at least $H(X)$ and at most $H(X)+1$ bits in expectation. In our case, $H(X) = (1/16)\log_2(16)+(4/16)\log_2(16/4)+(6/16)\log_2(16/6)+(4/16)\log_2(16/4)+(1/16)\log_2(16) \approx 2.03$.

(b) We can use a Huffman code. For instance, $0 \mapsto 0000, 1 \mapsto 001, 2 \mapsto 1, 3 \mapsto 01, 4 \mapsto 0001$.

(c) Let $Y = (X_1,\ldots,X_{100})$. Then, $H(Y) = 100H(X)$. By Shannon's Noiseless Coding Theorem, Bob can encode his message using at least $H(Y)$ and at most $H(Y)+1$ bits in expectation. Therefore, the expected number of bits *per symbol* is between $H(X)$ and $H(X)+(1/100)$.

Solution Problem 3. **(a)** Note that

$$\mathbf{KL}(p||q) = \sum_{i=1}^{n} p_i \cdot (-\log_2(q_i/p_i)) \geq -\log_2(\sum_{i=1}^{n} p_i \cdot (q_i/p_i)) = -\log_2(1) = 0.$$

Here, we have used the convexity of the function $x \mapsto -\log(x)$. Equality holds if and only if equality holds in Jensen's inequality, which holds if and only if q_i/p_i does not depend on i, which is true if and only if $p = q$.

(b) We have

$$\mathbf{KL}(p||\text{Unif}) = \sum_{i=1}^{n} p_i \log_2(np_i) = -H(p) + \log_2(n).$$

Hence, by the previous part, we have $\log_2(n) \geq H(p)$, and equality holds if and only if $p = \text{Unif}$.

Solution Problem 4. Following the proof of the Chernoff bound, we have that for any $\lambda > 0$,

$$\Pr[X \geq (p+\delta)n] = \Pr[\exp(\lambda X) \geq \exp(\lambda(p+\delta)n)] \leq \exp(-\lambda(p+\delta)n)\,\mathrm{E}[\exp(\lambda X)].$$

Moreover, we know that

$$\mathrm{E}[\exp(\lambda X)] = \left(pe^{\lambda} + (1-p)\right)^n.$$

Minimizing the quantity

$$\left(\frac{pe^{\lambda} + (1-p)}{e^{\lambda(p+\delta)}}\right)^{n}$$

over $\lambda > 0$ gives us

$$e^{\lambda} = \frac{(1-p)(p+\delta)}{p(1-p-\delta)},$$

and substituting this into the previous expression shows that the minimum is

$$\left(\left(\frac{p}{p+\delta}\right)^{p+\delta}\left(\frac{1-p}{1-p-\delta}\right)^{1-p-\delta}\right)^{n}.$$

Finally, note that

$$\left(\frac{p}{p+\delta}\right)^{p+\delta}\left(\frac{1-p}{1-p-\delta}\right)^{1-p-\delta} = \exp(-\mathbf{KL}(\mathrm{Ber}(p+\delta)||\mathrm{Ber}(p))).$$