Handout 36 Information Theory and Coding
Final exam solutions Jan. 23, 2021

PROBLEM 1.

(a) $I(U_1; Y_1, Y_2) + I(U_2; Y_1, Y_2, U_1) = I(U_1; Y_1, Y_2) + I(U_2; Y_1, Y_2|U_1) = I(U_1, U_2; Y_1, Y_2)$
$= I(X_1 \oplus X_2, X_2; Y_1, Y_2) = I(X_1, X_2; Y_1, Y_2) = I(X_1; Y_1) + I(X_2; Y_2) = I(W_1) + I(W_2)$.

(b) $I(W^+) = I(X_2; Y_1, Y_2, U_1) \geq I(X_2, Y_2) = I(W_2)$. From part (a) we know $I(W^-) + I(W^+) = I(W_1) + I(W_2)$. Hence it must be true that $I(W^-) \leq I(W_1)$.

(c) Observe that if we exchange W_1 and W_2, Y_1 and Y_2 will be swapped. Hence, $I(W^-) = I(U_1; Y_1, Y_2)$ remains unchanged. From (b) we know $I(W^-) \leq I(W_1)$ and by exchanging W_1 and W_2, we know $I(W^-) \leq I(W_2)$. Therefore $I(W^-) \leq \min\{I(W_1), I(W_2)\}$ and $I(W^+) \geq \max\{I(W_1), I(W_2)\}$ follows from (a).

(d) $I(W^-) = I(U_1; Y_1Y_2) = H(Y_1Y_2) - H(Y_1Y_2|U_1)$. $H(Y_1Y_2) = H(Y_1) + H(Y_2) = h_2(\epsilon) + (1-\epsilon) + 1$. $H(Y_1Y_2|U_1 = 0) = H(Y_1Y_2|U_1 = 1) = 1 + h_2(\epsilon) + (1-\epsilon)h_2(p)$. Then, $I(W^-) = (1-\epsilon)(1-h_2(p))$. $I(W^+) = I(W_1) + I(W_2) - I(W^-) = 1 - \epsilon h_2(p)$.

PROBLEM 2.

(a) $I(XY; U) \geq I(X; U) \geq I(X; Y)$ since $X - U - Y$ is a Markov Chain. Since this is true for all $p_{U|XY} : X - U - Y$, $K(X; Y) \geq I(X; Y)$.

(b) Let $V = f(U)$ with $f(u) = u$ except $f(u_1) = u_2$. Since $p_{X|U}(.|u_1) = p_{X|U}(.|u_2)$, we have the Markov Chain $X - V - Y$. But since V is a function of U, we have $I(XY; V) \leq I(XY; U)$. Also $|\mathcal{V}| < |\mathcal{U}|$.

(c) Suppose U is a minimizer and there exists $u_1 \neq u_2$ such that $p_{X|U}(.|u_1) = p_{X|U}(.|u_2)$ and $p_{Y|U}(.|u_1) = p_{Y|U}(.|u_2)$. Construct V as in (b) and observe $I(XY; U) = I(XY; V)$. Repeatedly apply (b) until whenever $u_1 \neq u_2$ either $p_{X|U}(.|u_1) \neq p_{X|U}(.|u_2)$ or $p_{Y|U}(.|u_1) \neq p_{Y|U}(.|u_2)$.

(d) First, observe that for any u, either $p_{X|U}(1|u) = 0$ or $p_{Y|U}(0|u) = 0$. If there exists $u_1 \neq u_2$ such that $p_{X|U}(1|u_1) = p_{X|U}(1|u_2) = 0$, by using (b) we can merge u_1 and u_2 to decrease $I(XY; U)$. Hence there must exist at most one u such that $p_{X|U}(1|u) = 0$. With a similar argument, we argue that there must exist at most one u such that $p_{Y|U}(0|u) = 0$. Hence, we can choose $|U|$ at most 2.

(e) Let $p := \Pr(U = 1)$ and $q := 1 - p$. With the choice of U in part (d), we have $H(X|U) = ph_2(\frac{1}{3p})$ and $H(Y|U) = qh_2(\frac{1}{3q})$. Minimizing $I(XY; U)$ is equivalent to maximizing $H(X|U) + H(Y|U) = ph_2(\frac{1}{3p}) + qh_2(\frac{1}{3q}) \leq h_2(\frac{2}{3})$. The inequality follows by concavity of $h_2(.)$ and is attained when $p = q = 1/2$. Hence, $K(X; Y) = H(XY) - h_2(\frac{2}{3}) = 2/3$.

PROBLEM 3.

(a) Since B_n is a lower-triangular matrix with positive diagonal entries, its inverse B_n^{-1} exists and is lower-triangular. Consider the transform $Z^n = B_n^{-1}(X^n - \mu^n)$, where $\mu^n := \big[E[X_1], \ldots, E[X_n]\big]^T$ and observe $E[Z_i] = 0$ for all $1 \leq i \leq n$ and the covariance matrix of Z^n is $B_n^{-1} K_n (B_n^{-1})^T = I_n$. Finally, since B_n^{-1} is lower-triangular, we can relate $a_{ij} = b_{ij}$, $j \leq i$ and $m_j = E[X_j]$.

(b)

$$-\log f_n(X^n) = \frac{1}{2}\log((2\pi)^n |K_n|) + \frac{\log(e)}{2}(X^n - \mu^n)^T K_n^{-1}(X^n - \mu^n)$$

$$= \frac{1}{2}\log((2\pi)^n |K_n|) + \frac{\log(e)}{2}(Z^n)^T Z^n = \frac{1}{2}\log((2\pi)^n |K_n|) + \frac{1}{2}\sum_{i=1}^{n} Z_i^2$$

$$h(X^n) = E[-\log f_n(X^n)] = \frac{1}{2}\log((2\pi)^n |K_n|) + \frac{\log(e)}{2}\sum_{i=1}^{n} E[Z_i^2]$$

Hence,

$$\frac{1}{n}\big[-\log f_n(X^n) - h(X^n)\big] = \frac{\log(e)}{2n}\sum_{i=1}^{n}(Z_i^2 - E[Z_i^2]) = \frac{\log(e)}{2n}\sum_{i=1}^{n}(Z_i^2 - 1)$$

(c) From Strong Law of Large Numbers, we know that $\frac{1}{n}\sum_{i=1}^{n} Z_i^2 \to 1$ with probability 1. Thus, $\frac{1}{n}\sum_{i=1}^{n}(Z_i^2 - 1) \to 0$ with probability 1.

(d) No. $X_1 = Z_1 \sim N(0,1)$ and $X_2 = Z_1 + 2Z_2 \sim N(0,5)$.

(e) From part (b), we know $\frac{1}{n}h(X^n) = \frac{1}{2n}\log((2\pi)^n|K_n|) + \frac{\log(e)}{2} = \frac{1}{2}\log(2\pi) + \frac{1}{2n}\log(|K_n|) + \frac{\log(e)}{2}$. Therefore, we only need to check if $\lim_n \frac{1}{2n}\log(|K_n|)$ exists. Observe that $|K_n| = |B_n||B_n^T| = (n!)^2$, hence $\lim_n \frac{1}{2n}\log(|K_n|) = \lim_n \frac{1}{n}\sum_{i=1}^{n}\log n = \infty$, which implies $\frac{1}{n}h(X^n) \to \infty$.

(f) Yes. Observe that X_i's are Gaussian and K_n is uniquely factorized as $K_n = B_n B_n^T$ where B_n is a lower triangular matrix with positive diagonal entries and with its ijth entry being j if $j \leq i$ and 0 otherwise. Thus parts a,b,c can be repeated for this case.

Problem 4.

(a) $\frac{1}{n}H(U^n|\hat{U}^n) \leq \frac{1}{n}\sum_i H(U_i|\hat{U}^n) \leq \frac{1}{n}\sum_i H(U_i|\hat{U}_i) \overset{(1)}{\leq} \frac{1}{n}\sum_i h_2(P(U_i \neq \hat{U}_i)) \overset{(2)}{\leq} h_2(q_n)$ where (1) follows from Fano's inequality and (2) follows from convexity of $h_2(.)$.

(b) $I(U^n;\hat{U}^n) \overset{(1)}{\leq} I(U^n; W_n, V^n) = I(U^n;V^n) + I(U^n;W_n|V^n) \overset{(2)}{\leq} I(U^n;V^n) + H(W_n) = n(1-p) + H(W_n)$ where (1) follows from Data Processing inequality and (2) follows from the fact that $I(X;Y) \leq H(X)$.

(c) From (b) we have $= 1 - \frac{1}{n}H(U^n|\hat{U}^n) = \frac{1}{n}I(U^n;\hat{U}^n) \leq (1-p) + \frac{1}{n}H(W_n) \leq (1-p) + \frac{1}{n}\log|\mathcal{W}_n|$. Hence, $\frac{1}{n}\log|\mathcal{W}_n| \geq p - \frac{1}{n}H(U^n|\hat{U}^n) \leq p - h_2(q_n)$.

(d) Given $V^n = v^n$, define the set $C(v^n) = \{u^n : u_i = v_i$ whenever v_i is unerased$\}$. Observe that any $u^n \in C(v^n)$ is equally likely. Hence, let Bob choose one of them. Note that any other decision rule will have an error probability at least as this method's.

(e) Suppose $W_n(u^n)$ are chosen uniformly at random. Then,

$$\Pr(\hat{U}^n \neq U^n | K = k) \leq \Pr(\exists u^n \in C(V^n) : W_n(u^n) = W_n(U^n) | K = k)$$

$$\leq \frac{E[|C(V^n)||K = k]}{2^{nR}} = 2^{k-nR}$$

since $|C(V^n)| = 2^k$ given $K = k$. Pick $r \in (p, R)$ and write

$$\Pr(\hat{U}^n \neq U^n) = \Pr(\hat{U}^n \neq U^n | K > nr) \Pr(K > nr) + \Pr(\hat{U}^n \neq U^n | K \leq nr) \Pr(K \leq nr)$$

$$\leq \Pr(K > nr) + \Pr(\hat{U}^n \neq U^n | K \leq nr) \Pr(K \leq nr)$$

$$\leq \Pr(K > nr) + \Pr(\hat{U}^n \neq U^n | K \leq nr).$$

Since $K = \sum_{i=1}^{n} E_i$ where E_i are erasures that occur with probability p, we know that $\frac{K}{n} \to p$ with probability 1, hence $\Pr(K > nr) \to 0$. Also note that $\Pr(\hat{U}^n \neq U^n | K \leq nr) \leq 2^{n(r-R)}$, hence goes to 0 as well.

This concludes that the average error probability over the ensemble of labelings is small, hence there exists a labeling such that the error probability is small.