This is an 80 minute exam. You may have an additional 100 minutes to answer the following questions in the notebooks provided. The exam is closed book. Make sure that you have included your name, your personal random 4 digit code number and signature in each book used (*5 points*). Read each question carefully. All statements must be justified. Computations should be simplified as much as possible.

1. *25 points* Each day, it rains (event $R = 1$) or not (event $R = 0$). A TV station subscribes to a weather forecasting service which delivers a prediction: $Q = 1$ if the prediction is rain, or $Q = 0$ if no rain. Each day, the TV weatherman makes the weather announcement $A = Q$. Fortunately, Q and R are not independent and have the following PMF

$P_{R,Q}(r,q)$	$q = 1$	$q = 0$
$r = 1$	1/8	1/16
$r = 0$	3/16	10/16

   (a) *10 points* A student observes that the weatherman is correct with probability 12/16 but could be correct with probability 13/16 by always making the weather announcement $A = 0$, corresponding to "no rain." The student applies for the weatherman's job, but the boss, who is an information theorist, turns him down. Why?

   When the weatherman always makes the announcement $A = 0$, the mutual information $I(R; A) = 0$; i.e., the weatherman communicates no information about the rain. When the weatherman makes the prediction $A = Q$, the mutual information communicated to the TV audience about the weather is

$$
\begin{aligned}
I(R; A) = I(R; Q) &= E\left[\log \frac{P_{R,Q}(r,q)}{P_R(r)\, P_Q(q)}\right] \\
&= \frac{1}{8}\log\frac{1/8}{(3/16)(5/16)} + \frac{1}{16}\log\frac{1/16}{(3/16)(11/16)} \\
&\quad + \frac{3}{16}\log\frac{3/16}{(13/16)(5/16)} + \frac{10}{16}\log\frac{10/16}{(13/16)(11/16)} \\
&= \frac{1}{8}\log\frac{32}{15} + \frac{1}{16}\log\frac{16}{33} + \frac{3}{16}\log\frac{48}{65} + \frac{10}{16}\log\frac{160}{143} \\
&= 0.0906 \textit{ bits}
\end{aligned}
$$

   The boss believes that communicating some information, even 0.09 bits, will attract a bigger audience than communicating no information.

   (b) *10 points* The prediction Q is based on a maximum likelihood (ML) hypothesis test (using some unspecified observations X) as to whether $R = 0$ or $R = 1$. For what values of $p = P(R = 1)$ does the weatherman's announcement $A = Q$ also maximize the probability $P(C)$ of a correct prediction based on Q?

   This question really has not much to do with information theory; it's mostly a probability question. From the joint PMF of R and Q, we can conclude that R has marginal PMF

$$
P_R(r) = \begin{cases} 13/16 & r = 0, \\ 3/16 & r = 1. \end{cases}
$$

   The conditional PMF of Q given R is

$$
P_{Q|R}(q|r) = \frac{P_{R,Q}(r,q)}{P_R(r)}.
$$

   In particular,

$$
P_{Q|R}(q|0) = \begin{cases} 10/13 & q = 0, \\ 3/13 & q = 1, \end{cases} \qquad P_{Q|R}(q|1) = \begin{cases} 1/3 & q = 0, \\ 2/3 & q = 1. \end{cases}
$$

We can view these conditional probabilities as a channel from R to Q. From the output Q, the weatherman constructs a channel from Q to the announcement A. If the goal of the weatherman is to maximize the information communicated, then the weatherman should do no processing and announce $A = Q$. However, the probability of a correct announcement is maximized using the maximum a posteriori decision rule. Given $Q = q$, the MAP decision is $\hat{r}$ where $\hat{r}$ is the argument of

$$\max_{\hat{r}=0,1} P_{R|Q}(\hat{r}|q) = \max_{\hat{r}=0,1} \frac{P_{R,Q}(\hat{r},q)}{P_Q(q)} = \max_{\hat{r}=0,1} P_{R,Q}(\hat{r},q) = \max_{\hat{r}=0,1} P_{Q|R}(q|\hat{r}) P_R(\hat{r})$$

When the weatherman learns $Q = 0$, the MAP rule becomes

$$\max_{\hat{r}=0,1} P_{Q|R}(0|\hat{r}) P_R(\hat{r}) = \max \left\{ \underbrace{\frac{10}{13}(1-p)}_{\hat{r}=0}, \underbrace{\frac{1}{3}p}_{\hat{r}=1} \right\}$$

The MAP rule is $A = \hat{r} = 0 = Q$ iff

$$\frac{10(1-p)}{13} > \frac{p}{3}, \quad \text{or} \quad \frac{30}{43} > p.$$

When the weatherman learns $Q = 1$, the MAP rule becomes

$$\max_{\hat{r}=0,1} P_{Q|R}(1|\hat{r}) P_R(\hat{r}) = \max \left\{ \underbrace{\frac{3}{13}(1-p)}_{\hat{r}=0}, \underbrace{\frac{2}{3}p}_{\hat{r}=1} \right\}$$

The MAP rule is $A = \hat{r} = 1 = Q$ iff

$$\frac{3(1-p)}{13} < \frac{2p}{3}, \quad \text{or} \quad \frac{9}{35} < p$$

Thus, for $p = P(R = 1)$, the probability of correct decision is maximized by the announcement $A = Q$ iff

$$\frac{9}{35} < p < \frac{30}{43}$$

Otherwise, if $p < 9/35$, then the probability of a correct announcement is maximized by the announcement $A = 0$ corresponding to "no rain;" while if $p > 30/43$, the probability of a correct announcement is maximized by $A = 1$, corresponding to "rain." Note that in the original problem, we were given that $p = 3/16$, which is less than $9/35$, and announcing "no rain" maximized the probability of a correct announcement.

(c) *5 points* Was it necessary in the preceding step to specify that the prediction Q was based on maximum likelihood?

It was necessary to specify that Q was based on a maximum likelihood decision because it made it clear that the output Q did not depend on the prior probabilities $p = P_R(1) = 1 - P_R(0)$. Otherwise, in examining how to maximize the probability of a correct decision based on Q, we would need to know how Q depends on the a priori distribution of R.

2. *30 points* Let X, Y, Z be an ensemble of discrete random variables. In each of the following problems, there exists an equality or inequality between the two quantities. Fill in the blank $\underline{\quad ? \quad}$ with the appropriate relationship ($\leq$, $=$, or $\geq$) and justify the correctness of that relationship.

(a) $I(X,Y;Z) \ \underline{\quad ? \quad}\ I(X;Z)$

The correct answer is

$$I(X,Y;Z) \ \underline{\quad \geq \quad}\ I(X;Z).$$

Proof: Since

$$\begin{aligned} I(X,Y;Z) &= H(Z) - H(Z|X,Y) \\ &\geq H(Z) - H(Z|X) \\ &= I(X;Z) \end{aligned}$$

Note that $H(Z|X,Y) \leq H(Z|X)$ because conditioning reduces entropy.

(b) $H(X|Z)$ $\underline{\quad ? \quad}$ $H(X,Y|Z)$

The correct answer is

$$H(X|Z) \quad \underline{\quad \leq \quad} \quad H(X,Y|Z).$$

Proof: By the chain rule

$$H(X,Y|Z) = H(X|Z) + H(Y|X,Z) \geq H(X|Z)$$

since $H(Y|X,Z) \geq 0$ *because entropy is always non-negative.*

(c) $H(X,Z) - H(X)$ $\underline{\quad ? \quad}$ $H(X,Y,Z) - H(X,Y)$

The correct answer is

$$H(X,Z) - H(X) \quad \underline{\quad \geq \quad} \quad H(X,Y,Z) - H(X,Y).$$

Proof: First we observe by the chain rule that

$$H(X,Z) - H(X) = H(X) + H(Z|X) - H(X) = H(Z|X).$$

Again by the chain rule,

$$H(X,Y,Z) - H(X,Y) = H(X,Y) + H(Z|X,Y) - H(X,Y) = H(Z|X,Y)$$

The inequality follows since $H(Z|X) \geq H(Z|X,Y)$.

3. *30 points* The process $X_1, X_2, \ldots$ is an iid Bernoulli (p) random sequence. Let $R_n = (X_1 + \cdots + X_n)/n$ denote the success rate of the process.

(a) In terms of R_n, characterize the set $A_\epsilon^{(n)}$ of typical sequences.

A sequence $x^n = (x_1, \ldots, x_n) \in A_\epsilon^{(n)}$ *if*

$$2^{-n(H(X)+\epsilon)} \leq P(x_1, \ldots, x_n) \leq 2^{-n(H(X)-\epsilon)},$$

or equivalently,

$$-n(H(X)+\epsilon) \leq \log P(x_1, \ldots, x_n) \leq -n(H(X)-\epsilon).$$

Since each X_i *is Bernoulli, we can write the Bernoulli PMF in the form*

$$P(X_i = x_i) = (1-p)^{1-x_i} p^{x_i}, \qquad x_i = 0, 1.$$

Since the X_i *are independent,*

$$\begin{aligned} P(x_1, \ldots, x_n) &= \prod_{i=1}^n P(X_i = x_i) \\ &= (1-p)^{n - \sum_{i=1}^n x_i} p^{\sum_{i=1}^n x_i} \\ &= (1-p)^{n - nr_n} p^{nr_n} \\ &= (1-p)^n \left(\frac{p}{1-p} \right)^{nr_n} \end{aligned}$$

where $r_n = (x_1 + \cdots + x_n)/n$. *Thus,*

$$\log P(x_1, \ldots, x_n) = n \log(1-p) + nr_n \log \frac{p}{1-p}.$$

Combining these facts, we see that $x^n \in A_\epsilon^{(n)}$ *iff*

$$-n(H(X)+\epsilon) \leq n\log(1-p) + nr_n \log \frac{p}{1-p} \leq -n(H(X)-\epsilon),$$

or, equivalently, $x^n \in A_\epsilon^{(n)}$,

$$-\epsilon \leq H(X) + \log(1-p) + r_n \log \frac{p}{1-p} \leq \epsilon.$$

Note that

$$H(X) + \log(1-p) = -p\log p - (1-p)\log(1-p) + \log(1-p)$$
$$= -p\log\frac{p}{1-p}.$$

It follows that $x^n \in A_\epsilon^{(n)}$ *if*

$$-\epsilon \le (r_n - p)\log\frac{p}{1-p} \le \epsilon.$$

(b) When $p = 1/2$, what sequences are typical?
When $p = 1/2$, *we see that* $\log(p/(1-p)) = 0$ *and a sequence* x^n *is typical if*

$$-\epsilon \le (r_n - p)\cdot 0 \le \epsilon.$$

That is, all sequences x^n *are typical for all* $\epsilon > 0$. *This should not be surprising since every sequence has the same probability.*

(c) For what values of $p > 1/2$, if any, does $A_\epsilon^{(n)}$ include the most probable sequence? For such p, does $A_\epsilon^{(n)}$ include the most probable sequence *for all* $\epsilon > 0$?
For $p > 1/2$, *the most probable sequence is the all one sequence* $x^n = (1, 1 \ldots, 1)$. *For the all one sequence,* $r_n = 1$, *implying the all one sequence is typical iff*

$$-\epsilon \le (1-p)\log\frac{p}{1-p} \le \epsilon.$$

For $p > 1/2$, *this condition simplifies to*

$$g(p) = (1-p)\log\frac{p}{1-p} \le \epsilon.$$

Note that $g(1/2) = 0$ *and* $\lim_{p\to 1} g(p) = 0$. *Note that* $g(p) > 0$ *for* $1/2 < p < 1$. *Also note that*

$$g''(p) = -\frac{1}{p^2(1-p)} < 0,$$

implying that $g(p)$ *is a concave function. Thus, there exists* δ_1 *and* δ_2 *such that the all one sequence is typical if*

$$1/2 \le p \le 1/2 + \delta_1, \quad \textit{or} \quad 1 - \delta_2 \le p \le 1.$$

Note that we can lower bound δ_1 *since* $\log x \le (x-1)\log e$ *implies*

$$g(p) \le (1-p)\left(\frac{p}{1-p} - 1\right)\log e = (2p-1)\log e.$$

Thus $(2p-1)\log e \le \epsilon$, *or equivalently,*

$$p \le \frac{1}{2} + \frac{\epsilon}{2\log e}$$

is a sufficient condition to ensure that the all-one sequence is typical. A similar sufficient condition can be derived for p *close to* 1. *Note that for* p *close to* $1/2$, *the all-one sequence is typical because every sequence with* $r_n \ge p$ *is typical. For* p *close to* 1, *the all one sequence is typical because the typical sequences are those with* r_n *close to* 1.

4. *20 points* Consider the code $\{0, 10, 01\}$ for a ternary source. Justify your answers to the following questions:

(a) *5 points* Is the code instantaneous?
No because 0 *is a prefix of* 01.

(b) *5 points* Is the code nonsingular?
Yes, because all the code words are unique.

(c) *5 points* Is the code uniquely decodable?

No because 010 *can be decoded as* 0, 10 *or* 01, 0.

(d) *5 points* Is there an instantaneous code with the same codeword lengths? If so, find an example of such a code.

The code has codeword lengths $l_1 = 1, l_2 = 2, l_3 = 3$. *Since*

$$\sum_{i=1}^{3} 2^{-l_i} = 2^{-1} + 2^{-2} + 2^{-2} = 1,$$

the Kraft inequality is satisfied with equality. Thus an instantaneous code with these lengths is satisfied. A simple example is $\{0, 10, 11\}$.

5. *50 points* The outcome of a roulette wheel is either red $X = 1$ or black $X = 0$, equiprobably and independently from spin to spin. By observing the ball until the last instant that bets can be placed, a gambler can predict X with some accuracy. Given the gambler's prediction, $Y = 0$ or $Y = 1$, conditional probabilities for X are given by

$$P_{X|Y}(1|1) = P_{X|Y}(0|0) = 3/4.$$

(a) Calculate the mutual information $I(X;Y)$.

The channel from X *to* Y *is a BSC with crossover probability* $\epsilon = 1/4$. *With equiprobable inputs, the outputs are equiprobable and* $H(y) = 1$. *For each input* $X = 0, 1$, *the conditional entropy of* Y *is*

$$H(Y|X=1) = H(Y|X=0) = H(\epsilon) = -\frac{1}{4}\log\frac{1}{4} - \frac{3}{4}\log\frac{3}{4}.$$

This implies $H(Y|X) = H(\epsilon)$ *and*

$$I(X;Y) = H(Y) - H(Y|X) = 1 - H(\epsilon) = \frac{3}{4}\log 3 - 1 = 0.1887\ \textit{bits}.$$

(b) The gambler has some initial capital C_0. On each spin, she bets a fraction $1-q$ of her total capital on the predicted color and a fraction q on the other color. Let $Z_n = 1$ if the gambler's prediction is correct on trial n. After N spins, the gamblers capital is the random variable C_N. Express C_N in terms of $Z_1, \ldots, Z_N$.

Given capital C_{n-1} *after* $n-1$ *steps, the gambler's capital after step* n *will be*

$$C_n = \begin{cases} 2C_{n-1}(1-q) & Z_n = 1 \\ 2C_{n-1}q & Z_n = 0 \end{cases}$$

In terms of the outcome $Z_n = 0, 1$, *this can be expressed as*

$$C_n = [2C_{n-1}(1-q)]^{Z_n}[2C_{n-1}q]^{1-Z_n} = 2C_{n-1}(1-q)^{Z_n}q^{1-Z_n}$$

It follows that

$$C_N = C_0 2^N \prod_{n=1}^{N} \left((1-q)^{Z_n} q^{1-Z_n}\right)$$

Comment: An equivalent and perhaps simpler formulation is

$$C_N = C_0 2^N \prod_{n=1}^{N} \left((1-q)Z_n + q(1-Z_n)\right)$$

(c) Find q_C^*, the value of q that maximizes the expected value $E[C_N]$.

Note that $Z_1, Z_2, \ldots$ is a seuqnece of iid Bernoulli ($p = 3/4$) random variables. This implies

$$
\begin{aligned}
E[C_N] &= C_0 2^N E\left[\prod_{n=1}^{N}\left((1-q)^{Z_n} q^{1-Z_n}\right)\right] \\
&= C_0 2^N \prod_{n=1}^{N} E\left[(1-q)^{Z_n} q^{1-Z_n}\right] \\
&= C_0 2^N \prod_{n=1}^{N} \left(\frac{3}{4}(1-q) + \frac{1}{4}q\right) \\
&= C_0 2^N \left(\frac{3}{4} - \frac{q}{2}\right)^N \\
&= C_0 \left(\frac{3}{2} - q\right)^N
\end{aligned}
$$

Note that $E[C_N]$ is maximized by choosing $q = q_C^* = 0$, i.e., she bets all her money on the prediction.

(d) Define the rate of growth as

$$
R_N = \frac{1}{N}\log_2 \frac{C_N}{C_0}.
$$

Find q_R^*, the value of q that maximizes the expected value $E[R_N]$. For $q = q_R^*$, compare $E[R_N]$ and $I(X;Y)$.

$$
\begin{aligned}
R_N &= \frac{1}{N}\left(N + \sum_{n=1}^{N} \log\left[(1-q)^{Z_n} q^{1-Z_n}\right]\right) \\
&= 1 + \frac{1}{N}\sum_{n=1}^{N} \left[Z_n \log(1-q) + (1-Z_n)\log(q)\right].
\end{aligned}
$$

Since the expected value of a sum is the sum of the expected values,

$$
\begin{aligned}
E[R_N] &= 1 + \frac{1}{N}\sum_{n=1}^{N} E\left[Z_n \log(1-q) + (1-Z_n)\log(q)\right] \\
&= 1 + \frac{1}{N}\sum_{n=1}^{N} \left[E[Z_n]\log(1-q) + (1-E[Z_n])\log(q)\right] \\
&= 1 + \left[\frac{3}{4}\log(1-q) + \frac{1}{4}\log(q)\right]
\end{aligned}
$$

By taking the derivative of $E[R_N]$ with respect to q, we find that $E[R_N]$ is maximized at $q = q_R^* = 1/4$. In this case, the expected rate of return is

$$
E[R_N] = 1 + \frac{3}{4}\log\frac{3}{4} + \frac{1}{4}\log\frac{1}{4} = 1 - H(\epsilon) = I(X;Y).
$$

(e) If you were the gambler, would you use $q = q_R^*$ or $q = q_C^*$? Explain why.

If you use $q = q_C^* = 0$, then you go broke as soon as your prediction is wrong. The probability you have zero capital at time N is

$$
P(C_N = 0) = 1 - (1-\epsilon)^N,
$$

which goes to 1 as $N \to \infty$. Thus, using $q = q_C^*$ guarantees she eventually goes broke. On the other hand, if you use $q = q_R^* = 1/4$, then

$$
R_N = 1 + \frac{1}{N}\sum_{n=1}^{N}\left[Z_n \log\frac{3}{4} + (1-Z_n)\log\frac{1}{4}\right].
$$

If we define

$$W_n = Z_n \log \frac{3}{4} + (1 - Z_n) \log \frac{1}{4},$$

then

$$R_N = 1 + \frac{1}{N} \sum_{n=1}^{N} W_n.$$

Since $W_1, W_2, \dots$ *is an iid sum, we know by the law of large numbers that* R_N *will converge to* $1 + E[W_n] = E[R_N] = I(X;Y)$. *In this case,* C_N *will be close to* $C_0 2^{NI(X;Y)}$. *The law of large numbers promises us that* $q = q_R^*$ *is a good solution.*