問題3

A君がコイン投げ実験を複数回行う．試行は互いに独立であるとする．各試行に対して，あなたは表が出るのか裏が出るのか，賭けをする．B君が「A君の投げる動作を見ただけで結果を予測できる．君が賭ける前に僕の予測を教えよう．」と言ってきた．実際にB君が，A君の投げる動作を見ただけで予測したところ，B君の予測精度は下表の通りであった．X はB君の予測，Y はA君の実験結果であり，表は結合確率分布 $P(X,Y)$ である．この表を既知として以下の問に答えよ．解答には $\log_2 3$，$\log_2 5$，$\log_2 7$ の表現を用いてよい．

$P(X,Y)$		X: Bの予測	
		表	裏
Y: 結果	表	0.45	0.25
	裏	0.15	0.15

(1) B君の予測を参照せずに賭ける．各試行において，実験結果を知ることにより得られる情報量の期待値（即ち，エントロピー $H(Y)$）は幾らか？計算手順も示せ．

(2) 賭ける前にB君の予測を参照する．B君の予測を参照した後に，実験結果を知ることにより得られる情報量の期待値（即ち，条件付きエントロピー $H(Y|X)$）は幾らか？計算手順も示せ．

(3) B君の予測を参照することで得られる情報量（即ち，相互情報量 $I(Y;X)$）は幾らか？計算手順も示せ．

(4) あなたは10回賭けをする．B君の予測を参照した後で賭ける．賭けに勝つと，毎回，賞金100円をもらえる．以下の問に答えよ．

(4-1) 賞金総額の期待値を最大にしたい．常にB君の予測通りに賭けることは有益か？理由を付して答えよ．

(4-2) 各試行に対して，B君の予測を参照した後に表・裏・パスのいずれかを選択できることとなった．パスとは，賭けを行わず，次のコイン投げ実験に移ることを意味する．そして賭けを10回行ったら実験を終了する．この時，B君の予測を参照することは有益か？理由を付して答えよ．

コイン投げ実験は，2種類のシンボル $\{z_1, z_2\}$ を値として持つ確率変数 Z としてモデル化できる．シンボル数を $N(>2)$ に増やした場合の Z を考える．Z の値を知ることにより得られる情報量に関する，以下の問に答えよ．

(5) 情報量の期待値（即ち，エントロピー$H(Z)$）が最大となる確率分布 $P(Z)$ を答えよ．解答には数学的証明を含めること．

(6) 事象 i の自己情報量（自己エントロピー）$-\log_2 p_i$ は，それが生起したことに対する「驚きの度合い（degree of surprise）」と解釈される．但し p_i は事象 i の生起確率である．今，英単語が提示されることを期待している観測者を考える．この観測者に情報理論的に最大の驚きを与える語，即ち，最も英単語とは思えない語を提示したい．ここでは，小文字を5個繋げて構成される語のみを考える（英単語の平均文字数は5であり，また $N=26$ となる）．C君が「問(5)の $P(Z)$ に従って文字を5つ生成し，それ

らを繋げて語を作るべきだ．」と主張した．彼の主張が妥当か否かを，理由を付して論ぜよ．主張が妥当な場合，N が 26 より大きいと，同じ方法を用いた時の驚きの度合いがより大きくなるか否かを，理由を付して論ぜよ．主張が妥当でない場合，情報理論的に最大の驚きを与える語の生成方法を論ぜよ．

Problem 3

Mr. **A** conducts a series of coin tossing experiments, which are assumed to be independent of one another. In each experiment, you make a bet on the result, namely, heads or tails. Mr. **B** said "I can predict the result only by observing **A**'s motion of tossing and I'll tell you my prediction before you make a bet." It was examined how correctly **B** can predict the result only by observing **A**'s motion of tossing. **B**'s performance of prediction is shown in the table below as joint probability $P(X,Y)$ of X for **B**'s predictions and Y for the results of **A**'s experiments. Answer the following questions, where you can use this table as prior knowledge. Your answer can include the expressions of $\log_2 3$, $\log_2 5$, and $\log_2 7$.

$P(X,Y)$		X: B's prediction	
		heads	tails
Y: Results	heads	0.45	0.25
	tails	0.15	0.15

(1) You make a bet without referring to **B**'s prediction. Answer the expected quantity of information that you can obtain in each experiment by knowing the result of the experiment, namely, entropy $H(Y)$. Show the procedure of calculation.

(2) You make a bet after referring to **B**'s prediction. Answer the expected quantity of information that you can obtain by knowing the result after you refer to **B**'s prediction, namely, conditional entropy $H(Y|X)$. Show the procedure of calculation.

(3) Answer the quantity of information that you can obtain by referring to **B**'s prediction, namely, mutual information $I(Y;X)$. Show the procedure of calculation.

(4) You make a bet ten times. After you refer to **B**'s prediction, you make a bet. Each time you win, you receive a reward of 100 yen. Answer the following questions.

(4-1) You want to maximize the expected value of the total amount of reward. Is it beneficial for you to make a bet by always following **B**'s prediction? Answer with a reason.

(4-2) In each experiment, you can select heads, tails or pass after you refer to **B**'s prediction. Here, pass means that you do not make a bet and proceed to the next experiment. When you make a bet ten times eventually, the tossing experiments end. In this condition, is it beneficial for you to refer to **B**'s predictions? Answer with a reason.

The coin tossing experiment can be modeled as random variable Z that takes one of the two symbols $\{z_1, z_2\}$ as its value. Here, we consider a general case where the number of symbols is N ($N > 2$) for Z. Answer the following questions related to the quantity of information that you can obtain by knowing the value of Z.

(5) Answer the probability distribution of $P(Z)$ that maximizes the expected quantity of

information, namely, entropy $H(Z)$. Your answer has to be with a mathematical proof.

(6) The quantity of self-information (self-entropy) of event i, $-\log_2 p_i$, is interpreted as "degree of surprise" when that event happens. Here, p_i is the probability that event i happens. Now, we assume an observer who is expecting an English word to be presented. We would like to present a word to the observer that gives the biggest surprise in terms of the information theory, namely, a word that looks least probable as an English word. Here, we consider only the words that are composed of five small letters (The average number of letters in English words is five and $N = 26$). Mr. **C** claimed that we have to generate five letters by obeying $P(Z)$ of Question (5), and concatenate them to form such a word. Examine whether his claim is valid or not with a reason. If it is valid, examine with a reason whether the degree of surprise becomes larger when N is larger than 26, assuming the same strategy. If it is not valid, describe a method to make a word that gives the biggest surprise in terms of the information theory.