黒インクでご記入ください．
持込欄の不可・可のどちらかに 必ず○ をつけてください．

<table>
<tr><td colspan="5">2008 年度春・秋学期中間試験</td><td>問題枚数</td><td colspan="2">1/1</td></tr>
<tr><td colspan="2">科目名</td><td>出題者氏名</td><td>受験クラス</td><td colspan="2">学生証番号</td><td colspan="2">氏名</td></tr>
<tr><td colspan="2">情報理論</td><td>山本宙</td><td>DD-2B,DD-3,
DM-2B,DM-3,
その他</td><td colspan="2"></td><td colspan="2"></td></tr>
<tr><td rowspan="2">持込</td><td>不可</td><td>◇可の場合は，記入</td><td>開講曜日・時限</td><td rowspan="2">現在使用している授業教室</td><td rowspan="2">6B — 104</td><td rowspan="2">採点</td><td rowspan="2"></td></tr>
<tr><td>可</td><td>関数電卓のみ</td><td>木曜 3,4 限</td></tr>
</table>

注意事項

答案は解答用紙に書け．答案用紙の裏を使用する場合は表の最後に「裏に続く」と記入せよ．単に log と書いた場合の対数の底は 2 であるとする．また，断らない限り情報量の単位は 2 を底とした “ビット” を使用する．解答に変数が含まれない場合，整数または小数で表し，四捨五入して小数点以下 2 桁まで求めよ．分母，分子とも整数の分数で表せる場合は既約分数で答えてもよい．関数電卓の使用を許可する．但し，電卓としてであっても携帯電話の使用は認めない．和は省略記法 $+\cdots+$ は用いず，$\sum$ を用いて表現せよ．対数の値として，$\log 3 = 1.584962, \log 5 = 2.321928, \log 7 = 2.807355, \log 11 = 3.459431, \log 13 = 3.700440, \log 17 = 4.087463, \log 19 = 4.247927, \log 23 = 4.523562, \log 29 = 4.857981, \log 31 = 4.954196, \log 37 = 5.209453, \log 41 = 5.357552, \log 43 = 5.426264, \log 47 = 5.554589, \log 53 = 5.727920$ を使用してもよい．$P(E_2/E_1)$ は事象 E_1 が起こった条件の下で E_2 が起こる条件付確率を表す．$P(E_1, E_2)$ は E_1, E_2 の順番に事象が続けて起こる確率を表す．

問 1 (11 点)

$x_1, x_2, \ldots, x_q$ と $y_1, y_2, \ldots, y_q$ を 2 組の確率とする．以下の式を証明せよ．

$$\sum_{i=1}^{q} x_i \left(\frac{y_i}{x_i} - 1 \right) = 0$$

問 2 (各 4 点，計 16 点)

通報シンボルが $S = \{s_1, s_2, s_3, s_4\}$ でそれぞれの生起確率が $P(s_1) = a, P(s_2) = P(s_3) = 1/8, P(s_4) = b$ である無記憶情報源 S を考える．s_1, s_2, s_3, s_4 をそれぞれ $00, 01, 10, 11$ に対応させる符号を $\mathcal{A}$，s_1, s_2, s_3, s_4 をそれぞれ $0, 10, 110, 111$ に対応させる符号を $\mathcal{B}$ とする．以下の問いに答えよ．

1. 符号 $\mathcal{A}$ を用いたときの符号語の平均長 (単位はビニット) を求めよ．
2. 符号 $\mathcal{B}$ を用いたときの符号語の平均長 (単位はビニット) を a で表せ．
3. 符号 $\mathcal{B}$ を用いたときの符号語の平均長が符号 $\mathcal{A}$ を用いたときより小さくなる条件を求めよ．
4. $a = 1/8$ であるとき， s_1 の発生を知ったときの情報量を求めよ．

問 3 (1-3 各 6 点, 4 20 点，計 38 点)

2 つの無記憶情報源 S_1 および S_2 があり，それらの情報源シンボルの個数をそれぞれ q_1 および q_2 とする．また S_1 のシンボルの生起確率を P_i, $(i = 1, 2, \ldots, q_1)$; S_2 のシンボルの生起確率を Q_i, $(i = 1, 2, \ldots, q_2)$; S_1 および S_2 のエントロピーをそれぞれ H_1 および H_2 とする．ここで新しく，1 つの無記憶情報源 $S(\lambda)$, すなわち S_1 と S_2 の混合 (mixture) を定義する．この情報源 $S(\lambda)$ は $q_1 + q_2$ 個のシンボル s_i, $(1 \le i \le q_1 + q_2)$ を有し，$S(\lambda)$ の最初の q_1 個のシンボルの生起確率 $P(s_i)$, $(i = 1, 2, \ldots, q_1)$ がそれぞれ λP_i で与えられ，また残りの q_2 個のシンボル生起確率 $P(s_i)$, $(i = q_1 + 1, q_1 + 2, \ldots, q_1 + q_2)$ がそれぞれ $\overline{\lambda} Q_{i-q_1}$ で与えられるものとする (ただし，ここで $\overline{\lambda} = 1 - \lambda$ とする)．以下の問に答えよ．

1. H_1 を P_i, $(i = 1, 2, \ldots, q_1)$ から求める式を書け．
2. $1 \le i \le q_1$ なる i について $I(s_i)$ を求める式を書け．
3. $q_1 + 1 \le i \le q_1 + q_2$ なる i について $I(s_i)$ を求める式を書け．
4. $H[S(\lambda)] = \lambda H_1 + \overline{\lambda} H_2 + H(\lambda)$ を証明せよ．

問 4 (各 5 点，計 30 点)

情報源シンボル $S = \{0, 1\}$ が単純マルコフ情報源であり，$P(0/0) = 3/4, P(0/1) = 1/8$，また $P(0) = 1/3, P(1) = 2/3$ であるとする．以下の問に答えよ．

1. $P(00), P(01), P(10), P(11)$ の値を求めよ．
2. 状態が s_j であるときに s_i が生じたときに受けとる情報量を $I(s_i/s_j)$ と書くとき，$I(0/0)$, $I(1/0)$, $I(0/1)$, $I(1/1)$ を求めよ．
3. 状態が 0 であるときのシンボルあたりの平均情報量 $H(S/0)$, 状態が 1 であるときのシンボルあたりの平均情報量 $H(S/1)$ を求めよ．
4. S のエントロピー $H(S)$ を求めよ．
5. S の随伴情報源 $\overline{S}$ のエントロピー $H(\overline{S})$ を求めよ．
6. S の 2 次拡大 $S^2 = \{s_1, s_2, s_3, s_4\}$ を考える．但し $s_1 = 00, s_2 = 01, s_3 = 10, s_4 = 11$ とする．$P(s_1/s_1), P(s_2/s_1), P(s_3/s_1), P(s_4/s_1)$ の値を求めよ．

問 5 (5 点)

英語に使われる文字が，$S = \{\mathrm{A}, \mathrm{B}, \ldots, \mathrm{Z}, \_\}$ の 27 文字だけであると仮定する．ここで “_” はスペースの意味とする．

手順 十分に長い英語の文章 (テキストとよぶ) から無作為にに 1 つの文字を選び出す．これが例えば U であったならば，次に何行かをとばしてふたたび U という文字が現れるところまで読み，この U の文字のすぐ次の文字を選び出す．これが例えば R であったならばふたたび何行かをとばし，UR という文字が現れるまで読み，この UR のすぐ次の文字を選び出す．これが例えば K であったならばふたたび何行かをとばし，RK という文字が現れるまで読み，この RK のすぐ次の文字を選び出す．このように先行する 2 つの文字に従って，つぎの文字を選び出してゆく手順を繰り返す．

上記手順で生成された文字列は英語を何重マルコフ情報源とみなして近似したものか答えよ．