

## 2007 年度 情報数理 レポート 2

学生用

学籍番号： \_\_\_\_\_

氏名： \_\_\_\_\_

下記の注意事項を守り、次ページ以降の問いに答え、レポートを完成させなさい。

提出期限： 2007 年 11 月 13 日 (火) 13:00 まで

提出場所： 理学部棟 正面玄関内に設置のレポートボックス

### 注意事項：

- (1) このページを印刷し、必要事項を記入の上 (学籍番号欄と氏名欄は 2 箇所あるので忘れずに記入すること)、レポートの表紙として提出すること。
- (2) ~~文章処理ソフトウェアや図形処理ソフトウェア等を駆使してレポートを作成し (問→解答→問→解答→… の順になるように記述すること)、A4 サイズの用紙に印刷して提出すること (手書きは不可)。~~
- (3) クラスメイトのレポートを参考にしたり、クラスメイトと協力してレポートを作成した場合は、教員控の協力者氏名欄にクラスメイトの氏名を記入すること。これらの場合も、自分の言葉で表現し直すこと。**コピー禁止。**
- (4) 情報数理について、あなたの声を聞かせてください (教員控の意見・質問欄に記入のこと)。気軽にどうぞ (成績には一切影響しません)。

出題者： 幸山 直人

出題日： 2007 年 10 月 31 日 (水)

得点：

/6

----- 切り取り線 -----

## 2007 年度 情報数理 レポート 2

教員控

学籍番号： \_\_\_\_\_

氏名： \_\_\_\_\_

協力者氏名： \_\_\_\_\_ , \_\_\_\_\_ , \_\_\_\_\_

レポート作成に要した時間： \_\_\_\_\_ . \_\_\_\_\_ 時間

得点：

/6

意見・質問：

**問 1** テキスト (14 ページ) の「図 1.9: 3 個ずつまとめた Huffman の符号化法 II」について、「(a) 構成法」を描きなさい。ただし、右の「(c) 符号」に一致するようにすること。(2 点)

**解答例** まず、3 個ずつまとめた情報源記号の生起確率 (独立に生起) を計算すると以下のようになる。

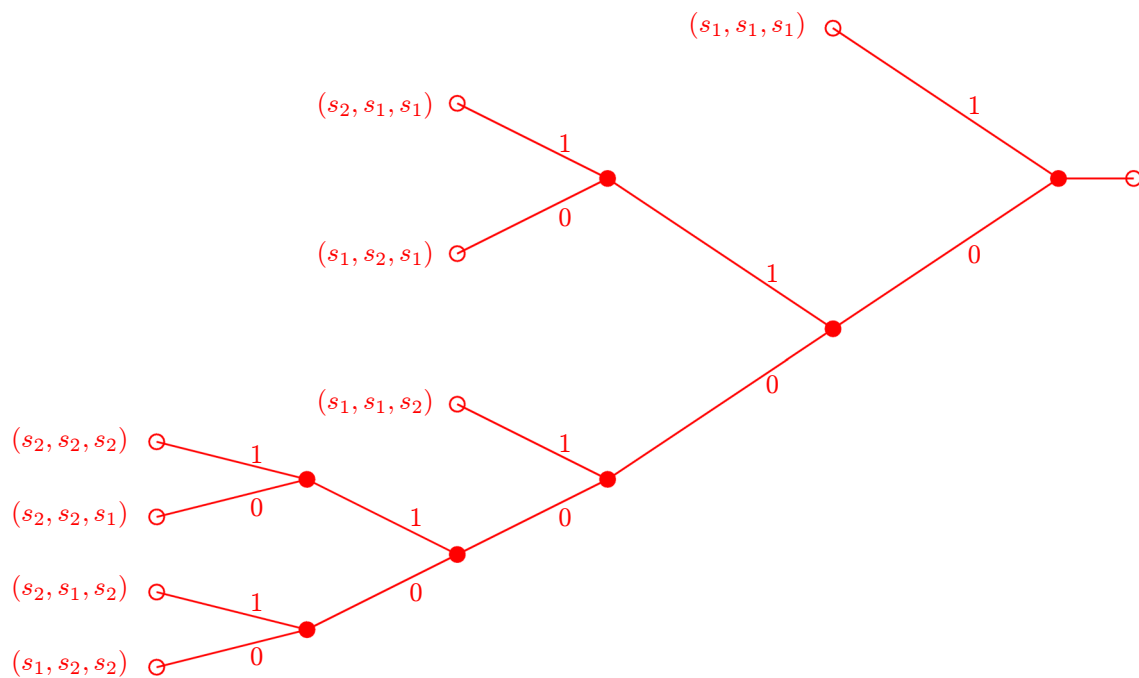
$$p(s_1, s_1, s_1) = \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{27}{64},$$

$$p(s_1, s_1, s_2) = \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{1}{4} = \frac{9}{64}, \quad p(s_1, s_2, s_1) = \frac{3}{4} \cdot \frac{1}{4} \cdot \frac{3}{4} = \frac{9}{64}, \quad p(s_2, s_1, s_1) = \frac{1}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{9}{64},$$

$$p(s_1, s_2, s_2) = \frac{3}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} = \frac{3}{64}, \quad p(s_2, s_1, s_2) = \frac{1}{4} \cdot \frac{3}{4} \cdot \frac{1}{4} = \frac{3}{64}, \quad p(s_2, s_2, s_1) = \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{64},$$

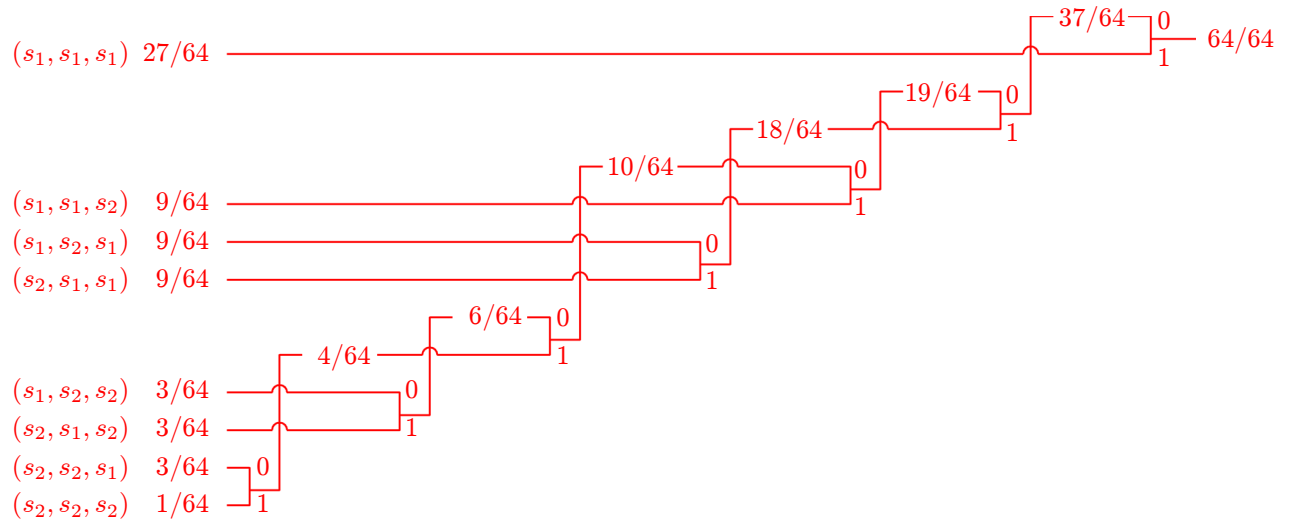
$$p(s_2, s_2, s_2) = \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{64}.$$

次に、「(c) 符号」の構造を調べるために「(b) 符号の木」を描くと以下のようになる。



(b) 符号の木

描いた「(b) 符号の木」を参考に、生起確率の高い記号から順に並べ、生起確率の低い 2 つの記号を組にしながら合成確率が 1 になるまで繰り返す (Huffman の符号化法 I を実行する)。さらに、「(b) 符号の木」と一致するように 2 元符号 (0 と 1) を割り当てる。



(a) 構成法

**評価基準** 解答例に準じた解答であれば 2 点。

**問 2** テキスト (17 ページ) の補助定理 2.2 の証明を完了しなさい。テキストに記した  $n = 2$  の場合を参考にするとよい。(2 点)

**解答例**

$$\begin{aligned}
 (\text{左辺}) &= H(S^n) = \sum_{i=1}^{M^n} p(\sigma_i) \log_r \frac{1}{p(\sigma_i)} \\
 &= \sum_{i_1=1}^M \sum_{i_2=1}^M \cdots \sum_{i_n=1}^M p(s_{i_1}, s_{i_2}, \dots, s_{i_n}) \log_r \frac{1}{p(s_{i_1}, s_{i_2}, \dots, s_{i_n})}
 \end{aligned}$$

となる。ここで、記号が独立に生起することから  $p(s_{i_1}, s_{i_2}, \dots, s_{i_n}) = p(s_{i_1})p(s_{i_2}) \cdots p(s_{i_n})$  を代入すると

$$\begin{aligned}
 (\text{与式}) &= \sum_{i_1=1}^M \sum_{i_2=1}^M \cdots \sum_{i_n=1}^M p(s_{i_1}, s_{i_2}, \dots, s_{i_n}) \log_r \frac{1}{p(s_{i_1})p(s_{i_2}) \cdots p(s_{i_n})} \\
 &= \sum_{i_1=1}^M \sum_{i_2=1}^M \cdots \sum_{i_n=1}^M p(s_{i_1}, s_{i_2}, \dots, s_{i_n}) \log_r \frac{1}{p(s_{i_1})} \\
 &\quad + \sum_{i_1=1}^M \sum_{i_2=1}^M \cdots \sum_{i_n=1}^M p(s_{i_1}, s_{i_2}, \dots, s_{i_n}) \log_r \frac{1}{p(s_{i_2})} \\
 &\quad \vdots \\
 &\quad + \sum_{i_1=1}^M \sum_{i_2=1}^M \cdots \sum_{i_n=1}^M p(s_{i_1}, s_{i_2}, \dots, s_{i_n}) \log_r \frac{1}{p(s_{i_n})}
 \end{aligned}$$

となる。また、確率の性質から第  $j$  項 ( $j = 1, 2, \dots, n$ ) に対して

$$p(s_{i_j}) = \sum_{i_1=1}^M \sum_{i_2=1}^M \cdots \sum_{i_{j-1}=1}^M \sum_{i_{j+1}=1}^M \cdots \sum_{i_n=1}^M p(s_{i_1}, s_{i_2}, \dots, s_{i_n})$$

が成り立つので

$$\begin{aligned}
 (\text{与式}) &= \sum_{i_1=1}^M \left( \sum_{i_2=1}^M \sum_{i_3=1}^M \cdots \sum_{i_n=1}^M p(s_{i_1}, s_{i_2}, \dots, s_{i_n}) \right) \log_r \frac{1}{p(s_{i_1})} \\
 &\quad + \sum_{i_2=1}^M \left( \sum_{i_1=1}^M \sum_{i_3=1}^M \cdots \sum_{i_n=1}^M p(s_{i_1}, s_{i_2}, \dots, s_{i_n}) \right) \log_r \frac{1}{p(s_{i_2})} \\
 &\quad \vdots \\
 &\quad + \sum_{i_n=1}^M \left( \sum_{i_1=1}^M \sum_{i_2=1}^M \cdots \sum_{i_{n-1}=1}^M p(s_{i_1}, s_{i_2}, \dots, s_{i_n}) \right) \log_r \frac{1}{p(s_{i_n})} \\
 &= \sum_{i_1=1}^M p(s_{i_1}) \log_r \frac{1}{p(s_{i_1})} + \sum_{i_2=1}^M p(s_{i_2}) \log_r \frac{1}{p(s_{i_2})} + \cdots + \sum_{i_n=1}^M p(s_{i_n}) \log_r \frac{1}{p(s_{i_n})}
 \end{aligned}$$

となる。なお、与式の各項はそれぞれ

$$\sum_{i=1}^M p(s_i) \log_r \frac{1}{p(s_i)}$$

の形をしており、 $H(S)$  に等しい。ゆえに、

$$H(S^n) = nH(S)$$

が成り立つ。 ■

**評価基準** 解答例に準じた解答であれば 2 点。

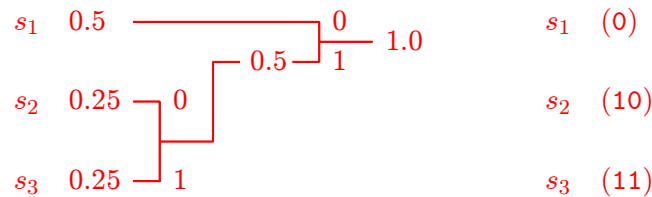
**問 3** 情報源

$$S = \left\{ \begin{array}{ccc} s_1, & s_2, & s_3 \\ p(s_1) = \frac{1}{2}, & p(s_2) = \frac{1}{4}, & p(s_3) = \frac{1}{4} \end{array} \right\}$$

について、以下の (1)～(2) の問いに答えなさい。ただし、 $s_1, s_2, s_3$  は独立に生起するものとする。

(1) 情報源  $S$  を Huffman の符号化法で符号化し、 $S$  の平均符号長  $L$  を求めなさい。(1 点)

**解答例** Huffman の符号化法 I を実行すると



となる。このとき、符号化された  $s_1, s_2, s_3$  の符号長はそれぞれ  $l_1 = 1, l_2 = 2, l_3 = 2$  となる。したがって、平均符号長  $L$  は

$$L = \sum_{i=1}^3 p(s_i) l_i = \frac{1}{2} \times 1 + \frac{1}{4} \times 2 + \frac{1}{4} \times 2 = 1.5 \text{ (ビット)}$$

である。

(2) 情報源  $S$  の平均符号長の下限  $H(S)$  を求めなさい。(1 点)

（この問いは、全ての生起確率が  $1/r^k$  ( $r = 2; k$  は正の整数) の形をしているとき、平均符号長  $L$  と平均符号長の下限  $H(S)$  が一致する具体例である。このような場合、Huffman の符号化法 II を適用しても 1 文字あたりの平均符号長は変わらない。

**解答例**

$$\begin{aligned} H(S) &= \sum_{i=1}^3 p(s_i) \log_2 \frac{1}{p(s_i)} \\ &= \frac{1}{2} \log_2 2 + \frac{1}{4} \log_2 4 + \frac{1}{4} \log_2 4 = \frac{1}{2} \times 1 + \frac{1}{4} \times 2 + \frac{1}{4} \times 2 = 1.5 \text{ (ビット)}. \end{aligned}$$

\* 符号長  $l_i$  と生起確率  $p(s_i)$  の間に、関係式  $l_i = \log_2 \frac{1}{p(s_i)}$  が成り立つ特殊な例である。

**評価基準** 解答例に準じた解答であれば各 1 点。