

信源表示：率失真特性

4

本讲和下一讲关注信息源发出的消息的表示，如第1讲图1.1所示。假设信息源发出的消息属于一个有限集合 $S = \{a_1, a_2, \dots, a_{|S|}\}$ 。那么，如果希望用一串二进制数字，即比特，来无歧义地表示该消息，显然这串数字的长度至少应为 $\lceil \log_2 |S| \rceil$ 。但我们能用更短的字符串来完成这个任务吗？如果我们允许表示中存在一些模糊性，答案是肯定的。例如，考虑 $S = \{1, 2, 3, 4, 5, 6, 7, 8\}$ ，并假设希望再现消息的目的地不在乎再现的消息与源消息的差异是否超过1。那么，我们可以安全地将1、2、3中的任何一个表示为2，将4、5、6中的任何一个表示为5，将7、8中的任何一个表示为8。因此，长度为 $\lceil \log_2 3 \rceil = 2$ 的二进制字符串就足够了，而不是在不允许任何模糊性时所需的 $\lceil \log_2 8 \rceil = 3$ 。

从上面的简单例子中，我们清楚地看到，用于表示源消息的资源量（例如，二进制字符串的长度）与再现消息的质量（例如，模糊程度）之间存在一种权衡。如果考虑到信源的概率性质，这种权衡将变得更加有趣。率失真理论描述了这种权衡的基本极限，是本讲的主题。它的极端情况，即失真为零或几乎为零的情况，需要特殊处理，并将在下一讲中进行研究。

4.1 问题描述 43 4.2 香农信源编码基本定理 46 4.3 逆定理部分的证明 53 4.4 可达性部分的证明 55 4.4.1 码本的生成 56 4.4.2 编码 56 4.4.3 期望失真分析 58 4.4.4 编码失败概率的估计 58 4.4.5 总结步骤 60

4.1 问题描述

在第1讲图1.1的一般通信系统模型中，我们将信息源建模为一个产生随机过程 S_1, S_2, \dots 的概率设备。一个消息则对应于该随机过程的一个指定长度 n 的段，即 $S = [S_1, S_2, \dots, S_n]$ 。

在本讲中，我们关注信息源是离散无记忆信源（DMS）的情况；也就是说， S_1, S_2, \dots 是一系列独立同分布（i.i.d.）的随机变量，每个变量的概率质量函数（pmf）为 $P_S(s)$ ，字母表为 \mathcal{S} 。

44

4 信源表示：率失真特性

为了表示一个信源消息，我们需要为 \mathcal{S} 的每一个可能值 $s \in \mathcal{S}$ 分配一个从某个有限集合中选出的索引，不失一般性，这个集合可以固定为 $\{1, 2, \dots, M_n\}$ 。这个分配通过一个映射完成：

$$f_n(s): \mathcal{S} \mapsto \{1, 2, \dots, M_n\}, \quad (4.1)$$

我们称之为**信源编码器**。这里的下标 n 用来强调对信源消息长度的依赖性，上标 (s) 用来表示该映射用于信源编码，以区别于第6讲中的信道编码。给定 $f_n(s)$ ，索引 $W = f_n(s)(S)$ 就是由信源消息随机向量 S 导出的一个随机变量。

假设索引 W 已被告知目的地。目的地需要将信源消息再现为 $\hat{S} = [\hat{S}^1, \hat{S}^2, \dots, \hat{S}^n]$ ，这也是一个长度为 n 的随机向量。但我们允许 $\hat{S}^i, i=1, 2, \dots, n$ 在一个字母表 $\hat{\mathcal{S}}$ 中取值，这个字母表通常可能与信源字母表 \mathcal{S} 不同。这种再现通过一个映射完成：

$$g_n(s): \{1, 2, \dots, M_n\} \mapsto \hat{\mathcal{S}}^n, \quad (4.2)$$

我们称之为**信源译码器**。总结来说，我们有以下马尔可夫链：

$$S \leftrightarrow W = f_n(s)(S) \leftrightarrow \hat{S} = g_n(s)(W) = g_n(s)(f_n(s)(S)). \quad (4.3)$$

图4.1展示了信源编码和译码过程。

图4.1：信源编码和译码示意图。DMS（离散无记忆信源）产生 S ，经过信源编码器得到 W ，再经过信源译码器得到 \hat{S} 。

图4.1：信源编码和译码示意图。

注4.1 这个问题的关键特征是马尔可夫链关系 (4.3)。实际上，可以用一些条件概率分布 $P_{W|S}$ 和 $P_{S^{\wedge}|W}$ 分别替代确定性映射 $(f_n(s)$ 和 $g_n(s))$ ，而本讲的核心结果，即下一节的香农信源编码基本定理，仍然成立。

注4.2 允许再现字母表 S^{\wedge} 与信源字母表 S 不同，初看起来可能有些奇怪。但这大大增加了问题描述的适用性，因为它使得目的地能够完成与信源相关的不同任务。例如，在图像分类任务中，信源是一个图像，而目的地感兴趣的是决定该图像属于哪个类别（例如，动物、人物或风景）。

--- 第 3 页 ---

在这个例子中，信源消息（即图像）是一个像素阵列，而再现的消息只是一个指示信源消息类别的标签。

一个信源编码器将每个长度为 n 的信源消息段表示为 M^n 个索引中的一个，这可以存储为一个长度为 $\lceil \log_2 M^n \rceil$ 的二进制字符串。平均而言，每个信源符号因此由 $\lceil \log_2 M \rceil / n$ 个比特表示。我们澄清一下，这里的术语“比特”对应于计算机等设备中的存储单位，不应与我们在第2讲中介绍的信息度量混淆。

因此，我们将信源编码器/译码器对的**速率**定义为：

$$R = n \lceil \log_2 M^n \rceil \text{ 比特/信源符号。 (4.4)}$$

由于我们允许在目的地的再现消息中存在一定程度的模糊性，我们在这里引入**失真**的概念。**失真度量** d 是一个映射，它为每对 $(s, s^{\wedge}) \in S \times S^{\wedge}$ 分配一个非负数 $d(s, s^{\wedge})$ ，用以量化将 s 再现为 s^{\wedge} 所产生的成本。在我们的讲义中，除非另有说明，我们考虑有界失真度量；也就是说， $d_{\max} := \max_{(s, s^{\wedge}) \in S \times S^{\wedge}} d(s, s^{\wedge}) < \infty$ 。

例 4.1 对于**汉明失真**，我们有 $S^{\wedge} = S$ ，并设如果 $s = s^{\wedge}$ 则 $d(s, s^{\wedge}) = 0$ ，否则为 1。注意对于汉明失真， $E[d(S, S^{\wedge})] = P(S \neq S^{\wedge})$ 。

对于一个信源编码器/译码器对 $(f_n(s), g_n(s))$ ，我们进一步对 S 和 S^{\wedge} 之间的失真施加一个可加结构：

$$d(s, s^{\wedge}) = n^{-1} \sum_{i=1}^n d(s_i, s_i^{\wedge}); \quad (4.5)$$

也就是说，一个信源消息与其对应的再现消息之间的失真是每个信源符号与其对应的再现符号之间成对失真的平均值。对于例4.1中的汉明失真， $d(s, s^{\wedge})$ 则是再现消息 s^{\wedge} 中“错误”的比例。

拥有一个低速率和小失真的信源编码器/译码器对无疑是理想的。由马尔可夫链 $S \leftrightarrow W \leftrightarrow S^{\wedge}$ 导出，失真 $d(S, S^{\wedge})$ 是一个随机变量。这一考虑导致了以下可达率失真对的定义。

定义 4.1 如果存在一个信源编码器/译码器对的序列，

4.1 问题描述 45

--- 第 4 页 ---

46

4 信源表示：率失真特性

图 4.2：率失真域和 $R(D)$ 函数示意图。图中横轴为失真 D ，纵轴为速率 R 。阴影区域为率失真域，其左上边界曲线为 $R(D)$ 函数。

图4.2：率失真域和 $R(D)$ 函数示意图。

$\{(f_n(s), g_n(s))\}_{n=1,2,\dots}$ 满足

$$R \geq n \lceil \log_2 M^n \rceil \text{ 比特/信源符号, (4.6)}$$

$$\lim_{n \rightarrow \infty} E[d(S, S^{\wedge})] \leq D. \quad (4.7)$$

则称一个率失真对 (R, D) 是**可达的**。

取 (R,D) 平面第一象限中可达率失真对集合的闭包 $*$ ，我们得到**率失真域**。因此，率失真域的边界引出了**率失真函数**及其逆函数**失真率函数**的概念。见图 4.2 的示意图。

定义 4.2 率失真函数 $R(D)$ 是对于每个给定的 D ，使得 (R,D) 在率失真域中的速率的下确界；而**失真率函数** $D(R)$ 是对于每个给定的 R ，使得 (R,D) 在率失真域中的失真的下确界。

值得注意的是，率失真域是在渐近意义上定义的：信源编码器/译码器对在消息大小 $n \rightarrow \infty$ 时对任意大的消息都有效。采用渐近分析是信息论的一个关键，并且在渐近状态下常常会出现奇观。

4.2 香农信源编码基本定理

在本节中，我们将介绍一个基本结果，该结果首次出现在香农的文章[16]中，它描述了率失真函数。

- 对我们的目的而言，将一个集合的闭包理解为该集合与其边界的并集就足够了。

--- 第 5 页 ---

4.2 香农信源编码基本定理 47

让我们定义信息率失真函数如下。

定义 4.3 对于一个概率质量函数为 $PS(s)$ 的离散无记忆信源 S 和失真度量 $d(s,s^{\wedge})$, $(s,s^{\wedge}) \in S \times S^{\wedge}$ ，以下约束优化问题的解：

$$\min_{PS^{\wedge}} I(S;S^{\wedge}), \quad (4.8)$$

$$\text{s.t. } E[d(S,S^{\wedge})] \leq D \quad (4.9)$$

被称为**信息率失真函数** $RI(D)$ 。

香农信源编码基本定理断言信息率失真函数等于率失真函数。

定理 4.1 对于一个概率质量函数为 $PS(s)$ 的离散无记忆信源 S 和有界失真度量 $d(s,s^{\wedge})$, $(s,s^{\wedge}) \in S \times S^{\wedge}$ ，我们有

$$R(D) = RI(D) \quad (4.10)$$

请注意，定理4.1中的率失真函数不是以闭合形式给出的。它涉及一个优化问题 (4.8)-(4.9)。目标函数是信源 S 与其再现 S^{\wedge} 之间的互信息 $I(S;S^{\wedge})$ 。由于 $PS, S^{\wedge}(s,s^{\wedge}) = PS(s)PS^{\wedge}(s^{\wedge} | s)$ 且 PS 是给定的，最小化 $I(S;S^{\wedge})$ 是相对于 $PS^{\wedge} | S$ 进行的。约束 $E[d(S,S^{\wedge})] \leq D$ 对 $PS^{\wedge} | S$ 的可行选择施加了限制。

根据定义4.2，率失真函数 $R(D)$ 具有编码解释，即在一个足够长的信源消息块上，受期望失真约束 D 的一系列编码器/译码器对所能达到的速率的下确界。另一方面，由定义4.3给出的信息率失真函数 $RI(D)$ 只是一个数学对象，一个依赖于 PS 和 d 的函数，没有任何编码解释。是定理4.1将这两者联系起来，断言它们实际上是相等的。有了定理4.1，当我们被要求求解率失真函数时，我们可以转而评估信息率失真函数。定理4.1的证明将在接下来的两节中提供。在本节的剩余部分，我们建立信息率失真函数 $RI(D)$ 的一些有用性质，并提供一些简单的说明性例子来计算它。

--- 第 6 页 ---

48

4 信源表示：率失真特性

首先，由于约束 (4.9) 的左侧对于任何 $PS^{\wedge} | S$ 都满足

$$E[d(S,S^{\wedge})] \geq D_{\min} := \sum s \in S PS(s) \min_{s^{\wedge} \in S^{\wedge}} d(s,s^{\wedge}), \quad (4.11)$$

因此对于 $D < D_{\min}$ ，优化问题 (4.8) 是不可行的，我们可以在该区域简单地定义 $RI(D) = \infty$ 。

其次，请注意 $I(S;S^{\wedge}) = 0$ 当且仅当 S 和 S^{\wedge} 是独立的；见推论 3.3。因此，如果存在一个概率质量函数 PS^{\wedge} 使得

$$\sum_{(s,s^*) \in S \times S^*} PS(s)PS^*(s^*)d(s,s^*) \leq D \quad (4.12)$$

成立，我们可以在优化问题 (4.8) 中简单地令 $PS^* | S = PS^*$ ，从而得到 $RI(D)=0$ 。由于总是成立

$$\sum_{(s,s^*) \in S \times S^*} PS(s)PS^*(s^*)d(s,s^*) \geq D_{\max} := \min_{s^* \in S^*} \sum_{s \in S} PS(s)d(s,s^*) \quad (4.13)$$

对于 $RI(D)=0$ ，必须有 $D \geq D_{\max}$ 。另一方面，这个条件也是充分的。为了看到这一点，将达到 D_{\max} 的 s^* 的值记为 s^{*} ，并让 $S^*=s^{*}$ 的概率为1。那么我们有 $I(S;S^*)=0$ ，并且

$$E[d(S,S^*)] = E[d(S,s^{*})] = \sum_{s \in S} PS(s)d(s,s^{*}) = D_{\max} \quad (4.14)$$

对于任何 PS 。因此，对于任何 $D \geq D_{\max}$ ，这种确定性选择 $S^*=s^{*}$ 是可行的，导致 $RI(D)=0$ 。总而言之，我们有以下关于 $RI(D)$ 为零的充要条件。

推论 4.1 信息率失真函数 $RI(D)$ 为零当且仅当

$$D \geq D_{\max} = \min_{s^* \in S^*} \sum_{s \in S} PS(s)d(s,s^*) \quad (4.15)$$

$RI(D)$ 的一些进一步特征在下一个推论中给出。

推论 4.2 信息率失真函数 $RI(D)$ 是一个关于 D 的非增凸函数。如果 $RI(D_{\min}) > 0$ ，那么 $RI(D)$ 在 $D \in [D_{\min}, D_{\max}]$ 上是严格递减的，并且不等式约束 (4.9) 可以被一个等式约束替换。

证明： 考虑任何不小于 D_{\min} 的 $D_0 < D_1$ 。将满足 $E[d(S,S^*)] \leq D$ 的 $PS^* | S$ 的集合记为 PD 。那么很明显 $PD_0 \subseteq PD_1$ ，因此 $RI(D_0) \geq RI(D_1)$ 。这证明了 $RI(D)$ 是关于 D 的非增函数。

--- 第 7 页 ---

4.2 香农信源编码基本定理 49

对于凸性，我们需要证明对于任何不小于 D_{\min} 的 $D_0 < D_1$ 和任何 $0 \leq \lambda \leq 1$ ，

$$RI((1-\lambda)D_0 + \lambda D_1) \leq (1-\lambda)RI(D_0) + \lambda RI(D_1) \quad (4.16)$$

成立。

用 $PS^* | S(i), i=0,1$ 表示达到 $RI(D_i)$ 的条件概率分布。

定义

$$PS^* | S(\lambda)(s^* | s) = (1-\lambda)PS^* | S(0)(s^* | s) + \lambda PS^* | S(1)(s^* | s) \quad (4.17)$$

注意 $PS^* | S(\lambda)$ 对于求解 $RI((1-\lambda)D_0 + \lambda D_1)$ 的优化问题是可行的，因为在 $PS^* | S(\lambda)$ 下

$$E[d(S,S^*)] = (1-\lambda) \sum_{(s,s^*) \in S \times S^*} PS(s)PS^* | S(0)(s^* | s)d(s,s^*) + \lambda \sum_{(s,s^*) \in S \times S^*} PS(s)PS^* | S(1)(s^* | s)d(s,s^*) \leq (1-\lambda)D_0 + \lambda D_1 \quad (4.18)$$

其中不等式来自假设 $PS^* | S(0)$ 和 $PS^* | S(1)$ 分别达到 $RI(D_0)$ 和 $RI(D_1)$ 。因此，根据约束优化问题公式 (4.8)-(4.9)，我们有

$$RI((1-\lambda)D_0 + \lambda D_1) \leq I(\lambda)(S;S^*) \quad (4.19)$$

这是通过令条件概率分布 $PS^* | S$ 为 $PS^* | S(\lambda)$ 所达到的互信息。

另一方面，根据定理 3.7， $I(S;S^*)$ 在固定 PS 的情况下是关于 $PS^* | S$ 的凸函数。所以在 $PS^* | S(\lambda)$ 下，我们有

$$I(\lambda)(S;S^*) \leq (1-\lambda)I(0)(S;S^*) + \lambda I(1)(S;S^*) = (1-\lambda)RI(D_0) + \lambda RI(D_1) \quad (4.20)$$

将 (4.19) 和 (4.20) 结合起来，我们得到 (4.16) 并证明了 $RI(D)$ 的凸性。

我们然后假设 $R(D_{\min}) > 0$ 并证明 $R(D)$ 在 $[D_{\min}, D_{\max}]$ 上是严格递减的。假设情况并非如此。因为我们已经证明 $R(D)$ 是非增的，唯一的可能性是存在一些“平台”，即 $D_{\min} \leq D_0 < D_1 \leq D_{\max}$ 且对于所有 $D \in [D_0, D_1]$ 都有 $R(D) = R(D_0)$ 。见图 4.3 的示意图。因为我们已经证明 $R(D)$ 是

--- 第 8 页 ---

50

4 信源表示：率失真特性

凸的，通过重写

$$D_1 = D_{\max} - D_0 D_{\max} - D_1 D_0 + D_{\max} - D_0 D_1 - D_0 D_{\max}, \quad (4.21)$$

我们有

$$R(D_1) \leq D_{\max} - D_0 D_{\max} - D_1 R(D_0) + D_{\max} - D_0 D_1 - D_0 R(D_{\max}). \quad (4.22)$$

注意到根据我们的假设， $R(D_1) = R(D_0)$ 成立，并且对于所有 $D \geq D_{\max}$ ， $R(D) = 0$ 成立，我们因此得出唯一的可能性是 $R(D_0) = 0$ ，从而对于所有 $D \geq D_0$ ， $R(D) = 0$ 。但根据推论 4.1，对于 $R(D) = 0$ ，必须有 $D \geq D_{\max}$ 。这因此导致了一个矛盾，所以 $R(D)$ 必须在 $[D_{\min}, D_{\max}]$ 上是严格递减的。

图 4.3: $R(D)$ 在 $[D_{\min}, D_{\max}]$ 上严格递减的示意图。图中展示了一个函数 $R(D)$ ，它在 D_0 和 D_1 之间有一个平台（水平线），这与函数的凸性相矛盾（虚线表示了凸性应有的走势）。

图 4.3: $R(D)$ 在 $[D_{\min}, D_{\max}]$ 上严格递减的示意图。

为了证明我们最后的论断，即不等式约束 (4.9) 可以被等式约束所取代，让我们假设存在某个 $D \in [D_{\min}, D_{\max}]$ ，使得在达到 $R(D)$ 时，相应的 $P_{S^* | S}$ 满足一个严格不等式 $E[d(S, S^*)] = D' < D$ 。这便导致 $R(D') = R(D)$ ，这与我们刚刚证明的 $R(D)$ 在 $[D_{\min}, D_{\max}]$ 上的严格递减性质相矛盾。

例 4.2 在这个例子中，我们计算一个服从参数为 δ 的伯努利分布的信源 S 在汉明失真下的信息率失真函数。

让我们从寻找 D_{\min} 和 D_{\max} 开始。很明显 $D_{\min} = 0$ 。根据定义 $D_{\max} = \min_{S^* \in \mathcal{S}^*} \sum_{s \in \mathcal{S}} P_S(s) d(s, s^*)$ ，我们有

$$D_{\max} = \min_{s^* \in \{0, 1\}} [(1-\delta)d(0, s^*) + \delta d(1, s^*)] = \min\{\delta, 1-\delta\}. \quad (4.23)$$

为确定起见，假设 $\delta \leq 1/2$ 。因此 $D_{\max} = \delta$ 且对于所有 $D \geq \delta$ ， $R(D) = 0$ 。 $\delta > 1/2$ 的另一种情况可以类似处理，留给读者。

--- 第 9 页 ---

4.2 香农信源编码基本定理 51

现在让我们研究对于 $D \in [0, \delta]$ 的约束优化问题 (4.8)-(4.9)。为此，我们根据

$$I(S; S^*) = H(S) - H(S | S^*), \quad (4.24)$$

来展开 $I(S; S^*)$ ，其中 $H(S) = h_2(\delta)$ 是给定的，并且

$$H(S | S^*) = (a) H(S \oplus S^* | S^*) \leq H(S \oplus S^*) \quad (4.25)$$

其中 (a) 是由于推论 3.6，因为在给定 S^* 的条件下， S 和 $S \oplus S^*$ 之间存在双射关系，而 (b) 是由于定理 3.6，“条件作用降低熵”。

但 $S \oplus S^*$ 是什么？实际上， $\{S \oplus S^* = 1\}$ 等价于 $\{S \neq S^*\}$ ，在汉明失真下，这又等价于 $\{d(S, S^*) = 1\}$ 。所以约束 (4.9)（由于推论 4.2，现在是一个等式约束）等价于 $P(S \oplus S^* = 1) = D$ 。因此，我们得到了 $I(S; S^*)$ 的一个下界

$$I(S; S^*) \geq h_2(\delta) - h_2(D) \quad (4.26)$$

这对任何 $PS^\wedge | S$ 都成立。

如果我们能找到某个 $PS^\wedge | S$ 来达到下界 (4.26)，那么这个下界就恰好是 $RI(D)$ 。检查到目前为止的推导步骤，我们看到所寻求的 $PS^\wedge | S$ 应该在 (4.25) 的 (b) 中达到等式；也就是说，它应该使得 S^\wedge 和 $S \oplus S^\wedge$ 独立。受到 $S = S^\wedge \oplus (S \oplus S^\wedge)$ 这一事实的启发，我们将 $S \oplus S^\wedge$ 视为一个与 S^\wedge 无关的辅助随机变量 Z ，根据上一段的讨论，它是一个参数为 D 的伯努利变量。注意，这指定了“后向”条件概率分布 $PS | S^\wedge$ ，而不是约束优化问题 (4.8)-(4.9) 中所需的“前向”条件概率分布 $PS^\wedge | S$ 。由于 $PS^\wedge | S = PS^\wedge PS | S^\wedge$ ，为了指定 $PS^\wedge | S$ ，我们还需要找到相应的 PS^\wedge 。经过一些计算，不难验证这样的 PS^\wedge 存在，并且由一个参数为 $(\delta - D)/(1 - 2D)$ 的伯努利分布给出。

因此，下界 (4.26) 是可达的，对于参数为 $\delta \leq 1/2$ 的伯努利信源和汉明失真，信息率失真函数由下式给出：

$$RI(D) = h_2(\delta) - h_2(D) \text{ 如果 } 0 \leq D \leq \delta, \text{ 否则为 } 0, \quad (4.27)$$

如图 4.4 所示。

在例 4.2 中，首先指定 $PS | S^\wedge$ 然后找到相应的 PS^\wedge 来实现信息率失真函数的过程是一种有用的技术。这样构造的 $PS | S^\wedge$ 通常

--- 第 10 页 ---

52

4 信源表示：率失真特性

图 4.4：带汉明失真的伯努利离散无记忆信源的率失真函数。图中显示了两条曲线，分别对应 $\delta=0.1$ 和 $\delta=0.5$ 。两条曲线都是从 $R(0) > 0$ 开始，随着 D 的增加而单调递减，直到 $D=\delta$ 时 $R(D)$ 变为 0。

图4.4：带汉明失真的伯努利DMS的率失真函数。

被称为“测试信道”。回想起来，因为我们是从 $I(S; S^\wedge) = H(S) - H(S | S^\wedge)$ 开始的，所以很自然地要考虑构造合适的 $PS | S^\wedge$ 和 PS^\wedge ，以便在满足失真约束和对 S 的概率分布的约束下，最大化 $H(S | S^\wedge)$ 。

应该认识到，除了一些具有高度对称结构的极少数 PS 和 d （例如例 4.2 和本讲中的一些练习），测试信道的显式构造是困难的，我们不能期望信息率失真函数通常能得到闭合形式的解。下一个例子通过仅轻微修改例 4.2 的设置来说明这一点。

例 4.3 我们考虑一个服从参数为 $1/2$ 的伯努利分布的信源 S ，具有以下非对称失真度量：

$$d(0,0)=d(1,1)=0, d(0,1)=1, \text{ 并且 } d(1,0)=A > 1 \quad (4.28)$$

对于此设置， $D_{\min}=0$ ， $D_{\max}=1/2$ 。我们然后关注 $D \in [0, 1/2]$ 。根据推论 4.2，约束 (4.9) 是一个等式

$$E[d(S, S^\wedge)] = 21PS^\wedge | S(1 | 0) + 2APS^\wedge | S(0 | 1) = D. \quad (4.29)$$

我们然后令 $\alpha = PS^\wedge | S(1 | 0) \in [0, 1]$ 和 $\beta = PS^\wedge | S(0 | 1) \in [0, 1]$ 。利用它们，我们可以评估 $I(S; S^\wedge)$ 得到

$$I(S; S^\wedge) = H(S) - H(S | S^\wedge) = \log 2 - [21 - \alpha + \beta h_2(1 - \alpha + \beta\beta) + 21 + \alpha - \beta h_2(1 + \alpha - \beta\alpha)] \quad (4.30)$$

这应该在 $(\alpha, \beta) \in [0, 1]^2$ 上，在 $\alpha + A\beta = 2D$ 的约束下最小化。不幸的是，这个优化问题似乎没有

--- 第 11 页 ---

4.3 逆定理部分的证明 53

一个封闭形式的解，只能通过数值方法来解决。图 4.5 描绘了 $A=1.5$ 和 3 时的 $RI(D)$ 。它还包括了 (4.27) 作为参考，对应于 $A=1$ 的情况。

图 4.5: 具有非对称失真 (4.28) 的伯努利(1/2)离散无记忆信源的率失真函数。图中显示了 $A=1.0$ (虚线)、 $A=1.5$ (实线) 和 $A=3.0$ (实线) 三种情况下的 $R(D)$ 曲线。所有曲线都从 $D=0$ 时的 $R(D)=1$ 开始, 随 D 增加而下降, 在 $D=0.5$ 时达到0。 A 值越大, 曲线在中间部分越高。

图 4.5: 具有非对称失真 (4.28) 的伯努利(1/2) DMS 的率失真函数。

尽管如此, 存在用于计算信息率失真函数的有效数值算法, 我们将在第9讲中介绍它们。

4.3 逆定理部分的证明

在本节中, 我们建立定理4.1的逆定理部分, 即 $R(D) \geq R_I(D)$ 。为此, 我们将证明, 对于任何信源编码器/译码器对, 对于任何 $R < R_I(D)$, 都不可能达到不大于 D 的期望失真。

让我们固定一个任意的信源编码器和译码器对 $(f_n(s), g_n(s))$, 信源消息长度为 n 。由于信源消息 S 是一个随机向量, $(f_n(s), g_n(s))$ 在 (S, S^\wedge) 上导出一个联合概率分布。假设所产生的期望失真满足 $E[d(S, S^\wedge)] \leq D$ 。我们现在考察这样一个信源编码器/译码器对的速率。我们按以下步骤进行:

$$\begin{aligned} nR &\geq (a)H(W) \geq (b)I(S; W) \geq I(S; S^\wedge) = H(S) - H(S | S^\wedge) = (c) \sum_{i=1}^n H(S_i) - \sum_{i=1}^n H(S_i | S^\wedge, S_{i-1}, \dots, S_1) \geq (d) \sum_{i=1}^n H(S_i) \\ &\quad - \sum_{i=1}^n H(S_i | S^\wedge_i) = \sum_{i=1}^n I(S_i; S^\wedge_i), \quad (4.31) \end{aligned}$$

--- 第 12 页 ---

54

4 信源表示: 率失真特性

其中, (a) 是因为根据 (4.1), W 是 M^n 个索引之一, 因此根据推论 3.1, 其熵最多为 $\log_2 M^n \leq nR$ 比特, 根据定义 4.1, (b) 是由于定理 3.5, 即 DPI, (c) 是由于定理 3.1, 即熵的链式法则, 而 (d) 是因为条件作用降低熵 (见定理 3.6)。

为了进一步进行, 我们引入一个辅助的“分时”随机变量 Q , 它在 $\{1, 2, \dots, n\}$ 上均匀分布, 并且与 S 无关。因此, 我们可以继续前面的界定步骤:

$$R \geq n^{-1} \sum_{i=1}^n I(S_i; S^\wedge_i) = (e) I(SQ; S^\wedge Q | Q) = H(SQ | Q) - H(SQ | S^\wedge Q, Q) \quad (4.32)$$

其中 (e) 正是条件互信息的定义 (见定义 2.15), 注意到对于所有 $1 \leq i \leq n$, $P_Q(i) = 1/n$ 。由于 DMS S 的概率分布不依赖于下标 Q , 我们有 $H(SQ | Q) = H(SQ)$ 。另一方面, 因为条件作用降低熵, $H(SQ | S^\wedge Q, Q) \leq H(SQ | S^\wedge Q)$ 。所以我们得到

$$R \geq H(SQ) - H(SQ | S^\wedge Q) = I(SQ; S^\wedge Q). \quad (4.33)$$

现在, 注意到 SQ 的概率分布就是 PS , 根据 $R_I(D)$ 的约束优化问题表述 (4.8)-(4.9), 我们可以更进一步得到

$$R \geq I(SQ; S^\wedge Q) \geq R_I(E[d(SQ, S^\wedge Q)]). \quad (4.34)$$

由于 Q 在 $\{1, 2, \dots, n\}$ 上是均匀的, 使用全期望定律 (见定理 2.3), 我们有

$$E[d(SQ, S^\wedge Q)] = n^{-1} \sum_{i=1}^n E[d(S_i, S^\wedge_i)] = (f) E[d(S, S^\wedge)] \leq (g) D, \quad (4.35)$$

其中 (f) 是由于失真的可加结构, 即 (4.5), 而不等式 (g) 则源于证明开始时对 $(f_n(s), g_n(s))$ 的假设。我们现在得出结论:

$$R \geq R_I(E[d(SQ, S^\wedge Q)]) \geq R_I(D) \quad (4.36)$$

--- 第 13 页 ---

4.4 可达性部分的证明 55

通过使用 $R_I(D)$ 的非增性质 (见推论 4.2)。这就完成了逆定理部分的证明。

注 4.3 检查证明过程，我们发现关于编码和译码的唯一要求是马尔可夫链 $S \leftrightarrow W \leftrightarrow S^\wedge$ 。我们不要求 W 在给定 S 的条件下是确定性的，也不要求 S^\wedge 在给定 W 的条件下是确定性的。也就是说，即使信源编码器和译码器是随机的，逆定理仍然成立。

4.4 可达性部分的证明

在本节中，我们建立定理 4.1 的可达性部分，即 $R(D) \leq R_I(D)$ 。为此，我们将证明，给定 D ，对于任何速率 $R > R_I(D)$ ，存在一个信源编码器/译码器对的序列，以信源消息长度 $n=1,2,\dots$ 为索引，使得 (R, D) 是一个可达的率失真对，其意义如定义 4.1 所述。

在开始之前，让我们把信源编码器/译码器对的概念具体化。注意到代表 S 的索引 W 是从 M_n 个索引中选出的 $f_n(s)$ (S)，并且每个 $W=w$ 对应一个再现的消息向量 $g_n(s)(w)$ 。我们称 $\{g_n(s)(w), w=1,\dots,M_n\}$ 为**码本**，并用 C 表示。 C 的每一项称为一个**码字**，第 w 个码字，即 $g_n(s)(w)$ ，也表示为 $C(w)$ ，其第 i 个元素进一步表示为 $C_i(w)$ ，对于 $i=1,\dots,n$ 。

编码器 $f_n(s)$ 指定了一个规则，将每个可能的信源消息向量分配给 C 中的一个码字，而译码器 $g_n(s)$ 只是输出这个分配的码字作为再现消息。 C 的维度参数 n 和 M_n 决定了速率 R 为 $\lceil \log M_n \rceil / n$ ；见 (4.4)。关于期望失真，我们有

$$E[d(S, S^\wedge)] = \sum_{s \in S} p_S(s) d(s, C(f_n(s)(s))) \quad (4.37)$$

在这一点上，对于一个特定的 $(f_n(s), g_n(s))$ 对，通常很难对 $E[d(S, S^\wedge)]$ 说些什么。两个关键思想使得进一步的分析成为可能。首先，让码本本身是随机生成的；其次，在分析期望失真时，让期望同时对信源消息和码本进行。

--- 第 14 页 ---

56

4 信源表示：率失真特性

4.4.1 码本的生成

我们固定某个满足 $E[d(S, S^\wedge)] \leq D$ 的 $PS^\wedge \mid S$ ，并固定一个速率 $R > I(S; S^\wedge)$ ，该互信息是使用 $PS^\wedge \mid S$ 计算的。 S^\wedge 的边缘概率分布是 $PS^\wedge(s^\wedge) = \sum_{s \in S} p_S(s) PS^\wedge(s^\wedge \mid s)$ 。如前所述，我们让码本是随机生成的。具体来说，我们让所有码字的所有元素都是服从 PS^\wedge 的独立同分布 (i.i.d.) 随机变量。我们将这样一个随机码本表示为 C ，以区别于它的实现 C 。

请注意，这样一个随机码本 C 与信源消息 S 是独立的。一旦 C 生成，它就会被告知编码器和译码器。

将 S^\wedge^n 视为一个空间，其中每个码字都是一个点。 C 的生成对应于根据 $\prod_{i=1}^n PS^\wedge_i$ 从 S^\wedge^n 中独立有放回地抽样 M_n 次，其中 $\{S^\wedge_i, i=1,\dots,n\}$ 是服从 PS^\wedge 的 i.i.d. 变量。

每个码字负责再现信源空间 S_n 中的一些点，而 S_n 中的每个点都需要由一个码字来再现。因此，情况是 S_n 的空间应被划分为 M_n 个不重叠的子集，并且每个这样的子集都应与给定码本 $C \subseteq S^\wedge^n$ 中的一个不同码字相关联。从直觉上可以清楚地看出，如果码本的大小 M_n 很小，并且 S^\wedge^n 只是被稀疏地采样，那么每个码字就需要再现 S_n 的一个“大”子集中的点，从而导致大的失真。因此，码本应包含足够多的码字，这一考虑转化为对最小速率 R 的要求。

4.4.2 编码

给定一个码本 C ，我们应该如何对一个信源消息 s 进行编码？显然，最优的编码方式应该将 s 分配给能最小化 s 与码字之间失真的码字；也就是说，选择

$$w^* = \operatorname{argmin}_{w=1,\dots,M_n} d(s, C(w)), \quad (4.38)$$

并令 $s^\wedge = C(w^*)$ 。这本质上是一种“聚类”操作，它将 S_n 中的点围绕它们各自“最近”的码字进行聚类。然而，最优编码过程的性能分析起来有些不便。因此，我们转向另一种编码过程，它通常不是最优的，但仍然足以完成我们的可达性证明。为了激发

--- 第 15 页 ---

4.4 可达性部分的证明 57

编码过程，让我们进行以下思想实验。如果 S 和 $C(W)$ 是根据 $PS, S^\wedge = \prod_{i=1}^n PS_i, S^\wedge_i$ 联合分布的，其中 $\{(S_i, S^\wedge_i), i=1, \dots, n\}$ 是 i.i.d. 且服从 PS, S^\wedge ，那么根据弱大数定律（见定理 2.4），当 n 无限增长时，以高概率 $d(S, C(W))$ 将会接近其期望 $E[d(S, S^\wedge)]$ ，这个期望在我们的证明开始时已经固定为不大于 D 。然而，现实并非如此，因为码本 C 是独立于 S 生成的。因此，我们希望编码器（在可能的情况下）选择某个码字，使得信源消息和所选码字的对表现得好像它们确实服从联合概率分布 PS, S^\wedge 。

受前面讨论揭示的关键洞察的启发，我们设计了以下编码过程：在 $\epsilon > 0$ 的条件下，给定 $S=s$ ，找到最小的索引 $w \in \{1, 2, \dots, M_n\}$ 使得

$$d(s, C(w)) \leq D + \epsilon, \quad (4.39) \quad |I(S; S^\wedge) - \epsilon \leq I(s; C(w)) \leq I(S; S^\wedge) + \epsilon, \quad (4.40)$$

其中

$$d(s, C(w)) = n \sum_{i=1}^n d(s_i, C_i(w)), \quad (4.41) \quad I(s; C(w)) = n \log_2 \frac{PS(s) PS^\wedge(C(w))}{PS^\wedge(s, C(w))}$$

$$= n \sum_{i=1}^n \log_2 \frac{PS(s_i) PS^\wedge(C_i(w))}{PS^\wedge(s_i, C_i(w))}. \quad (4.42)$$

如果存在这样的索引 w ，我们将其表示为 w^* 并使用对应的码字 $C(w^*)$ 来再现 s ；否则，编码过程失败，我们简单地设置 $w^*=1$ 并使用 $C(1)$ 来再现 S 。

标准 (4.39) 很容易理解。但我们为什么需要另一个标准 (4.40)？正如我们在思想实验中刚刚讨论的，我们希望编码器（在可能的情况下）选择某个码字，使得信源消息和所选码字的对表现得好像它们服从联合概率分布 PS, S 。鉴于弱大数定律，标准 (4.40) 有效地强制了这一点。正如接下来将看到的，标准 (4.40) 促进了期望失真分析中的一个关键步骤。为方便起见，对于给定的 C ，我们为每个 $s \in S_n$ 关联一个指示函数 $E(C, s)$ 来指示是否发生编码失败；也就是说，如果存在某个索引 w 同时满足 (4.39) 和 (4.40)，则 $E(C, s)=0$ ，否则 $E(C, s)=1$ 。

--- 第 16 页 ---

58

4 信源表示：率失真特性

4.4.3 期望失真分析

对于给定的码本 C ，期望失真 $E[d(S, S^\wedge)]$ 仍然难以评估。如前所述，除了随机生成码本外，一个密切相关的关键思想是研究码本集合上的期望失真。现在让我们考察期望失真，这个期望是对 S 和 C 两者同时取的：

$$E_C S[d(S, S^\wedge)] = \sum_{C, s \in S_n} PC(C) PS(s) d(s, C(w)) = \sum_{C, s \in S_n} E(C, s) PC(C) PS(s) d(s, C(w))$$

$$\bullet \sum_{C, s \in S_n} E(C, s) PC(C) PS(s) d(s, C(1)). \quad (4.43)$$

对于 (4.43) 中的第一个求和，由于每一项都满足 (4.39)，所以 $d(s, C(w)) \leq D + \epsilon$ 成立，因此该求和也被 $D + \epsilon$ 上界。对于 (4.43) 中的第二个求和，无论如何我们都可以用 d_{\max} 来上界 $d(s, C(1))$ ，因此该求和被 $P_f d_{\max}$ 上界，其中

$$P_f = \sum_{C, s \in S_n} E(C, s) PC(C) PS(s) \quad (4.44)$$

是编码失败的概率。

4.4.4 编码失败概率的估计

现在让我们估计可以重写为

P_f

$$P_f = \sum_{C, s \in S_n} PC(C) PS(s) E(C, s) = \sum_{s \in S_n} PS(s) \sum_C PC(C) E(C, s) \quad (4.45)$$

一个关键步骤是看到，对于任何 $s \in S_n$ ，

$$\sum_{C \in \mathcal{C}} P(C) E(C, s) = [1 - P((s, s^\wedge) \text{ 满足 (4.39)(4.40)})] M_n. \quad (4.46)$$

这是因为 $E(C, s) = 1$ 意味着 C 的 M_n 个码字中没有同时满足 (4.39) 和 (4.40)，并且这些码字是 i.i.d. 的，概率分布为 PS^\wedge 。在这一点上，我们看到了对 C 取期望的好处： P_f 的评估现在只涉及 PS, S^\wedge ，而不需要任何特定的码本。

--- 第 17 页 ---

4.4 可达性部分的证明 59

为了进一步进行，让我们引入另一个指示函数 $E_{\sim}(s, s^\wedge)$ 如下：如果 (s, s^\wedge) 同时满足 (4.39) 和 (4.40)，则 $E_{\sim}(s, s^\wedge) = 1$ ，否则 $E_{\sim}(s, s^\wedge) = 0$ 。所以

$$P((s, s^\wedge) \text{ 满足 (4.39)(4.40)}) = \sum_{s^\wedge \in S^\wedge} P(S^\wedge = s^\wedge) E_{\sim}(s, s^\wedge). \quad (4.47)$$

现在标准 (4.40) 发挥作用，从中我们有 $i(s; s^\wedge) \leq I(S; S^\wedge) + \epsilon$ ，这等价于

$$P(S^\wedge = s^\wedge) \geq 2^{-n(I(S; S^\wedge) + \epsilon)} P(S^\wedge = s^\wedge | s). \quad (4.48)$$

因此我们得到

$$P((s, s^\wedge) \text{ 满足 (4.39)(4.40)}) = \sum_{s^\wedge \in S^\wedge} P(S^\wedge = s^\wedge) E_{\sim}(s, s^\wedge) \geq 2^{-n(I(S; S^\wedge) + \epsilon)} \sum_{s^\wedge \in S^\wedge} P(S^\wedge = s^\wedge | s) E_{\sim}(s, s^\wedge). \quad (4.49)$$

将 (4.49) 代入 (4.45)，我们有

$$P_f \leq \sum_{s \in S} P(S = s) [1 - 2^{-n(I(S; S^\wedge) + \epsilon)} \sum_{s^\wedge \in S^\wedge} P(S^\wedge = s^\wedge | s) E_{\sim}(s, s^\wedge)] M_n \quad (4.50)$$

此时，我们应用不等式 $(1 - xy)^n \leq 1 - x + e^{-ny}$ ，对所有 $0 \leq x, y \leq 1$ 成立，得到

$$P_f \leq \sum_{s \in S} P(S = s) [1 - \sum_{s^\wedge \in S^\wedge} P(S^\wedge = s^\wedge | s) E_{\sim}(s, s^\wedge) + e^{-M_n 2^{-n(I(S; S^\wedge) + \epsilon)}}] = 1 - \sum_{(s, s^\wedge) \in S \times S^\wedge} P(S, s^\wedge) E_{\sim}(s, s^\wedge) + e^{-M_n 2^{-n(I(S; S^\wedge) + \epsilon)}}. \quad (4.51)$$

在 (4.51) 中，第二项正是 $E_{\sim}(S, S^\wedge)$ 在 PS, S^\wedge 下的期望，或者说是 (S, S^\wedge) 在 PS, S^\wedge 下同时满足 (4.39) 和 (4.40) 的概率。由于弱大数定律（见定理 2.4），对于任何 $\epsilon > 0$ ，当 $n \rightarrow \infty$ 时，这个概率趋近于 1。

在 (4.51) 中，第三项由 $M_n 2^{-n(I(S; S^\wedge) + \epsilon)}$ 随 n 无限增长时的渐近行为控制。忽略 M_n 中可能存在的对于大 n 变得无关紧要的舍入效应，我们有

$$M_n 2^{-n(I(S; S^\wedge) + \epsilon)} = 2^{nR} 2^{-n(I(S; S^\wedge) + \epsilon)} = 2^{n(R - I(S; S^\wedge) - \epsilon)} \quad (4.52)$$

由于我们在证明开始时固定了 $R > I(S; S^\wedge)$ ，我们总可以

--- 第 18 页 ---

60

4 信源表示：率失真特性

找到一个足够小的 $\epsilon > 0$ 使得 $R - I(S; S^\wedge) - \epsilon > 0$ ，从而 (4.51) 中的第三项当 $n \rightarrow \infty$ 时收敛到零。

总而言之，通过考虑随机码本 C 的集合，我们已经证明了对于任何足够小的 $\epsilon > 0$ ，当 $n \rightarrow \infty$ 时，编码失败的概率 $P_f \rightarrow 0$ 。

4.4.5 总结步骤

回到期望失真的分析，我们看到对于任何 $\epsilon > 0$ ，当 $n \rightarrow \infty$ 时， $E_C S[d(S, S^\wedge)] \leq D + 2\epsilon$ 成立。这意味着存在至少一个以 n 为索引的码本序列，对于这个序列，当 $n \rightarrow \infty$ 时， $E[d(S, S^\wedge)] \leq D + 2\epsilon$ 成立。[†] 让 $\epsilon \rightarrow 0$ ，我们便可以断言，对于我们选择的 $PS^\wedge | S$ ，任何 $R > I(S; S^\wedge)$ 和 D 构成一个可达的率失真对。通过选择能达到信息率失真函数的 $PS^\wedge | S$ ，即 $I(S; S^\wedge) = R_I(D)$ ，我们完成了可达性部分的证明。

注 4.4 回顾一下，唯一需要标准 (4.40) 的地方是对 P_f 的分析，从 (4.48) 开始。如果编码过程忽略 (4.40)，相应的 P_f 将会更小。但是包含 (4.40) 对于证明当 $n \rightarrow \infty$ 时 $P_f \rightarrow 0$ 是至关重要的。

注释

在满足失真约束条件下表示信源的基本思想出现在香农1948年的论文[1, 第五部分]中，并在他1959年发表的著作[16]中得到系统发展，该著作奠定了量化和有损压缩的基础。量化和有损压缩在信号（例如，图像、视频、音频、语音、生物医学、金融等）处理中得到了广泛应用。在信息论教学中，将率失真理论放在如此靠前的位置，甚至在信道传输之前，有点不寻常，但我们这样做是考虑到从香农的一般通信系统模型中的信源部分开始似乎更自然（见第1讲图1.1）。我们对香农信源编码基本定理可达性部分的证明是基于Robert J. McEliece的教科书[17]，并且没有明确依赖于许多其他教科书（例如，[18][8]）中的“标准”方法——典型性语言。我们对典型性工具的介绍将推迟到第8讲。

一些教科书（例如，[8][9]）采用了一个不同的可达率失真对定义，通过用数学上更强的超额失真约束 $\lim_{n \rightarrow \infty} P(d(S, S^n) \geq D) = 0$ 来替换期望失真约束 $\lim_{n \rightarrow \infty} E[d(S, S^n)] \leq D$ 。

一本专门讨论率失真理论的经典参考文献是Toby Berger的[19]。读者也被邀请阅读他与Jerry D. Gibson合著的更近期的综述[20]，其中提供了率失真理论发展（直到二十世纪末）的历史概览，包括俄罗斯信息论学派的贡献。近年来，随着机器学习的快速发展，有损信源编码的新前沿已经出现并正在积极研究中，例如扩散模型等生成式编码技术（例如，[21]）和面向感知的有损压缩（例如，[22]）。

† 考虑一个日常生活中的例子，一个班级学生的平均身高是170cm，那么班级里必然至少有一个学生的身高不低于170cm。

--- 第 19 页 ---

练习

1. 考虑一个有两个分量 (S_1, S_2) 的信源，其中 S_1 和 S_2 是独立的，并让失真度量形式为 $d((s_1, s_2), (s_1^*, s_2^*)) = d_1(s_1, s_1^*) + d_2(s_2, s_2^*)$ 。将 S_i 在失真 d_i 下的率失真函数表示为 $R_i(D)$ ， $i=1,2$ 。求率失真函数 $R(D)$ ，用 $R_1(D)$ 和 $R_2(D)$ 表示。
2. 如果一个DMS S 在失真度量 $d(s, s^*), (s, s^*) \in S \times S^*$ 下的率失真函数是 $R(D)$ ，那么当失真度量变为 $d_{k,b}(s, s^*) = kd(s, s^*) + b$ 时，对于某些 $k > 0, b \geq 0$ ，率失真函数是什么？
3. 我们可以用一个矩阵 D 来集体表示具有有限字母表的信源和再现的失真度量；也就是说， D 的第 i 行第 j 列对应于 $d(s_i, s_j^*)$ 。如果 D 满足其所有列都是某个向量 $[d_1, d_2, \dots, d_{|S|}]$ 的排列，证明率失真函数的以下下界：

$$R(D) \geq H(S) - H(V) \quad (4.53)$$

其中 V 是在所有满足

$$\sum_{i=1}^{|S|} P_V(i) d_i \leq D.$$

的 $\{1, 2, \dots, |S|\}$ 上的随机变量中熵最大的那个。

--- 第 20 页 ---

62

4 信源表示：率失真特性

此外，证明如果 S 是均匀的并且 D 的行是彼此的排列，那么下界 (4.53) 是紧的，即它恰好是率失真函数 $R(D)$ 。

1. 计算并绘制以下模型的率失真函数。 a) 信源字母表为 $S = \{-1, 0, 1\}$ ，再现字母表为 $S^* = \{-1, 1\}$ ，信源分布为均匀分布 $P_S(s) = \{1/3, 1/3, 1/3\}$ ，失真度量为

$$D = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

- b) 信源字母表为 $S = \{-1, 1\}$ ，再现字母表为 $S^* = \{-1, 0, 1\}$ ，信源分布为均匀分布 $P_S(s) = \{1/2, 1/2\}$ ，失真度量为

$$D=(011/21/210)$$

2. 如果一个DMS S 是参数为 δ 的伯努利分布，并且失真度量是汉明失真，当 $PS^\wedge | S$ 达到 $R(D)$ 时，描述编码过程 (4.38) 和 (4.39)-(4.40)，并找出它们的区别。

3. 考虑一个伯努利(1/2) DMS S 。如果我们使用以下码本：

$$C=\{0000,0101,1010,1111\}, (4.54)$$

用编码过程 (4.38) 计算期望汉明失真。

4. 假设有两个独立的伯努利 DMS S_1 和 S_2 ，参数分别为 $\delta_1 \leq 1/2$ 和 $\delta_2 \leq 1/2$ 。每次我们以无记忆的方式从其中一个 DMS 中采样，以得到一个新的 DMS，记为 S 。设采样 S_1 (或 S_2) 的概率为 λ (或 $1-\lambda$)。在汉明失真下， S 的率失真函数是什么？

5. 考虑一个在 $S=\{1,...,m\}$ 上均匀分布的 DMS S ，再现字母表 $S^\wedge=S$ ，以及汉明失真度量。计算率失真函数 $R(D)$ 。这也给出了一个费诺不等式（定理 3.9）达到等式的例子。

6. 在第 4.1 节的问题表述中，不假设失真度量 d 是从 $S \times S^\wedge$ 到 $[0,\infty)$ 的映射，而是让它是一个根据某个条件概率分布 $PU | S, S^\wedge(u | s, s^\wedge)$ 生成的随机变量 $U \in [0,\infty)$ 。推广定理 4.1 以解决这个扩展的问题

--- 第 21 页 ---

4.4 可达性部分的证明 63

表述，并指出在逆定理和可达性证明中必要的修改。

1. 我们对香农信源编码基本定理的陈述明确假设了失真度量 $d(s, s^\wedge)$ 是有界的，即 $d_{\max} < \infty$ 。

a) 解释为什么如果失真度量是无界的，即如果存在某个 $(s, s^\wedge) \in S \times S^\wedge$ 使得 $d(s, s^\wedge) = \infty$ ，定理的可达性部分的证明可能会失效。

b) 考虑存在 $s^\wedge_* \in S^\wedge$ 使得对于所有 $s \in S$ ， $d(s, s^\wedge_*) < \infty$ 的情况。论证在这种情况下，率失真函数 $R(D)$ 仍然由定理 4.1 给出，即在 $E[d(S, S^\wedge)] \leq D$ 的约束下 $\min PS^\wedge | S | S(S; S^\wedge)$ 。

c) 论证存在某些 $(S, S^\wedge, d(s, s^\wedge))$ 的情况，使得率失真函数必须高达 $\log | S |$ ，即不可能有有效的表示，并且 $R(D)$ 可能不同于 $R(D)$ 。

d) 计算以下设置的率失真函数： $S=\{0,1\}$ ， $S^\wedge=\{0,1,e\}$ ， S 是伯努利(1/2)，如果 $s^\wedge=s$ 则 $d(s, s^\wedge)=0$ ，如果 $s^\wedge=e$ 则为 1，如果 $s^\wedge \neq s$ 且 $s^\wedge \neq e$ 则为 ∞ 。