中国科学技术大学

信息论复习宝典

学　　院：网络空间安全学院

课　　程：信息论

主　　编：高源

定稿日期：2024 年春

目录

第二版前言

亲爱的读者：

你好！

如果你正在学习信息论课程，或者正在准备信息论课程的期末考试，那这本书可能会对你有所帮助。转眼间已经是作者担任信息论课程助教的第四个学期了，作者基于这四年间担任信息论课程助教期间答疑、批改作业等过程中收获的反馈以及触发的思考，对《信息论复习宝典》（后面简称“第一版”）的内容进行增补和修改，希望能够更好地契合同学们学习和复习时的需求从而提供更有效的支持。

在第一版的基础上对信息论解题模型、考点总结部分做了补充和完善，模拟试卷部分提供了 2024 年春季学期的最新制作。读者在使用本书时，可以根据不同的阶段尝试不同的使用方法，例如在预习阶段可以参考本书明确重要考点，从而更有针对性地完成课堂学习。

曾经在经验交流会上分享了一个观点，每个课程都有自己的特点，每个人也是这样。也许信息论不一定特别匹配你的学习习惯，也许学这门课的过程中遇到了重重困难，想说，请相信这只是暂时的。作者特别喜欢一首歌的歌词，“想见你，只想见你，未来过去，我只想见你”，这种热烈的情感是有感染力和生命力的，也是可以支持你变得更好。信息论学习，甚至本科课程学习，都只是这一阶段的事情，请保持热爱，请相信总有更多更重要的事情等你去欣赏，总有更多更适合你的精彩等你去创造！

由于作者能力有限，本书中也许仍然存在需要优化的地方，作者诚挚欢迎各位读者反馈建议和意见。同时也欢迎读者一起交流关于信息论课程的理解和思考。另外很期待有优秀的学弟学妹愿意继续参与到信息论课程的教学工作中，作者愿意提供力所能及的支持。谨以此书献给网安学院，祝愿学院的本科教学工作越来越好！

祝好！

高源

2024 年春

第一版前言

亲爱的读者：

你好！

如果你正在学习信息论课程，或者正在准备信息论课程的期末考试，那这本书可能会对你有所帮助。作者基于三年担任信息论课程助教期间答疑、批改作业等过程中收获的反馈，考虑到同学们考前复（预）习时候的需求，特别编著了这本《信息论复习宝典》。

这本书包括信息论解题模型、考点总结、专题复习计划、模拟试卷。信息论解题模型为作者基于对课程的理解总结而成，如果能够熟练掌握该模型，可以顺利解决信息论课程中的各种常见题目，在 2023 年信息论第一次习题课上作者也向班级同学验证了这一点。考点总结部分针对同学们在学习、复习以及考试中经常出现的问题进行总结，同时结合对课程的理解梳理了课程中的重要考点，并对难点进行讲解。如果读者处于复（预）习初期，可以对照着考点总结，筛选目前以及掌握的、还未熟练掌握的、不了解的，然后根据掌握程度安排复习计划。如果读者已经完成了复习，可以对照考点总结检查自己是否存在疏漏。专题复习计划专门针对考试题目对不同考点的考察方式进行设计，读者可以按照自己规划好的节奏对照着专题复习计划中的题目完成复习，这里特别提示，专题复习计划中的题目是作者模拟出题人的视角充分归纳不同的考点对应的考察形式，其中包括作者在习题课上反复强调的易错点等，建议读者在使用的时候主要检查自己在面对每道题目的时候是否有处理思路。模拟试卷主要包含近两年考前准备的模拟题目，建议需要做模拟考试练习的读者使用。

本书包括不同形式的复习材料，供不同需求的读者使用。将这些材料整理成这本书，也是作为对作者三年课程助教经历的总结，同时为学习信息论课程的同学提供可能的帮助。除此外，本书也可以作为将来担任信息论课程的助教同学们准备习题课的参考。作者谨祝愿学习信息论课程的同学都能有满意的收获，祝愿信息安全专业信息论课程教学工作越来越好！

祝好！

高源

2023 年春

信息论解题模型

作者在 2024 年春季学期第一次习题课上有详细地讲解了信息论解题模型的具体内容，并在讲解作业题目的时候展示了如何应用该模型。详见https://www.bilibili.com/video/BV1u1421Z7wP。另外由于讲解这次习题课的时候主要针对离散部分的内容，所以连续部分的内容没有具体谈及，读者可以自行补全。

1. 定义。定义法是一种万能解法，几乎可以适用于求解任何题目，但缺点在于往往使得求解过程比较繁琐。这一点也是合理的，很多事情都是这样，有一利就有一弊。作者把它放在首位，一方面是由于这是最本质的（这也是其万能的原因），一方面是由于在答疑和批改作业中发现一些同学遇到的困难往往是忽略了“解决信息论问题首先要搞清楚概率分布”。事实上，读者可以联想一下课程中学到的内容，第二章和第八章就是在讨论离散和连续的信息度量，自然题目中离不开信息量的定义；第三章主要考察渐进均分性定理和典型集，也是考察定义；第七章主要考察离散信道的信道容量求解，求解的一般范式就是对输入信源概率作一般性假设，实现显式表达互信息，进而按照定义用微积分的知识求解函数极值问题；第十章率失真函数对于初学者往往是最难的部分，难点主要在于对率失真的理解以及使用“先观察再验证”的方法引入失真求解条件熵/微分熵极值，而事实上这里可以通过直接对条件熵/微分熵作一般性假设，实现显式表达互信息，进而按照定义用拉格朗日乘子法求解有约束的函数极值问题。这里，作者仅仅列举了一些往往会选择定义法的情况，并不表示其他章节的内容不可以用定义法，只不过如最初所述，如果有其他解法的时候其他解法往往是更简洁的，此时我们就不会首选定义法了。在这里作者建议，读者能够熟练掌握离散和连续情形下的信息量的定义，并坚持一个想法“解决信息论问题首先要搞清楚概率分布”。

2. 性质和链式法则。信息论课程中一种常见的题目是信息量不等式的证明，性质和链式法则则是解决这类问题的基础。作者建议读者能够熟练掌握信息量相关的重要性质以及其证明方法，例如，熵的四条重要性质是，非负性（可以用定义证明）、极值性（可以构造任意分布和均匀分布的相对熵，进而利用相对熵非负证明）、独立界（可以用互信息非负证明）、条件使得熵减少（可以用互信息非负证明），相对熵重要的性质是非负性（可以用 Jensen 不等式证明），互信息的重要性质是非负性（可以写成联合分布和边缘分布乘积的相对熵，进而利用相对熵非负证明）。

同时可以比较离散情形和连续情形，例如在习题课上作者重点分析的一个思考题，函数变换对于熵的影响，如果应用到连续情形是否有一致的结论（如果没有思路可以尝试举简单的例子，比如考虑线性映射）。同时，不要忽略了链式法则的应用，作者在第一次习题课上用一道补充题目为了提醒大家不要忘记了相对熵的链式法则。

3. 结论/模型。这门课程中一些题目是可以通过直接应用模型解决的，例如存在平稳分布的马尔可夫链熵率、加权图上随机移动问题、哈夫曼编码、弱对称信道的信道容量、三种不同约束下的微分熵极值问题、并联高斯信道的信道容量、并联高斯信源的率失真等等。在这里作者特别强调，使用结论/模型解题时，首先要关注的是使用条件，如果不符合使用条件则不得应用相应的结论/模型。作者的高中数学老师杨老师曾经分享过一个思想“先有意义后求量”，这句话对作者的学习产生了深远的影响，在此向杨老师表示感谢，同时也想把这句话分享给读者，希望也能对你有所启发。

4. 重要不等式。在这门课程中我们主要学习了三个重要的不等式，分别是 Jensen 不等式、数据处理不等式和费诺不等式。作者建议读者熟练掌握三个不等式的内容以及证明方法以及等号成立条件。举个例子，在这门课程中解题时遇到了马尔可夫链，要么是计算熵率问题，要么就是要应用数据处理不等式或者其证明方法，在往年的考试题中有一些易错题就是考察了这一点。关于等号成立条件，这点对于研究不等式也是非常重要的，在往年答疑中就有同学在第十章试图使用费诺不等式求解离散均匀分布信源在汉明失真度量下的率失真问题时不清楚费诺不等式等号成立条件从而遇到问题。

5. 常用技巧。作者在 2024 春习题课上总结了三条，分别是引入示性变量（针对随机变量依概率选择问题）、构造一个信息量的不同展开式以及先观察再验证。其中第一条适用范围比较小，读者只需要观察题目中是否存在依概率选择问题即可（费诺不等式的证明中蕴含了依概率选择问题，相关的思考可以参考作者编著的《信息论学习指导》）；第二条则是广泛应用于信息量不等式证明以及信息量求解，例如数据处理不等式、费诺不等式的证明以及作业题目中一个联合熵的求解；第三条是一种数学方法，常见的应用是先观察并放缩得到一个上（下）界，然后给出一个具体的例子达到该上（下）界，进而得到了上（下）确界，例如第十章求解率失真函数的时候，课堂学习的引入失真后利用性质放缩再结合给出具体的虚拟信道实例其实就是这种方法的应用。关于第三条技巧，在求解最大（小）值问题上往往有奇效，但其缺点在于不容易观察出最佳的放缩目标并且给出实例。放缩部分需要读者熟练掌握信息量的性质和链式法则，同时这里也需要观察出题人的意图，有时候前一小问可能是为了这一问提供支持的。

作者提供一种学习思路，在复习的时候尝试自己把完整的信息论解题模型默写出

来，包括每项内容具体的表达、相关的证明方法，并且尝试总结每种方法使用的案例。同时建议读者按照第一次习题课上作者讲解习题时的思路，尝试分析部分题目的不同解法。作者相信，如果能够熟练地掌握该模型，是可以顺利解决信息论课程中绝大多数题目的（作者在这里想了想觉得是不是还是要谦虚一下比较好）。更进一步地，读者可以尝试把自己设想为助教在讲习题课，尝试着完整地把该模型讲给听众，也许会有不一样的收获哦。事实上，学会和讲清楚其实是两个不同的事情，在尝试讲清楚的时候也会在深入思考中获得更多的收获（作者也曾经建议同学们组织学习小组互相讲解）。

考点总结

4.1 第二章 熵、相对熵、互信息

这一章的考题往往是包含一道选择题（判断信息量不等式是否成立），一道填空题（可能正向考察信息量的计算，也可能逆向考察基于已知信息量推导概率分布），判断题加大题共两道（判断题主要考察对信息量性质的理解，大题主要考察信息量不等式的证明或者计算）。

1. 【掌握】信息量定义。例：比较 $H\left(p_1, p_2, \ldots, p_n\right)$ 和 $H\left(p_1, p_2, \ldots, p_m, q_m\right)+q_m \log(n-m)$ 的大小，其中 $q_m=\left(1-\sum_{i=1}^{m} p_i\right)$。

   - 虽然解题的时候往往会采用更高效的方法，如性质、链式法则等，但要时刻记得最基本的定义，如果其他方法都失效的时候就需要用定义来解决问题。而应用定义解题最关键的就是搞清楚概率分布，往往是需要根据题意对概率分布作假设，进而显式表达目标函数，例如建模为函数极值问题求解。
   - 一般来说，熵的定义式比较常用，二元相对熵的定义式中加权求和的表达 $H(Y \mid X)=\sum_{x \in \mathcal{X}} p(x) H(Y \mid X=x)$ 比较常用，相对熵的定义式在解决含相对熵的问题时较为常用，联合熵、互信息的定义式应用较少（往往是利用链式法则等进行展开处理）。
   - 关于熵的定义式的易错点包括：$H(p)$ 表示参数为 p 的伯努利分布的熵，其定义式为 $H(p)=-p\log p-(1-p)\log(1-p)$，而非 $-p\log p$；熵的定义式中是负对数期望形式，不要忽略了负号。
   - 注意信息量的单位！另外，注意如果需要做单位变化的时候，最好把表达式写成一个完整的对数函数表达式，这样不容易出错。例如，$\log 3+1$ 比特 $=\log 3+\log 2$ 比特 $=\log 6$ 比特 $=\ln 6$ 奈特，而非 $\log 3+1$ 比特 $=\ln 3+1$ 奈特 $=\ln 3e$ 奈特（这里出错原因在于“1”在变换时没有应用换底公式）。

2. 【掌握】信息量性质。例：比较 $H(Y \mid X=x)$ 与 $H(Y)$ 的大小。

   - 常用的性质参见作者编著的另一本教材《信息论学习指导》。

- 信息量的性质往往应用于讨论信息量不等式，这时候就要关注不等号的方向问题，建议读者熟练掌握不同性质中不等号的方向，便于快速选择合适的性质用于解题。
- 建议读者熟练掌握每条性质的完整内容（包括使用条件、具体数学描述以及等号成立条件等）。关于使用条件的问题，存在一些常见的易错点，例如“条件使得熵减少”中的“条件”应该为随机变量（即，条件熵）而非随机事件（随机变量固定了某个取值），“条件使得互信息减少”的成立条件是“构成了马尔可夫链”等。
- 一类常见的考试题是基于信息量的特殊取值推导随机变量的概率分布，例如有三个二元离散随机变量 X,Y,Z 均服从伯努利分布，若要使得 $I(X;Y)=1$ 比特，$I(X;Y\mid Z)=0$，推导 X,Y,Z 的联合概率分布（往往是填空题）。这类题目就需要读者熟练掌握信息量的性质。

3. 【掌握】重要的不等式。例：随机变量 X_1、X_2、X_3、X_4 构成了马尔可夫链 $X_1\to X_2\to X_3\to X_4$. 证明：$I\left(X_1;X_3\right)+I\left(X_2;X_4\right)\leqslant I\left(X_1;X_4\right)+I\left(X_2;X_3\right)$。

- 三个重要的不等式包括 Jesen 不等式、数据处理不等式和费诺不等式。
- Jesen 不等式主要用于证明信息量的一些性质（例如熵的极值性），数据处理不等式用于处理马尔可夫链的相关问题，费诺不等式主要用于处理离散均匀分布信源在汉明失真下的率失真问题。
- 数据处理不等式的证明方法中所蕴含的技巧非常重要，在信息论解题模型中有总结。
- 注意每个不等式的使用条件和等号成立条件。例如，Jensen 不等式讨论的对象是凸函数，等号成立需要随机变量 X 为常量；数据处理不等式讨论马尔可夫链 $X\to Y\to Z$ 问题，等号成立则是 $X\to Z\to Y$ 成立；费诺不等式讨论信道（虚拟信道）模型，等号成立时条件概率矩阵呈现“对角元相同，非对角元均匀分布”态势。

4. 【理解】信息量的物理含义。例：已知 X,Y 的联合分布，通过测定 X 可以消除关于 Y 的不确定度是多少。

- 作业题目中涉及信息量计算的问题往往直接给出待计算的对象，但如果题目通过信息量的物理含义进行考察，读者也应理解待求解对象具体是什么。

4.2 第三章 渐进均分性

这一章在近几年的期末考试题中没有出现，掌握作业题目中的三类问题即可。

1.【理解】弱大数定理和渐进均分性定理。例：作业 3.9 题。

- 掌握定理的内容即可。

2.【了解】给定集合的大小上下界证明。例：作业 3.4 题后两问。

- 采用和典型集大小的上下界证明相同的方法。详见《信息论学习指导》。

3.【理解】典型集相关计算。例：作业 3.13 题。

- 掌握典型集的基本概念即可。提醒读者注意，不要忽略连续情形的典型集。

4.3 第四章 随机过程的熵率

这一章往往就是一道大题计算熵率，其中主要考察存在平稳分布的马尔可夫链的熵率计算。

1.【掌握】存在平稳分布的马尔可夫链的熵率计算。例：已知马尔可夫链状态转移概率矩阵，计算熵率。

- 过程书写，第一步："由题意知，马尔可夫链有限状态、不可约、非周期，因此平稳分布存在"；第二步，计算平稳分布；第三步，代入公式求解熵率。
- 注意，题目中可能给出初始分布，这个往往是迷惑你的（但不排除，有可能会用于其他计算，比如计算初始条件下熵的大小），平稳马尔可夫链的熵率计算公式中一定要用平稳分布。

2.【了解】加权图上随机移动问题。例：作业 4.22。

- 这类问题如果考试出现，直接用公式。因此，需要读者记住公式，并且注意理解公式中每一项的含义。

3.【了解】一般随机过程的熵率计算。例：作业 4.18。

- 这类问题如果考试出现，只能用定义计算。考察可能性不大，不过还是建议读者记住熵率定义式。

4.4 第五章 数据压缩

这一章可能考察一道选择题（下列码字长度符合即时码要求的是、以下编码不可能是二元哈夫曼编码的是，也可能出成判断题），一道填空题（给出哈夫曼编码判断信源概率大小关系、计算期望码长等），一道大题（主要是给出概率分布进行哈夫曼编码）。

1. 【理解】唯一可译码、及时码、前缀码等概念。例：作业 5.37。

   - 掌握基本概念即可。

2. 【掌握】Kraft 不等式。例：下列码字长度符合即时码要求的是。

   - 掌握 Kraft 不等式的内容。
   - 注意不等式是即时码存在的充要条件，也是即时码的必要条件。换句话说，给了一个符合不等式的码长，说明存在即时码满足该码长，但不能说明满足该码长的都是即时码。

3. 【掌握】最优码长理论。例：若某信源有一个三元前缀码满足平均码长 $L = H_3(X)$，则其码字空间大小一定为奇数。

   - 掌握理论内容 $H_D(X) \leqslant L^* < H_D(X) + 1$。
   - 记住哈夫曼编码是最优码。这一性质往往会用于解题，例如上述问题的一种解法是“由于存在三元前缀码满足平均码长 $L = H_3(X)$，则可知该三元前缀码为最优码，又由于哈夫曼编码是最优码，因此该信源的三元哈夫曼编码的平均码长也是 $L_{huffman} = H_3(X)$，所以可知三元哈夫曼编码不需要增加零概率项。由哈夫曼编码的性质可知其码字空间大小 $|\mathcal{X}| = 1 + k(D-1)$，因此可知码字空间大小为 $(3-1)k + 1 = 2k + 1$ 一定是奇数”。

4. 【掌握】哈夫曼编码。例：根据给定的信源概率分布进行哈夫曼编码。

   - 掌握哈夫曼编码的性质 $|\mathcal{X}| = 1 + k(D-1)$。
   - 编码时首先检查是否需要增加零概率项，然后再进行概率合并过程。
   - 若题目要求给出两种编码方案，具有相同的平均码长和不同的方差，其产生方法是，在概率合并的时候改变相同概率项合并的顺序。
   - 若题目增加约束。不要慌，认真分析一下约束到底是什么含义。例如，要求三元哈夫曼编码的第一位可以从 3 个字母中选取，后续编码只能从其中的两个字母中选取，这也就是要求我们先进行二元哈夫曼编码，对最后 3 个待合并的概率项进行三元哈夫曼编码。
   - 这里可能换一种角度考察对哈夫曼编码的理解，例如给出哈夫曼编码结果要求推理概率分布需要满足的条件，参考作业题目 5.25。此外也可能给出一组编码，要求判断是否可能是哈夫曼编码，考察哈夫曼编码是前缀码、最优码（如果给出的码字能够变短，说明不是最优码，从而不是哈夫曼编码）。

5. 【了解】Shannon-Fano-Elias 编码及其相关的前缀码证明方法。例：作业 5.28。

   - 掌握 Shannon-Fano-Elias 编码的基本方法。
   - 掌握 Shannon-Fano-Elias 编码中证明前缀码的方法（码字空间不重叠）。

4.5 第七章 信道容量

这一章一般会是一道大题（考察给定信道的信道容量计算），偶尔会出现一道小题比如判断题（考察对信道容量的理解，例如某一年一道判断题问“达到信道容量的输入概率分布是唯一的”）。

1. 【掌握】信道容量计算。例：给定信道转移概率矩阵，计算信道容量。

    - 一般情况下会直接给出信道的数学描述（转移概率图或者矩阵），但也有时候会以应用题形式考察。遇到应用题不要慌，关键是从中抽象出信道概率模型，具体地，首先确定输入字母表，接着确定输出字母表，最后确定他们之间的对应关系。
    - 信道容量的一般求解，首先检查是否是弱对称信道，如果是的话可以应用结论，否则对于一般的信道，首先假设输入信源概率 $p(x) = (p_1, p_2, \ldots, p_{n-1}, 1-p_1-p_2-\cdots, -p_{n-1})$，进而表达输出的概率 $p(y)$；代入互信息分解式 $I(X;Y) = H(Y) - H(Y \mid X)$，实现用输入信源概率为变量给出互信息的显式表达；最后利用微积分技术求解函数极值问题，得到信道容量以及达到信道容量时的信源概率。**温馨提示：**关于信源概率一般性假设，作者更推荐的是用上述方式，如果读者想采用 $p(x) = (p_1, p_2, \ldots, p_n)$，则在进行函数极值求解时需要引入约束 $p1 + p2 + \cdots + p_n = 1$，即需要用拉格朗日乘子法求解。如果忽略了约束则容易出现无法获得解的困境。

2. 【了解】并联信道的信道容量。例：作业 7.32。

    - 关于并联信道的信道容量的结论在《信息论学习指导》中有讲解。如果考试的时候担心不能直接用，可以根据《学习指导》中的推导过程进行简化，省略中间求解过程，“经过计算得到，并联信道的信道容量满足”后面写上结论，然后就可以用了。这个考点考察可能性比较小，老师在布置作业的时候也要求可以直接参考作业题目中提供的结论，但这个问题的推导过程实际上就是引入示性变量技巧的应用，推荐读者作为练习题目来加强对这一技巧的理解。

4.6 第八章 微分熵

这一章一般会有一道选择题（和第二章一起组成一道选择），一道判断题（考察性质，往往是和离散情形不同的结论），一道大题（给定概率密度函数计算微分熵）。

1. 【掌握】和第二章类似，信息量的定义、性质和链式法则等。

    - 定义、性质、链式法则等请参考《信息论学习指导》。

- 注意和离散情形的区别，例如 $h(AX) = h(X) + \log\det(A)$。
- 注意，这一块容易和正态分布的性质结合考察，建议读者熟练掌握正态分布常用性质，包括正态分布的边缘分布是正态分布，正态分布的条件分布是正态，正态分布的和是正态分布等等。

2. 【理解】连续情形下的渐进均分性定理和典型集。

   - 掌握基本概念和定义即可。

3. 【掌握】不同约束下最大微分熵对应的分布。

   - 记住结论即可，考试可以直接使用。

4.7 第九章 高斯信道

这一章主要是一道填空题（考察带宽有限信道的结论或者并联高斯信道注水法）和一道大题（和填空题考察的内容互补）。

1. 【掌握】高斯信道的信道容量计算及其变形。例：计算在同时给定功率约束和均值约束条件下的信道容量。

   - 互信息分解.$I(X;Y) = h(Y) - h(Y \mid X) = h(Y) - h(Z)$。
   - 由于噪声已知，所以上述步骤将求解互信息最大值问题转化为求解输出最大微分熵问题. 考虑到微分熵最大值求解的方法是利用结论，所以我们需要根据题目中所给约束确定输出 Y 所满足的约束，进而得到 Y 的最大微分熵以及达到最大微分熵时候的概率密度函数。
   - 上述步骤已经得到最大微分熵并确定了 Y，此时只需要给出满足条件的 X，这个问题转化为求解随机变量函数 $X = Y - Z$ 的概率密度函数的问题. 如果是正态分布可以直接利用其性质，否则，可以利用概率论部分学习的方法，引入辅助变量构造一一变换，结合变量代换的公式得到联合概率密度后对辅助变量边缘化即可得到 X 的概率密度函数。

2. 【掌握】并联高斯信道注水法。例：给出独立并联高斯信道各个子信道噪声方差，求解总信道容量。

   - 理解并掌握注水法，详见《信息论学习指导》。
   - 其中一种常见的问法是，并联高斯信道中输入功率约束为何值时，该信道不再像一个信道而开始像一对信道。这类问法，本质上还是考察对注水法本身的理解。

3. 【理解】彩色高斯信道。例：给出独立并联高斯信道各个子信道噪声方差，求解总信道容量。

   - 掌握定义式 $I\left(X^n;Y^n\right)=\frac{1}{2}\log\frac{|K_X+K_Z|}{|K_Z|}$。
   - 了解如何转化为使用注水法解决的问题。

4. 【掌握】带宽有限信道。例：给出带宽限制和信噪比，计算信道容量。

   - 记住结论 $C=W\log\left(1+\frac{P}{N_0W}\right)$ 比特/ 秒，注意单位。
   - 无限带宽下 $C\to W\cdot\frac{P}{N_0W}\cdot\log_2 e=\frac{P}{N_0}\log_2 e$ 比特/ 秒。
   - 注意信噪比的概念，填空题可能出现，详见《信息论学习指导》。

5. 【了解】反馈的影响。

   - 了解无记忆噪声不会影响，有记忆噪声会产生影响，其中两个约束的结论可以记住。

4.8 第十章 率失真理论

这一章是这门课最难的部分，考察往往是作业题直接改编。这一部分可能会考察一道填空题，以及一道大题（常见是率失真函数计算，偶尔会考察信息价值计算）。

1. 【掌握】率失真函数计算。例：二维信源由两个独立且已知分布的伯努利信源组成，每个信源的失真度量都是汉明失真，失真限制是总的失真限制，求二维信源的率失真函数。

   - 课堂上学习的求解范式是基于先观察再验证的思想，首先在互信息分解式的条件熵/微分熵中引入失真，然后对其进行放缩得到互信息的下界，接下来给出条件分布的具体实例（虚拟信道），从而确定得到了下确界，最后给出临界值得到完整的率失真函数。
   - 事实上，率失真问题求解的难点在于如何引入失真并构造虚拟信道. 其实对于连续型变量来说并不算难，考虑到目标是求解一个条件微分熵的极值，而我们对于连续型微分熵极值仅有的了解是三个不同约束下的极值模型（对于解题来说的优势在于，需要考虑的空间变小了），因此在连续型的设定下引入失真的时候只需要根据题目中的约束明确放缩的目标. 但是对于离散型则没有这么简单。
   - 回顾我们在介绍信息论解题模型的时候指出，如果遇到一个问题发现难以求解的时候，不妨用定义法。考虑信息率失真函数的定义，事实上可以对条件概率 $p(\hat{x}\mid x)$ 作一般假设，即用一组变量 $\{p_1,p_2,...,p_m\}$ 表示该条件概率，进

而得到率失真函数的显式表达。这样，就可以将其建模为以 $\{p_1, p_2, \ldots, p_m\}$ 为变量的有约束的函数极值问题，用拉格朗日乘子法求解。读者经过尝试也许会发现这种方法计算比较繁琐，这也是作者在提出信息论解题模型的时候重点强调的，定义法虽然在求解上是“万能”的，但其缺点在于求解过程可能很繁琐，因此在实际应用的时候需要读者做出权衡。

- 需要定义法解决的问题还可能有其他考察方式，例如对失真度量线性变换后如何影响率失真函数。

2. 【理解】从信息的价值理解率失真。例：课堂讨论的产品检验问题。

   - 事实上，这类问题主要考察信息量的计算，进而建立信息价值的概念。

3. 【理解】带失真的信源信道分离定理。例：有一伯努利分布的信源 X，$p(X=0)=1/4$，该信源每 2 秒发出 3 个符号，通过一个错误概率 $\varepsilon < 0.5$ 的二元对称信道进行传输，信道每秒使用 2 次。假设该信源能在允许失真 D 的前提下在该信道中进行传输, 则失真 D 要满足的条件是。

   - 事实上这一定理对信息论来说是非常重要的，不过从考试角度讲，记住 $C > R(D)$ 这个公式即可。

专题复习计划

5.1 第一关

1. 选择题。

   (a) 设 X, Y, Z 为离散随机变量，则以下不等式正确的是 (　)

$$(A) H(Y \mid X=x) \leqslant H(Y)$$
$$(B) I(X;Y) \geqslant I(X;Y \mid Z)$$
$$(C) H(X) < H(2X)$$
$$(D) I(X,Y;Z) \geqslant (X;Z) + I(Y;Z)$$

   (b) 设 X, Y, Z 均为离散随机变量, 则以下等式或不等式成立的是 (　)

$$(A) H(X+Y) \geqslant H(X)$$
$$(B) H(X \mid Y) = H(Y \mid X)$$
$$(C) H(g(X)) \leqslant H(X)$$
$$(D) H(f(X) \mid g(X)) = 0$$

   (c) 以下编码不可能是二元哈夫曼编码的是 (　)

$$(A) \{0, 10, 11\}$$
$$(B) \{00, 01, 10, 110\}$$
$$(C) \{01, 10\}$$
$$(D) \{1, 01, 10\}$$

(d) 以下 D 元字母表上的码字长度符合即时码要求的是 ( )

$$
\begin{aligned}
&(A) D=2, l_i=1,2,3,3,3\\
&(B) D=3, l_i=1,2,2,2,3,3,3,3\\
&(C) D=4, l_i=1,1,1,2,2,3,3,3,4,4\\
&(D) D=5, l_i=1,1,2,2,3,3,4,4
\end{aligned}
$$

2. 填空题。

(a) 有三个伯努利分布的离散随机变量 X, Y, Z, 若要使得 $I(X;Y)=0$ 比特,$I(X;Y \mid Z)=1$ 比特, 则 X, Y, Z 的联合概率分布为__________________ 。

(b) 设 $\{X_n\}$ 为平稳的马尔可夫链。则存在 $k=$ ______________ 满足

$$H\left(X_{-n} \mid X_0, X_1\right)=H\left(X_k \mid X_0, X_1\right) \quad (n>1)$$

(c) 设 X 是 $\{0,1,2\}$ 上均匀分布的随机变量，随机变量 Y 的字母表为 $\{0,1\}$。互信息 $I(X;Y)$ 的最大值为_________ 。

(d) 已知信源概率分布 $p=\left(\frac{1}{100}, \frac{1}{100}, \ldots, \frac{1}{100}\right)$，则其最优二元编码的期望长度为_________ 。

(e) 设有一批电阻，按阻值分 85% 是 $2k\Omega$，15% 是 $5\mathrm{k}\Omega$；按功耗分 64% 是 1/8 W，其余是 1/4 W 。现已知 $2k\Omega$ 阻值的电阻中 80% 是 1/8 W，则通过测量阻值可以减少关于功耗的平均不确定度是______________ 。

3. 判断题。

(a) 存在两个互不相同的概率分布 p, q 使得 $D(p\|q)=D(q\|p)$。

(b) 考虑平稳随机过程 $X_1, X_2, \cdots, X_n$ 。对某个函数 ϕ, 定义 $Y_1, Y_2, \cdots, Y_n$ 为 $Y_i=\phi\left(X_i\right), i=1,2, \cdots$。则有

$$H(\mathcal{Y}) \leqslant H(\mathcal{X})$$

(c) 对平稳随机过程 $X_1, X_2, \cdots, X_n$, 有

$$\frac{H\left(X_1, X_2, \cdots, X_n\right)}{n} \leqslant \frac{H\left(X_1, X_2, \cdots, X_{n-1}\right)}{n-1}$$

(d) 设随机变量 X 的字母表为 $\{0,1,2\}$，对应的概率为 $(0.6,0.3,0.1)$，则使得 D 元字母表上的香农码为最优码的最小整数 D 不小于 9。

(e) 对于离散无记忆信道, 达到信道容量时的输入概率分布 $p(X)$ 是唯一的。

(f) 添加一行到信道转移矩阵不会降低容量。

(g) 设信道 1 的转移概率矩阵为

$$p(y \mid x)=\begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$

其信道容量记为 C_1，信道 2 的转移概率矩阵为

$$p(y \mid x)=\begin{bmatrix} 1-q & q & 0 \\ 0 & q & 1-q \end{bmatrix}$$

其信道容量记为 C_2，信道 3 的转移概率矩阵为

$$p(y \mid x)=\begin{bmatrix} 1-p & p & 0 & 0 & 0 \\ p & 1-p & 0 & 0 & 0 \\ 0 & 0 & 1-q & q & 0 \\ 0 & 0 & 0 & q & 1-q \end{bmatrix}$$

其信道容量记为 C_3。则有 $C_3 \geqslant C_1+C_2$。

(h) 有 n 个完全相同的离散的独立二元对称信道串联，其中每个信道误差概率为 $p=0.01$。该信道以 5000 个二元符号/秒的速度传输输入符号。现有一消息序列共有 15000 个二元符号，并设在该消息中 $p(0)=p(1)=0.5$，从消息传输的角度来考虑，若该信道可以在 10 秒内将这消息序列无失真地传输完，则 n 最大不可超过 20。

4. X_1, X_2, Y 均为离散随机变量并满足 $Y = X_1 + X_2$，在什么条件下以下等式成立。

$$H(Y) = H\left(X_1\right) + H\left(X_2\right)$$

当 X_1 和 X_2 都只能取两个可能的值时, 给出满足上述条件的一个例子。

5. 设 $\{X_i, i=1,2,\cdots\}$ 为时间不变的马尔可夫链，初始状态的概率分布为 $P(X_1=1)=0.33$，$P(X_1=2)=0.46$，$P(X_1=3)=0.21$. 转移概率矩阵为

$$P=\begin{bmatrix} 5/12 & 1/4 & 1/3 \\ 1/4 & 7/12 & 1/6 \\ 1/2 & 1/2 & 0 \end{bmatrix}$$

请计算：

(a) 联合熵 $H(X_1,X_2,X_3)$。

(b) 马尔可夫链的熵率。

6. 设一离散信源只能产生“黑色”和“白色”两种消息，“黑色”出现的初始概率 p(黑色)=0.7，白色出现的初始概率是 p(白色)=0.3。

   (a) 假设消息出现前后没有关联，求该信源的熵率。

   (b) 假设消息出现前后有关联，其依赖关系为 p(白 | 白)=0.7，p(黑 | 白)=0.3，p(白 | 黑)=0.2，p(黑 | 黑)=0.8。求该信源的熵率。

   (c) 求当信源进入平稳状态后，以上两个信源的平均冗余度。

7. 设离散信源的概率分布为

$$\begin{bmatrix} S \\ P(S) \end{bmatrix} = \begin{bmatrix} s_1 & s_2 & s_3 & s_4 & s_5 & s_6 & s_7 & s_8 \\ 0.4 & 0.2 & 0.1 & 0.1 & 0.05 & 0.05 & 0.05 & 0.05 \end{bmatrix}$$

(a) 请给出该信源的三元哈夫曼编码，并计算平均码长。要求给出两种编码方案，平均码长相同，但码长方差不同。

(b) 要求输出码字的第一位可以从 $\{0,1,2\}$ 中选取，而后续位只能从 $\{0,1\}$ 中选取，请计算满足此约束下信源的三元最优编码的码长。

8. 一名间谍按下述方式与他的联系人通信：每个小时，该间谍要么不打电话，要么打电话并只允许电话响一定的次数 (不多于 N 次). 他的联系人不接听电话，只记录电话是否响起，以及响起次数. 由于电话系统的不足，每次打电话，电话正常连通的概率为 $p(0 < p < 1)$，且在不同通话中是独立的. 由于联系人不接听电话，间谍并不知道哪个电话正常连通了. 设该间谍打电话的概率为 q，请计算该通信系统的信道容量 (写成 p 和 q 的函数)。

9. 胶片是由碘酸银晶体按照泊松分布组成，每平方英寸的粒子密度 λ 已知。在不知道碘化银粒子位置的情况下对该胶片随机地进行光照。当碘酸银粒子感光 (接受光照) 后，其所在的区域被点亮，没有碘酸银粒子的区域和有碘酸银粒子但没有感光的区域未被点亮。我们做如下的假设，将胶片的区域划分为若干小格子，大小为 dA。假设每个格子中至多一个碘酸银粒子并且不在格子的边界上。于是，胶片可以看作是一系列具有交叉概率 $1-\lambda dA$ 的并联二元非对称信道。通过必要的近似，计算该胶片的信道容量 (量纲为比特/平方英寸)。

10. 离散信源 X 发出消息经过有噪声信道输出为 Y，且 $Y = X + Z(\bmod 2)$，其中 Z 为干扰噪声，X 和 Z 相互独立，取值空间为 $\{0, 1\}$，$p(X = 0) = \omega$，$p(Z = 0) = \varepsilon$。发送方功率受限，有 $E(X^2) \leqslant 1/2$；干扰方功率也受限，有 $E(Z^2) \leqslant 1/4$。

    (a) 求平均互信息 $I(X;Y)$。

    (b) 当信道噪声分布给定时，求使得 $I(X;Y)$ 最大的信源 X 分布。

    (c) 当信源分布给定时，求使得 $I(X;Y)$ 最小的信道噪声 Z 分布。

5.2 第二关

1. 填空题。

   (a) 由四个子信道构成的独立并联高斯信道的各子信道的噪声方差分另为 1,3,6,9, 总功率限制为 $P=8$, 则总信道容量为 C=_____________ 。

   (b) 设一带宽有限信道的带宽为 5kHZ ，信噪比 (P/N_0W) 为 10 dB, 则该信道的信道容量为 C=_____________ 。

   (c) 设 X,Y 是均值为 0 的联合高斯分布随机变量，且条件概率服从高斯分布 $X \mid Y \sim \mathcal{N}(\frac{a\sigma_x}{\sigma_y}, \sigma_x^2(1-a^2))$，其中 σ_x^2 和 σ_y^2 分布为 X,Y 的方差，a 为一常数。则 $I(X+Y;X-Y)=$________________ 。

2. 判断题。

   (a) 对任意连续随机变量 X 和 $a \neq 1, h(aX) \neq h(X)$ 。

   (b) 反馈不会对信道容量产生影响。

   (c) 高斯信源采用失真度量 $d(u,v)=|u-v|$，其率失真函数为

   $$R(D)=\begin{cases}\frac{1}{2}\log\frac{\sigma^2}{D}, & 0 \leqslant D \leqslant \sigma^2, \\ 0, & D > \sigma^2.\end{cases}$$

   (d) 完善保密性的条件是 $I(K;E)=H(E)-H(M \mid E)$，其中 M,K,E 分别表示明文、密钥和密文。

3. 给定两个连续随机变量 X 和 Y，它们的联合概率密度函数为

$$f(x,y)=\frac{1}{2\pi\sigma_x\sigma_y}\exp\left[-\frac{(x-m_x)^2}{2\sigma_x^2}-\frac{(y-m_y)^2}{2\sigma_y^2}\right],-\infty<x,y<\infty$$

设 $U=X+Y$，$V=X-Y$. 请计算 $h(U)$，$h(V\mid U)$，$D(U\|V)$，$I(U;V)$.

4. 假设有一无记忆加性指数噪声信道，输入随机变量 X 是非负的，输出为 $Y = X + Z$ ，其中随机变量 Z 为独立于输入的指数分布噪声

$$f(Z) = \frac{1}{\mu} \exp(-z/\mu)$$

如果对信道输入的均值进行限制，即 $E[X] \leqslant \lambda$，求该信道的信道容量。

5. 在高斯信道中，除了功率限制，还多了另外一个均值限制，即 $EX^2 \leqslant P, EX \geqslant \alpha$ 。计算在两个限制条件下的高斯信道的信道容量，并给出达到信道容量的输入 X 的分布。

6. 考虑一个在集合 $\{1,2,3,4,5,6\}$ 上均匀分布的信源 X，失真度量为

$$d(x,\hat{x}) = \begin{cases} 3, \text{ 如果} x = \hat{x} \\ 66, \text{ 如果} x \neq \hat{x} \end{cases}$$

试求解该信源的率失真函数 $R(D)$。

7. 设有一个均匀分布的离散无记忆信源的失真度量矩阵如下，求解其率失真函数。

$$d(x,\hat{x}) = \begin{bmatrix} 0 & \infty & \infty & 1 \\ \infty & 0 & \infty & 1 \\ \infty & \infty & 0 & 1 \end{bmatrix}$$

8. 设有一个离散无记忆信源 X，概率分布为 $p(X=0)=p(X=1)=2/5, p(X=2)=1/5$，失真度量采用汉明失真，求解其率失真函数。

$$d(x,\hat{x})=\begin{cases}0, & \text{如果} x=\hat{x}\\ 1, & \text{如果} x\neq\hat{x}\end{cases}$$

9. 考虑在集合 $\{1, 2, \ldots, m\}$ 上均匀分布的信源 X. 若失真度量为汉明失真，即

$$d(x, \hat{x}) = \begin{cases} 0, & \text{如果} x = \hat{x} \\ 1, & \text{如果} x \neq \hat{x} \end{cases}$$

求信源的率失真函数。

10. 某印刷电路板 (PCB) 加工厂的产品合格率约为 98% 。一块好的 PCB 板出厂价约为 100 元, 但如果客户发现一块不合格的板子可向厂方索赔 10000 元。已知厂方检验员检验的正确率约为 95%, 试用信息率失真理论来分析检验的作用并作比较。假设合格品出厂、废品报废都不造成损失。

模拟试卷

1. (每题 4 分，共 4 分) 选择题 (多选题)

   (a) 设 X,Y,Z 均为离散随机变量，U,V 为连续随机变量，f 为一一映射，则以下等式或不等式中不成立的是 (　)

   $$(A)H(X)+H(Y)+H(X,Y,Z)-H(Y,Z)\geqslant H(X,Z)+H(Y,X)-H(Z)$$
   $$(B)h(f(U))=h(U)$$
   $$(C)D(p(x)\|p(y))+D(p(y)\|p(z))\geqslant D(p(x)\|p(z))$$
   $$(D)h(U+V\mid U)=h(V\mid U)$$

2. (每题 4 分，共 8 分) 判断题 (若判断为对，简要说明或证明；若判断为错，简要说明或举出反例)

   (a) 设信源
   $$\begin{bmatrix} X \\ p(x) \end{bmatrix}=\begin{bmatrix} x_1 & x_2 & \cdots & x_6 \\ p_1 & p_2 & \cdots & p_6 \end{bmatrix},\quad \sum_i p_i=1$$
   将此信源编码为 r 元唯一可译变长码,其对应的码长为 $(l_1,l_2,\cdots,l_6)=(1,1,2,3,2,3)$。则 r 值的最佳下限为 3。

   (b) $H(E)=I(K;E)+H(M\mid E)$ 是完善保密性的充分必要条件。

3. (每题 4 分，共 16 分) 填空题

   (a) 有三个服从伯努利分布的离散随机变量 X, Y, Z, 若要使得 $I(X;Y) = 1$ 比特，$I(X;Y \mid Z) = 1$ 比特，则 X, Y, Z 的联合概率分布为__________________ 。

   (b) 设随机变量 X 的字母表为 $\{0, 1, 2\}$，对应的概率为 $(0.6, 0.3, 0.1)$，则使得 D 元字母表上的香农码为最优码的最小整数 D 应该为__________________ 。

   (c) 在图片传输中, 每帧约为 2.25×10^6 个像素，为了能很好地重现图像，需分 16 个亮度电平，并假设亮度电平等概率分布。使用信噪比 30 dB 的信道传输，每秒钟传送 30 帧图片所需信道的带宽为__________________ 。

   (d) 设一带宽有限信道的带宽为 10kHZ，信噪比为 20dB，则该信道的信道容量为__________________ 。

4. (10 分) 设 X 是 $\{0, 1, 2\}$ 上均匀分布的随机变量，随机变量 Y 的字母表为 $\{0, 1\}$。请计算互信息 $I(X;Y)$ 的最大值，给出达到最大值时的联合分布并计算 $H(X+Y \mid X-Y), I(X+2Y; 2X-Y)$。

5. (8 分) 设 $\{X_i, i=1,2\cdots\}$ 为时间不变的马尔可夫链，初始状态概率分布为 $(1/3,1/6,1/4,1/4)$，状态概率转移矩阵为

$$\mathbf{P}=\begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/2 & 0 & 0 & 1/2 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix}$$

请计算：

(a) $H(X_1,X_2,X_3,X_4)$；

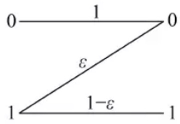
(b) 该马尔科夫链的熵率。

6.（10 分）若有一信源

$$\begin{bmatrix} X \\ p(x) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \\ 0.8 & 0.2 \end{bmatrix}$$

每秒钟发出 2.66 个信源符号。将此信源的输出符号送入某一个二元信道中进行传输（假设信道是无噪无损的），而信道每秒钟只传递 2 个二元符号。试问信源不通过编码能否直接与信道连接？若通过适当编码能否在此信道中进行无失真传输？若能连接，试说明如何编码并说明原因。

7. (10 分) 有一个离散 Z 形信道，其信道转移概率如图所示。试求：



(a) 信道容量；

(b) 若将两个同样 Z 形信道串联，请计算串联后的信道容量；

(c) 若将两个同样 Z 形信道并联，请计算并联后的信道容量。

8. (10 分) 连续型随机变量 X 和 Y 的联合概率密度为

$$f(x,y)=\frac{1}{2\pi\sqrt{SN}}\exp\left\{-\frac{1}{2N}\left[x^2\left(1+\frac{N}{S}\right)-2xy+y^2\right]\right\}$$

(a) 计算 $h(X),h(Y),h(Y\mid X),h(X+Y)$；

(b) $U=X+2Y$，$V=2X-Y$，请计算 $I(U;V)$。

9. (12 分) 设信道输入是连续型的随机序列 $\boldsymbol{X}=\left(X_1, X_2, \cdots, X_N\right)$，输出也是连续型的随机序列 $\boldsymbol{Y}=\left(Y_1, Y_2, \cdots, Y_N\right)$，信道传递概率密度函数为 $f(\boldsymbol{y} \mid \boldsymbol{x})$ 。试证明

(1) 信源无记忆时则存在

$$I(\boldsymbol{X} ; \boldsymbol{Y}) \geqslant \sum_{i=1}^N I\left(X_i ; Y_i\right)$$

当且仅当信道无记忆时等式成立。

(2) 信道无记忆时则存在

$$I(\boldsymbol{X} ; \boldsymbol{Y}) \leqslant \sum_{i=1}^N I\left(X_i ; Y_i\right)$$

当且仅当信源无记忆时等式成立。

10. (12 分) 考虑在集合 $\{0,1,2,3\}$ 上均匀分布的信源 X，再生点 $\hat{X}$ 来自集合 $\hat{\mathcal{X}} = \{0,1,2,3,4,5,6\}$，采用失真度量

$$d\left(x_i, \hat{x}_j\right) = \begin{cases} 0, & i = j, \\ 1, & i = 0 \quad 1 \text{ 且} j = 4, \\ 1, & i = 2 \quad 3 \text{ 且} j = 5, \\ 3, & j = 6 \quad i \text{ 任意}, \\ \infty, & \text{其他}. \end{cases}$$

求信源的率失真函数 $R(D)$。考虑另一种失真度量

$$\tilde{d}(x, \hat{x}) = d(x, \hat{x}) - \min_{\hat{x} \in \hat{\mathcal{X}}} d(x, \hat{x}) \quad x \in \mathcal{X}, \hat{x} \in \hat{\mathcal{X}}$$

它相当于失真矩阵中某一行各元素减去这行中的最小失真值。请证明此失真度量下的率失真函数为 $\tilde{R}(D)$ 满足 $\tilde{R}(D) = R\left(D + D_{\min}\right)$,其中 $D_{\min} = \sum_{\mathcal{X}} p\left(x_i\right) \min_j d\left(x_i, \hat{x}_j\right)$。

后记

转眼间已经是作者担任信息论课程助教的第四个学期了，临别之际更新了一下本书，算是一种告别，也是留给网安学院的一份小小的礼物吧。很久很久，耳边萦绕的是“三叩老臣当去远，难以谋事于殿前，陛下明思决断莫牵怀挂念”。从 2020 年春天第一次接触信息论每周作业至少要一天起步，到 2021 年春天下定决心要参与这门课程的教学以分享一点有限的经验希望能避免更多人遇到同样的困境，再到 2024 年春天即将说再见，认识了很多优秀的人，对这门课程也有了更多的认识。所以决定把这些思考以这本书的形式呈现，希望能帮助到更多有需要的人。谈起这本书的创作过程，自然离不开刘斌老师的指导和同学们在答疑和作业中的反馈，在此作者表示衷心的感谢！

最后，祝愿网安学院的本科教学越来越好！