中国科学技术大学

信息论学习指导

学　　院：信息科学技术学院

专　　业：信息安全

主　　编：高源

指导教师：刘斌老师

目录

课程综述

写在全书开篇的一句话：“如果能把这本书的内容都搞懂，这门课就没问题.”课程中会考察的点在本书中都能找到，如果本书没有列出的基本不会考；另一方面，本书中的内容都是考点，无需担心冗余问题.

1.1 主要内容

从总体上看，信息论课程主要分为两部分，分别是：

- **离散分布**的信息度量、信源编码理论、信道编码理论
- **连续分布**的信息度量、信源编码理论、信道编码理论

按照教学内容划分章节，课程主要包括 9 章内容，分别介绍

- 第 2 章——熵、相对熵、互信息

  这一章主要介绍离散分布的信息度量的方法——**熵，并介绍相关的概念**，如联合熵、条件熵、相对熵、条件相对熵、互信息、条件互信息等，以及熵、相对熵、互信息的**链式法则及其性质**.

  基于此，介绍三个重要的不等式：Jensen 不等式，数据处理不等式，费诺不等式. 这些不等式在**解题中有着重要的应用**.

- 第 3 章——渐进均分性

  这一章提出了渐进均分性定理，以及典型集的概念. 渐进均分性定理为我们提供了数据压缩可行性的理论基础，这也是其重要意义所在. 考核方面，**掌握定理的内容、理解定理的意义，并掌握典型集的概念及性质**即可.

- 第 4 章——随机过程的熵率

  这一章提出了随机过程熵率的概念. 介绍了平稳过程、马尔可夫链、马尔科夫链不可约和非周期等概念，提出了熵率的计算方法，并介绍了加权图上随机移动的熵率这一例子. 时间不变的、不可约的、非周期的马尔科夫链的**熵率计算**是这一章的重点.

- 第 5 章——数据压缩

  这一章研究离散分布的信源编码. 首先介绍信源编码期望长度的定义，非奇异码、唯一可译码、前缀码 (即时码) 的**概念**，接着提出了即时码存在的充要条件——**Kraft 不等式** (一般是直接考察概念，给出码长问是否可以构成即时码). 基于香农第一定理提出最优码的期望长度的估计，然后介绍了一种常见的最优码——哈夫曼编码. **哈夫曼编码的编码规则、D 元哈夫曼码的码字个数征 (与之直接相关的是多元码在编码前必须首先检查是否需要添加零概率项)** 都是这一章的重点. 最后介绍了另外一种编码方案.

- 第 7 章——信道容量

  这一章研究离散分布的信道编码. 首先提出信道容量的定义，以二元对称信道、二元擦除信道等常见信道为例介绍信道容量计算的方法. 提出对称信道的概念，并介绍对称信道的**信道容量计算方法**. 介绍了信道容量的几条基本性质. 提出码率的概念，并说明信道容量是所有可达码率的上界. 介绍了联合典型序列的概念 (和第 3 章所述类似)，提出联合渐进均分性定理. 本章的另一个重点是信道编码定理，信道编码逆定理以及零误差情况下的逆定理的证明是需要了解的内容. 最后提出了反馈信道的概念，并给出结论——对于离散信道，反馈容量和信道容量相同.

- 第 8 章——微分熵

  这一章介绍了连续分布的信息度量的方法——**微分熵，并介绍相关的概念**，如联合熵、条件熵、相对熵、条件相对熵、互信息、条件互信息等，以及熵、相对熵、互信息的链式法则及其性质. 正态分布是重要的连续分布，**多元正态分布的熵**也是这一章的重点，基于**任意非负定矩阵都可以是某个正态分布的协方差矩阵**这一重要性质，了解基于多元正态分布的熵及其性质来证明矩阵行列式不等式 (看一下老师讲的例子大概就知道怎么做了，我在学习的时候认真研究了一下这一块，目前的作业题和考试题还没见到，但是觉得也许不排除判断题给出来一个). **微分熵和离散熵的不同**是这一章学习中需要关注的重点，例如**线性变换对于微分熵的影响**的相关公式是需要熟练掌握的.

- 第 9 章——高斯信道

  这一章介绍连续分布的信道编码理论——提出高斯信道的概念，介绍高斯信道容量的定义以及**计算方法** (学习高斯信道信道容量的计算时要注意理解其方法，考察的时候可能考察其他加性噪声信道)；介绍带宽有限信道的概念，并提出**带宽有限信道的信道容量的计算方法** (往年题目中有考察，见本书最后的期末试卷. 这一考点感觉应该就是在填空题考察，只需要记住公式即可). 另一个重点是并联高斯信道，提出了其信道容量计算方法——**注水法**. 接着介绍高斯彩色噪声信道及其**信道容量计算**. 最后研究带反馈的高斯信道，分析**反馈对于信道容量的影响**——和离散信道不同.

- 第 10 章——率失真理论

  这一章介绍连续分布的信源编码理论——介绍失真度量方法，率失真函数、失真率函数的概念，并提出基于信息率失真函数**计算率失真函数**的方法. 离散和连续分布的率失真函数计算是这一章的重点. 针对并联高斯信源，提出**反注水法**，这也是考察的重点. 这一章是这门课程的难点，不过我已经总结好求解方法，并且根据分类，分别以一道例题展示求解方法，应该就可以顺利搞定了.

- 末章——信息论和信息安全

  这一章介绍信息安全的信息论模型，介绍冗余度、唯一解距离、Equivocation、完善保密性的概念，并提出完善保密性的充要条件.

1.2 课程特点

这门课程理论性较强，需要微积分 (数学分析)、线性代数、概率论的基础.

从**主要内容上看，课程主要研究的是**离散和连续分布的信息度量方法、信源编码理论、信道编码理论. 离散情形和连续情形所分析的内容是类似的，一般也有着对应关系，但同时也有一些概念和性质存在着不同.

从考察角度来看，考试题目注重考察性质、链式法则、重要定理的理解和求解方法的掌握. 部分题目可能会难点较大，不过我觉得理解了这些基本能够应付得来.

部分考试题目和作业题目比较类似，认真完成作业有助于取得优异成绩（手动狗头）.

1.3 学习方法

明确了主要内容和课程特点，就可以讨论课程的学习方法.

- **关注离散情形和连续情形的不同**，比较学习 (首先学习离散情形，然后在学习连续情形的时候可以对比离散情形进行学习，关注二者概念、性质等方面的异同)
- 熟练掌握基本概念——熵、联合熵、条件熵、相对熵、互信息、熵率、信道容量、率失真函数等
- 熟练掌握重要的不等式、性质、链式法则等的推导
- 注意性质的成立条件 (例如条件使得熵减少对于随机变量本身没有要求，但是条件使得互信息减少要求三个随机变量构成马尔可夫链)
- 注意不等式等号成立条件

- 熟练掌握联合熵、相对熵、互信息等的链式法则，掌握分解、构造的技巧
- 总结信道容量的求解流程，并根据信道特点分类掌握其求解细节
- 记住任意非负定矩阵都可以是某个多元正态分布的协方差矩阵，进而利用多元正态分布的熵来处理矩阵不等式
- 单纯记住重要但不会考察细节的公式——例如我觉得反馈对于高斯信道容量的影响就是一个
- 总结率失真函数求解流程，注意细节 (这一部分是难点，也是易错点，老师在讲课时候也强调过这里往年失分情况)

1.4 信息论解题模型

作者在 2023 年春季学期任助教期间，通过在批改作业和答疑中获得的反馈，决定系统地给班级同学总结出信息论课程常用解题方法组合成的信息论解题模型. 本质上，是在熟练掌握各个重点内容后的比较、凝练以及在解决具体题目时的经验总结. 由于作者不再继续担任该课程的助教，希望用这种形式把这些经验传递下去，希望信息安全专业的信息论课程教学越来越好.

读者可以在本书的正文中找到这里的总结所对应的具体内容，并可以从每一道作业题中检验该模型的使用方法和价值.

1. 定义. 定义是最朴素也是在很多时候最有效的解决方法. 作者在答疑过程中发现同学们解题遇到困难往往都是由于没有弄清楚题目中所对应的概率模型. 作者建议，如果遇到一个题目不知道从何处切入的时候，首先问一下自己，**概率分布是否搞清楚了**.

2. 性质. 信息量的一些重要性质，例如离散情形熵、相对熵、互信息非负性，熵的极值性、独立界等等. 要注意的是，每一条性质的完整描述，例如“条件使得熵减少”中条件是“随机变量”而非“事件”.

3. 链式法则. 信息量的展开处理是处理二元乃至多元问题的常用手段. 这一点也不难解释，我们常用的结论往往都是关于单个变量之间的关系.

4. 重要不等式.Jesen 不等式、数据处理不等式、费诺不等式等重要不等式，是这门课程中常用的解题手段.

5. 信息论常用技巧. 遇到“随机变量依概率选择”问题时往往需要引入示性变量（费诺不等式的证明）. 构造一个信息量的不同展开式也是一种非常常见的技巧（数据处理不等式、费诺不等式的证明）

6. 一般性的解题技巧. 例如利用上界和上确界的关系：题目要求求解最大值，直接求解往往需要建模成函数极值问题，如果发现求解起来有些麻烦，可以考虑先利用信息论的技术对其进行放缩（例如用性质、重要不等式），然后给出等号成立条件，也就是举例子说明这个上界能取到，即为上确界（这个是必要的，不然前面的过程只能得到上界，而不是上确界. 关于例子，可以通过观察题目前几问提供的信息、基于性质的猜测等等得到）. 一些具有多个小问的题目不同的小问之间可能是存在一定的关系的，有时候前面的小问会是给后面问题的辅助. 要学会猜测出题人意图，比如作业题目中有一道题给了一个情景并要求计算估计量和验证费诺不等式，那出题人可能是会希望你给出紧致的结论，所以要考虑费诺不等式的加强版本.

这个信息论解题模型是作者在课程学习和三个学期担任助教期间的经验总结，并在2023 年春季学期的习题课中分享给班级同学，在讲解每一道题目的时候都展示了该模型是如何使用的. 作者希望这个模型能够给读者带来帮助，或是把作者总结的模型消化吸收，或者以此为参考自己总结适合自己学习的模型，或者以此为参考自己选择合适的信息论学习方法.

重要数学基础——概率论与数理统计

2.1 事件的概率

2.1.1 概率的加法定理

若干个**互斥**事件之和的概率, 等于各事件的概率之和：

$$P\left(A_1+A_2+\cdots\right)=P\left(A_1\right)+P\left(A_2\right)+\cdots$$

2.1.2 条件概率

设有两事件 A,B 而 $P(B)\neq 0$. 则“在给定 B 发生的条件下 A 的条件概率”，记为 $P(A\mid B)$，定义为：

$$P(A\mid B)=P(AB)/P(B)$$

2.1.3 事件独立

两事件 A,B 若满足

$$P(AB)=P(A)P(B)$$

则称 A,B 独立.

注意，概率论意义上的独立可能和我们直观理解上的独立存在一定差异. 在信息论课程中，涉及“独立”概念的时候，一定要按照概率论中独立的定义来验证!

举例说明. 定义随机事件：$A=\{$ 三个骰子掷出的点数中至少有两个一样 (即不全相异)$\}$，$B=\{$ 至少有一个骰子郑出 1$\}$. 问 A,B 是否独立?

初一看使人的倾向于相信 A,B 独立，理由如下：知道 B 发生，即知道掷出的点中有 1，对 A 而言, 似与知道掷出的点中有 2 (或 $3,4,5,6$ 都可以) 一样. 故 1 这个数并不相对地更有利于或更不利于 A 发生. 经过计算发现不然：A,B 并不独立.（计算过程省略，请读者自行验证）

2.1.4 多个事件的独立

设 $A_1, A_2, \cdots$ 为有限或无限个事件. 如果从其中**任意**取出有限个 $A_{i_1}, A_{i_i}, \cdots, A_{i_i}$ 都成立

$$P\left(A_{i_1} A_{i_2} \cdots A_{i_m}\right) = P\left(A_{i_1}\right) P\left(A_{i_2}\right) \cdots P\left(A_{i_m}\right)$$

则称事件 $A_1, A_2, \cdots$ 相互独立或简称独立.

此时，乘法定理可以写作：若干个独立事件 $A_1, \cdots, A_n$ 之积的概率, 等于各事件概率的乘积：

$$P\left(A_1 \cdots A_n\right) = P\left(A_1\right) \cdots P\left(A_n\right)$$

2.1.5 贝叶斯公式

定义完备事件群：

$$B_i B_j = \varnothing(\text{ 不可能事件}), \text{ 当} i \neq j$$

$$B_1 + B_2 + \cdots = \Omega \text{ (必然事件)}$$

有全概率公式

$$P(A) = P\left(B_1\right) P\left(A \mid B_1\right) + P\left(B_2\right) P\left(A \mid B_2\right) + \cdots$$

对于事件 A, B，有

$$P(B \mid A) = P\left(AB_i\right) / P(A) = P\left(B_i\right) P\left(A \mid B_i\right) / \sum_j P\left(B_j\right) P\left(A \mid B_j\right)$$

2.2 随机变量及其概率分布

2.2.1 离散随机变量

随机变量就是“其值随机会而定”的变量. 设 X 为离散型随机变量, 其全部可能值为 $\{a_1, a_2 \cdots\}$. 则

$$p_i = P\left(X = a_i\right), i = 1, 2, \cdots$$

称为 X 的概率函数. 除了用概率函数来描述离散随机变量以外，还可以使用概率分布表

可 能 值	a_1	a_2	$\cdots$	a_i	$\cdots$
概 率	p_1	p_2	$\cdots$	p_i	$\cdots$

形式上类似于我们在数字逻辑电路课程中学习的真值表.

2.2.2 连续随机变量

连续性随机变量的最根本要求是，存在概率密度函数. 设连续性随机变量 X 有概率分布函数 $F(x)=P(X\leqslant x),-\infty<x<\infty$），则 $F(x)$ 的导数 $f(x)=F'(x)$, 称为 X 的概率密度函数. 概率密度函数的重要性质有：

1. $f(x)\geqslant 0$
2. $\int_{-\infty}^{\infty}f(x)\mathrm{d}x=1$
3. 对任何常数 $a<b$ 有

$$P(a\leqslant X\leqslant b)=F(b)-F(a)=\int_a^b(x)\mathrm{d}x$$

2.2.3 重要的离散随机变量分布——伯努利分布

如果随机变量 X 只取 0 和 1 两个值，并且相应的概率为

$$P(X=1)=p,P(X=0)=1-p,0<p<1$$

则称随机变量 X 服从参数为 p 的伯努利分布，记作 $X\sim B(p)$.

2.2.4 重要的连续随机变量分布——正态分布

如果一个随机变量具有概率密度函数

$$f(x)=(\sqrt{2\pi}\sigma)^{-1}\mathrm{e}^{-(x-\mu)^2/2\sigma^2},-\infty<x<\infty$$

则称 X 为正态随机变量并记为 $X\sim N\left(\mu,\sigma^2\right)$.

2.2.5 多维随机变量

离散型多维随机变量：每一维都是离散型随机变量.

连续型多维随机变量：存在概率密度函数（注意，每一维都是连续型随机变量不一定是连续型多维随机变量）.

2.2.6 重要的多维连续随机变量分布——多维正态

如果一个多维随机变量具有概率密度函数

$$f_{\mathbf{x}}\left(x_1,\ldots,x_k\right)=\frac{1}{\sqrt{(2\pi)^k|\boldsymbol{\Sigma}|}}\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

则称 X 为多维正态随机变量并记为 $X \sim N(\mu, \Sigma)$. 其中，常见的是二维正态分布

$$f(x,y)=\left(2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}\right)^{-1}\exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2}-\frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right)\right]$$

其中 $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$ 都是常数，我们称 (X_1, X_2) 服从参数为 $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$ 的二维正态分布，常把这个分布记作 $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho))$.$\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$ 的范围分别为 $-\infty < \mu_1 < +\infty; -\infty < \mu_2 < +\infty; -1 < \rho < 1$ $\sigma_1 > 0; \sigma_2 > 0$. 这个函数在三维空间中的图像是一个椭圆切面的钟倒扣在 Ox_1x_2 平面上，其中心在 (μ_1, μ_2) 点.
其重要性质包括：

1. 正态分布的边缘分布是正态分布，均值和方差对应不变.
2. 正态分布的条件分布是正态分布. 设 (X_1, X_2) 服从二维正态分布 $N(a, b, \sigma_1^2, \sigma_2^2, \rho)$，$X_2 \mid X_1 \sim \mathcal{N}\left(b+\rho\sigma_2\sigma_1^{-1}(x_1-a), \sigma_2^2(1-\rho^2)\right)$
3. 正态分布的和是正态分布. 设 (X_1, X_2) 服从二维正态分布 $N(a, b, \sigma_1^2, \sigma_2^2, \rho)$，则 $X_1+X_2 \sim \mathcal{N}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2)$. 特别地，若 X_1, X_2 独立，则有 $X_1+X_2 \sim \mathcal{N}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$.

2.3 小结

概率论是信息论课程的重要基础课程. 为了更好地学习信息论课程，建议读者合理安排时间温习概率论相关基础知识. 这本《学习指导》中只介绍最常用的概念和性质，也许还不足以支撑信息论课程学习. 读者可以进一步参考作者写的“概率论整理”，有必要的时候建议参考陈希孺老师编著的《概率论与数理统计》.

除了概率论以外，这门课程还需要一定的微积分和线性代数基础. 在学习指导中相关章节补充了一些数学基础知识整理，供读者参考. 如有进一步需求，请参考相关课程的教材.

第 2 章　熵、相对熵、互信息

3.1　主要内容

第 2 章是课程的基础，也是重要章节之一. 这一章主要需要掌握

3.1.1　熵

$$H(X) = -\sum_{x\in\mathcal{X}} p(x)\log p(x)$$

注意量纲问题 (**不写量纲是错误的**). 我们习惯上用以 2 为底的对数，相应熵的单位是比特 (不过在第 8 章部分，会为了求解方便在求解过程中使用以 e 为底). 注意的是，在计算熵的时候按照这个公式，选取不同的底对应不同的单位，被求取对数的函数本身不需要作任何变化 (即，以第 8 章部分为例，首先以 e 为底进行计算，最后需要换成单位为比特时，只需要把底换成 2 即可，不需要作其他任何改变). 事实上，这个问题比较容易理解. 我们希望度量信息量，并通过一些推理选取了一种合适的信息度量函数——负对数期望. 一个基本的原则是，信息量本身肯定不会因为度量方式的选取而改变，因此使用不同的对数底的时候肯定也不会带来信息量的改变. 但是我们也知道，一个量在使用不同底的对数的时候结果是不一样的，因此需要用一个手段消弭这个差距，在这里这个手段就是“单位”. 使用不同底的时候单位不同，单位之间的换算关系解决了这个差异，在实际使用的时候我们就可以安心地使用具有良好定义的负对数期望函数，而不需要再额外做任何改变.

关于不同信息量单位之间的转换，我们需要掌握一个重要的数学基础知识——对数换底公式

$$\log_a x = \frac{\log_b x}{\log_b a}$$

由此可以得到

$$1\text{奈特} = \frac{\log_2 X}{\log_e X}\text{比特} = \log_2 e\text{比特}$$

注意，本书中所有没有标注底的 log 都是表示以 2 为底.

3.1.2 熵和期望的关系

$$H(X) = E_p\left[-\log p(X)\right]$$

在这里可能觉得这个表达没有什么特别的. 不过在后面条件熵的时候，这样的处理是有帮助的.

3.1.3 熵的性质

1. 非负性：$H(x) \geqslant 0$(由熵的定义式、对数函数性质可知)
2. 极值性：$H(X) \leqslant \log|\mathcal{X}|$，等号成立当且仅当 X 服从均匀分布. 其中 $|\mathcal{X}|$ 为 X 字母表中元素个数.(证明见后，基于 Jensen 不等式)
3. 条件使得熵减少：$H(X \mid Y) \leqslant H(X)$，等号成立当且仅当 X、Y 独立.(由互信息非负可知)
4. 熵的独立界：$H\left(X_1, X_2, \cdots, X_n\right) \leqslant \sum_{i=1}^{n} H\left(X_i\right)$，等号成立当且仅当各个变量互相独立.(由条件使得熵减少可知)
5. 熵的值只和概率分布有关，和随机变量的取值、取值的顺序无关.(由定义式可知)

3.1.4 联合熵

$$\begin{aligned} H(X,Y) &= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x,y) \\ &= -E\log p(X,Y) \end{aligned}$$

3.1.5 条件熵

$$\begin{aligned} H(Y \mid X) &= \sum_{x\in\mathcal{X}} p(x)H(Y \mid X = x) \\ &= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(y \mid x) \\ &= -E\log p(Y \mid X) \end{aligned}$$

其中特别注意条件熵的定义式，在实际解题的时候最常用到

$$H(Y \mid X) = \sum_{x\in\mathcal{X}} p(x)H(Y \mid X = x)$$

3.1.6 熵的链式法则

$$\begin{aligned}
H(X,Y) &= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x,y) \\
&= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\left[\log p(x)p(y\mid x)\right] \\
&= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x) - \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(y\mid x) \\
&= -\sum_{x\in\mathcal{X}} p(x)\log p(x) - -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(y\mid x) \\
&= H(X) + H(Y\mid X)
\end{aligned}$$

类似地有

$$H(X,Y) = H(Y) + H(X\mid Y)$$

以及

$$H(X,Y\mid Z) = H(X\mid Z) + H(Y\mid X,Z)$$

$$H(X,Y\mid Z) = H(Y\mid Z) + H(X\mid Y,Z)$$

推广到多个变量

$$\begin{aligned}
H\left(X_1,X_2,\cdots,X_n\right) &= -\sum_{(x_1,x_2,\ldots,x_n)\in\mathcal{X}^n} p\left(x_1,x_2,\ldots,x_n\right)\log p\left(x_1,x_2,\ldots,x_n\right) \\
&= -\sum_{(x_1,x_2,\ldots,x_n)\in\mathcal{X}^n} p\left(x_1,x_2,\ldots,x_n\right)\log\prod_{i=1}^{n} p\left(x_i\mid x_{i-1},\ldots,x_1\right) \\
&= -\sum_{(x_1,x_2,\ldots,x_n)\in\mathcal{X}^n} p\left(x_1,x_2,\ldots,x_n\right)\sum_{i=1}^{n}\log p\left(x_i\mid x_{i-1},\ldots,x_1\right) \\
&= -\sum_{(x_1,x_2,\ldots,x_n)\in\mathcal{X}^n}\sum_{i=1}^{n} p\left(x_1,x_2,\ldots,x_n\right)\log p\left(x_i\mid x_{i-1},\ldots,x_1\right) \\
&= -\sum_{(x_1,x_2,\ldots,x_n)\in\mathcal{X}^n}\sum_{i=1}^{n} p\left(x_1,x_2,\ldots,x_i\right)\log p\left(x_i\mid x_{i-1},\ldots,x_1\right) \\
&= -\sum_{i=1}^{n}\sum_{(x_1,x_2,\ldots,x_i)\in\mathcal{X}^i} p\left(x_1,x_2,\ldots,x_i\right)\log p\left(x_i\mid x_{i-1},\ldots,x_1\right) \\
&= \sum_{i=1}^{n} H\left(X_i\mid X_{i-1},\cdots,X_1\right)
\end{aligned}$$

这里所体现的对于**联合熵的不同展开方式**，在解题中常常会用到！

3.1.7 相对熵

$$\begin{aligned} D(p\|q) &= \sum_{x\in\mathcal{X}} p(x)\log\frac{p(x)}{q(x)} \\ &= E_p \log\frac{p(X)}{q(X)} \end{aligned}$$

3.1.8 相对熵的性质

1. 非负性：$D(p\|q)\geqslant 0$，当且仅当对于任意的 x 有 $p(x)=q(x)$ 时等号成立.(不局限于单个变量，这里只是举个例子. 证明见后，基于 Jensen 不等式)
   $D(p(y\mid x)\|q(y\mid x))\geqslant 0$，即对于条件分布非负性也成立.
2. 可能无界 (和熵不同)：如果存在一个 $x\in\mathcal{X}$ 使得 $p(x)>0$，$q(x)=0$ ，则有 $D(p\|q)=\infty$
3. 不对称：一般情况下 $D(p\|q)\neq D(q\|p)$. 把相对熵看成距离只是为了帮助理解

3.1.9 相对熵的链式法则

条件相对熵的定义

$$D(p(y\mid x)\|q(y\mid x)) = \sum_x p(x)\sum_y p(y\mid x)\log\frac{p(y\mid x)}{q(y\mid x)}$$

链式法则

$$\begin{aligned} D(p(x,y)\|q(x,y)) &= \sum_x\sum_y p(x,y)\log\frac{p(x,y)}{q(x,y)} \\ &= \sum_x\sum_y p(x,y)\log\frac{p(x)p(y\mid x)}{q(y)q(y\mid x)} \\ &= \sum_x\sum_y p(x,y)\log\frac{p(x)}{q(x)} + \sum_x\sum_y p(x,y)\log\frac{p(y\mid x)}{q(y\mid x)} \\ &= \sum_x p(x)\log\frac{p(x)}{q(x)} + \sum_x\sum_y p(x,y)\log\frac{p(y\mid x)}{q(y\mid x)} \\ &= D(p(x)\|q(x)) + D(p(y\mid x)\|q(y\mid x)) \end{aligned}$$

3.1.10 互信息

$$\begin{aligned} I(X;Y) &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \\ &= D(p(x,y)\|p(x)p(y)) \\ &= E_{p(x,y)}\log\frac{p(X,Y)}{p(X)p(Y)} \end{aligned}$$

3.1.11 互信息的性质

1. 非负性：$I(X;Y)\geqslant 0$，当且仅当对于任意的 X,Y 独立时等号成立.(由相对熵的非负性可得)
   $I(X;Y\mid Z)\geqslant 0$，即对于条件相对熵非负性也成立.

2. 对称：$I(X;Y)=I(Y;X)$

3. 条件使得互信息减少 (**注意这个性质描述的结论的成立条件**)：若 $X\to Y\to Z$ 构成马尔科夫链，则 $I(X;Y\mid Z)\leqslant I(X;Y)$. 这个性质由数据处理不等式的证明过程直接可得.

$$\begin{aligned} I(X;Y,Z) &= I(X;Y)+I(X;Z\mid Y) \\ &= I(X;Z)+I(X;Y\mid Z) \end{aligned}$$

   由于 $I(X;Z\mid Y)=0$ 和 $I(X;Z)\geqslant 0$ 可得. **注意，这个性质描述的是变量构成马尔科夫链条件下有 $I(X;Y\mid Z)\leqslant I(X;Y)$ 成立**，若不满足马尔科夫链，则可能有 $I(X;Y\mid Z)\geqslant I(X;Y)$. 例如
   X,Y 独立，$Z=X+Y$. 则有 $I(X;Y)=0$ 和

$$I(X;Y\mid Z)=H(X\mid Z)-H(X\mid Y,Z)=H(X\mid Z)\geqslant 0$$

3.1.12 熵和互信息的关系

$$\begin{aligned}
I(X;Y) &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \\
&= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\frac{p(x\mid y)p(y)}{p(x)p(y)} \\
&= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\frac{p(x\mid y)}{p(x)} \\
&= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x\mid y) - \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x) \\
&= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x\mid y) - \sum_{x\in\mathcal{X}} p(x)\log p(x) \\
&= H(X) - H(X\mid Y)
\end{aligned}$$

类似可得

$$I(X;Y) = H(Y) - H(Y\mid X)$$

这里再一次出现**同一个量的不同分解方式！**

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

这些等量关系在等式、不等式证明中有着**重要**的应用.

3.1.13 互信息的链式法则

定义条件互信息

$$\begin{aligned}
I(X;Y\mid Z) &= H(X\mid Z) - H(X\mid Y,Z) \\
&= E_{p(x,y,z)}\log\frac{p(X,Y\mid Z)}{p(X\mid Z)p(Y\mid Z)}
\end{aligned}$$

链式法则

$$\begin{aligned}
I\left(X_1,X_2,\cdots,X_n;Y\right) &= H\left(X_1,X_2,\cdots,X_n\right) - H\left(X_1,X_2,\cdots,X_n\mid Y\right) \\
&= \sum H\left(X_i\mid X_{i-1},\ldots,X_1\right) - \sum H\left(X_i\mid Y,X_{i-1},\ldots,X_1\right) \\
&= \sum_{i=1}^{n} I\left(X_i;Y\mid X_{i-1},\cdots,X_1\right)
\end{aligned}$$

3.1.14 凸函数

对于任意

$$x_1, x_2 \in (a, b), \quad 0 \leqslant \lambda \leqslant 1$$

有

$$f\left(\lambda x_1 + (1-\lambda)x_2\right) \leqslant \lambda f\left(x_1\right) + (1-\lambda) f\left(x_2\right)$$

从几何角度看，这一性质描述的是：对于函数曲线上任意两点，二者之间的函数曲线位于两点连线的下方.

具体来说，$\left(x_1, f\left(x_1\right)\right)$ 和 $\left(x_2, f\left(x_2\right)\right)$ 两点之间连线所在直线方程为

$$y(x) = f\left(x_1\right) + \frac{f\left(x_2\right) - f\left(x_1\right)}{x_2 - x_1} \cdot \left(x - x_1\right)$$

在 $x = (\lambda x_1 + (1-\lambda)x_2)$ 处有

$$\begin{aligned} y &= f\left(x_1\right) + \frac{f\left(x_2\right) - f\left(x_1\right)}{x_2 - x_1} \cdot \left(\left(\lambda x_1 + (1-\lambda)x_2\right) - x_1\right) \\ &= f\left(x_1\right) + (1-\lambda) f\left(x_2\right) - f\left(x_1\right) \\ &= \lambda f\left(x_1\right) + (1-\lambda) f\left(x_2\right) \end{aligned}$$

所以上述性质等价于说明对于函数曲线上任意两点，二者之间的函数曲线位于两点连线的下方.

另一种形象化的记忆方法：**向下凸出的叫凸函数**，向上凸起的叫凹函数.

常见函数 $f(x) = x\log x$ 是凸函数，$f(x) = \log x$ 是凹函数，熵 $H(p)$ 是 p 的凹函数.

3.1.15 Jensen 不等式

对于给定的凸函数 f 和随机变量 X，有

$$Ef(X) \geqslant f(EX)$$

即先求函数值再求期望大于先求期望再求函数值. 关于二者大小关系的记忆，可以按照凸函数定义来记忆.

$$f\left(\lambda x_1 + (1-\lambda)x_2\right) \leqslant \lambda f\left(x_1\right) + (1-\lambda) f\left(x_2\right)$$

实际上可以看出描述的就是先求期望再求函数值 (左边) 小于先求函数值再求期望 (右边).

Jensen 不等式的证明可以使用数学归纳法，这个证明方法比较简单，并且在课程其他地方很少使用这种方法来证明 (一般都是利用链式法则、同一个量的不同分解、熵的

极值性、重要不等式等来证明)，所以这里不详细描述证明方法.

Jensen 不等式在这门课程的证明题目中有着重要应用，需要熟练掌握.

3.1.16 利用 Jensen 不等式证明相对熵的非负性

记 A 为 $p(x)$ 的支撑集，即 $A=\{x:p(x)>0\}$. 则可知 $\sum_A q(x)\leqslant 1$(因为 q 和 p 的支撑集可能不同). 相对熵的非负性证明如下：

$$\begin{aligned}D(p\|q)&=\sum_A p(x)\log\frac{p(x)}{q(x)}\\&=-\sum_A p(x)\log\frac{q(x)}{p(x)}\\&\geqslant-\log\sum_A p(x)\frac{q(x)}{p(x)}\\&=-\log\sum_A q(x)\\&\geqslant 0\end{aligned}$$

3.1.17 利用 Jensen 不等式证明熵的极值性

记 $\mu(x)=\frac{1}{|\mathcal{X}|}$ 服从均匀分布. 则有

$$\begin{aligned}D(p\|\mu)&=\sum p(x)\log\frac{p(x)}{\mu(x)}\\&=\sum p(x)\log p(x)+\sum p(x)\log|\mathcal{X}|\\&=\log|\mathcal{X}|-H(X)\\&\geqslant 0\end{aligned}$$

于是可得

$$H(X)\leqslant\log|\mathcal{X}|$$

当且仅当 X 服从均匀分布时，等号成立.

3.1.18 对数和不等式

对于非负数 $a_i,b_i(i=1,2,\ldots,n)$，有

$$\sum_{i=1}^n a_i\log\frac{a_i}{b_i}\geqslant\left(\sum_{i=1}^n a_i\right)\log\frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i}$$

这个老师说掌握这个不等式就行，应该是利用 Jensen 不等式可以证明. 不过做题和考试都没见过. 我的一个想法是如果感兴趣，想联系一下基于 Jensen 不等式的证明，可以把这个当成一个例题来做，教材上面有证明过程. 至于结论本身，感觉没用过.

3.1.19 马尔科夫链

$$p(x,y,z)=p(x)p(y\mid x)p(z\mid y)$$

则说 X、Y、Z 构成马尔科夫链. 记作

$$X\to Y\to Z$$

描述的是给定 Y 时，X、Z 独立. 注意到这一点，上述关系蕴含

$$Z\to Y\to Y$$

特别地有

$$X\to Y\to g(Y)$$

3.1.20 数学基础——马尔科夫链

如果对任何一列状态 $i_0,i_1,\cdots,i_{n-1},i,j$, 及对任何 $n\geqslant 0$, 随机过程 $\{X_n,n\geqslant 0\}$ 满足 Markov 性质

$$P\{X_{n+1}=j\mid X_0=i_0,\cdots,X_{n-1}=i_{n-1},X_n=i\}=P\{X_{n+1}=j\mid X_n=i\}$$

则称 X_n 为离散时间 Markov 链.

数据处理不等式及其证明方法在这门课程中有着重要的地位，为了学好数据处理不等式，建议读者认真理解马尔科夫链的基本概念及其性质. 简单来说，马尔科夫链的性质可以概括为“给定当前状态，过去与未来无关”.

3.1.21 数据处理不等式

若 $X\to Y\to Z$，则有

$$I(X;Y)\geqslant I(X;Z)$$

其证明方法的思想是：利用互信息的链式法则，构造一个量的不同展开式；进一步利用互信息的非负性和独立时互信息为零，完成证明. 具体过程如下

$$\begin{aligned} I(X;Y,Z) &= I(X;Y) + I(X;Z \mid Y) \\ &= I(X;Z) + I(X;Y \mid Z) \end{aligned}$$

由于 $X \to Y \to Z$ 构成马尔科夫链，有 $I(X;Z \mid Y) = 0$(给定 Y 时，X、Z 独立). 又由于互信息非负性，$I(X;Y \mid Z) \geqslant 0$，所以可得

$$I(X;Y) \geqslant I(X;Z)$$

在这里补充一道例题. 证明：$H(Y \mid X) \leqslant H(Y \mid f(X))$
解析：存在马尔可夫链

$$Y \to X \to f(X)$$

由数据处理不等式知

$$I(X;Y) \geqslant I(f(X);Y)$$

即

$$H(Y) - H(Y \mid X) \geqslant H(Y) - H(Y \mid f(X))$$

于是有

$$H(Y \mid X) \leqslant H(Y \mid f(X))$$

3.1.22 费诺不等式

对于满足 $X \to Y \to \hat{X}$ 的估计量 $\hat{X}$，记 $P_e = p\{X \neq \hat{X}\}$，有

$$H\left(P_e\right) + P_e \log |\mathcal{X}| \geqslant H(X \mid \hat{X}) \geqslant H(X \mid Y)$$

这个不等式在第十章讨论离散均匀分布的信源率失真函数的时候会用得到.

接下来我们讨论这一不等式的证明方法. 如前所述，这门课程中的不等式证明最常用的方法就是构造一个量的不同展开式.

观察这个不等式的形式：两项之和大于等于一项，和之前的一项大于等于另一项不同，这里的情况应该形式化描述为 $A + B = C + D$，其中 $D = 0$，$A \leqslant A'$，$B \leqslant B'$，则由 $A + B = C + D$ 可以得到 $A' + B' \geqslant C$.

另一个需要考虑的问题是，怎样构造这个被展开的量. 根据对应关系，$A' = H\left(P_e\right)$，$B' = P_e \log |\mathcal{X}|$，$C = H(X \mid \hat{X})$. 从 C 出发可以知道这个被展开的量形如 $H(X, Y \mid \hat{X})$. 需要确定的是 Y, 由 A' 可以进一步推断，Y 应该取示性变量. 至此，我们的分析已经结束了 (当然还需要进一步验证).

接下来我们展示详细的证明过程.

首先定义示性变量

$$E=\begin{cases}1, & X\neq\hat{X}\\ 0, & X=\hat{X}\end{cases}$$

接着构造展开式

$$\begin{aligned}H(E,X\mid\hat{X})&=H(X\mid\hat{X})+H(E\mid X,\hat{X})\\&=H(E\mid\hat{X})+H(X\mid E,\hat{X})\end{aligned}$$

由示性变量的定义可知，$H(E\mid X,\hat{X})=0$，又由于

$$H(E\mid\hat{X})\leqslant H(E)=H(P_e)$$

和

$$\begin{aligned}H(X\mid E,\hat{X})&=p(E=1)H(X\mid E=1,\hat{X})+p(E=0)H(X\mid E=0,\hat{X})\\&\leqslant p(E=1)\log|\mathcal{X}|\\&=P_e\log|\mathcal{X}|\end{aligned}$$

所以可得

$$H\left(P_e\right)+P_e\log|\mathcal{X}|\geqslant H(X\mid\hat{X})$$

而至于后半部分不等式，利用数据处理不等式、互信息和熵的关系即可得证. 这里不作详细描述.

费诺不等式还可以继续放缩得到

$$1+P_e\log|\mathcal{X}|\geqslant H(X\mid Y)$$

进而得到

$$P_e\geqslant\frac{H(X\mid Y)-1}{\log|\mathcal{X}|}$$

第 3 章　渐进均分性

4.1　主要内容

这一章的内容比较少，而且直接考察的题目并不多，从见过的两份试卷上看并没有特别对这一章进行考察. 但是这一章介绍的渐进均分性定理、典型集的概念有着重要的意义. 渐进均分性定理的一个重要意义在于，作为数据压缩的基础，也可以帮助我们理解数据压缩的原理. 而这里介绍的典型集的概念可以帮助我们理解联合典型序列的概念，以及联合典型译码 (信道编码定理证明中使用联合典型译码).

这一章主要需要掌握

4.1.1　依概率收敛

称 X_n 依概率收敛到 X，若对于任意的 $\varepsilon > 0$，有

$$p\{|X_n - X| > \varepsilon\} \to 0$$

在这一章节我们研究的都是依概率收敛.

4.1.2　弱大数定理

若 $X_1, X_2, \ldots, X_n$ i.i.d.$\sim p(x)$，当 n 足够大时有

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\text{依概率}} EX$$

4.1.3 渐进均分性定理

若 $X_1, X_2, \ldots, X_n$ i.i.d.$\sim p(x)$，当 n 足够大时有

$$\begin{aligned} -\frac{1}{n}\log p\left(X_1, X_2, \ldots, X_n\right) &= -\frac{1}{n}\log \prod p\left(X_i\right) \\ &= -\frac{1}{n}\sum_{i=1}^{n}\log p\left(X_i\right) \\ &\xrightarrow{\text{依概率}} E\left[-\log p\left(X\right)\right] \\ &= H(X) \end{aligned}$$

4.1.4 典型集

关于分布 $p(x)$ 的典型集定义 $A_\varepsilon^{(n)}$ 为序列 $x_1, x_2, \ldots, x_n \in \mathcal{X}^n$ 的集合，且满足

$$2^{-n(H(X)+\varepsilon)} \leqslant p\left(x_1, x_2, \cdots, x_n\right) \leqslant 2^{-n(H(X)-\varepsilon)}$$

注意这一个概念，并没有要求 $n \to \infty$，也没有要求 $\varepsilon \to 0$. 对于任意给定的 n 和 ε，都可以讨论典型集.

4.1.5 典型集的性质

1. 和渐进均分性定理比较：对于典型集中的元素 $(x_1, x_2, \cdots, x_n) \in A_\varepsilon^{(n)}$，有

$$H(X) - \varepsilon \leqslant -\frac{1}{n}\log p\left(x_1, \cdots, x_n\right) \leqslant H(X) + \varepsilon$$

2. 当 n 充分大时，以非零概率出现的元素 (几乎) 都是典型集中的元素，即

$$p\left\{A_\varepsilon^{(n)}\right\} > 1 - \varepsilon$$

这个由渐进均分性定理和典型集的概念可得.

3. 典型集元素个数上界

$$\left|A_\varepsilon^{(n)}\right| \leqslant 2^{n(H(X)+\varepsilon)}$$

这个性质的证明只需要利用典型集的概念，即

$$\begin{aligned}1 &= \sum_{x^n \in \mathcal{X}^n} p(x^n) \\ &\geqslant \sum_{x^n \in A_\varepsilon^{(n)}} p(x^n) \\ &\geqslant \sum_{x^n \in A_\varepsilon^{(n)}} 2^{-n(H(X)+\varepsilon)} \\ &= 2^{-n(H(X)+\varepsilon)} \left|A_\varepsilon^{(n)}\right|\end{aligned}$$

4. 典型集元素个数**在 n 足够大时的下界**：当 n 足够大时有

$$\left|A_\varepsilon^{(n)}\right| \geqslant (1-\varepsilon)2^{n(H(X)-\varepsilon)}$$

这个性质的证明需要利用典型集的概念和典型集的第 2 条性质，即

$$\begin{aligned}1-\varepsilon < p\left(x^n \in A_\varepsilon^{(n)}\right) &= \sum_{x^n \in A_\varepsilon^{(n)}} p(x^n) \\ &\leqslant \sum_{x^n \in A_\varepsilon^{(n)}} 2^{-n(H(X)-\varepsilon)} \\ &= 2^{-n(H(X)-\varepsilon)} \left|A_\varepsilon^{(n)}\right|\end{aligned}$$

第 4 章　随机过程的熵率

5.1　主要内容

第 4 章在考试中一般会有一道大题，以及可能会有一道小题. 大题常见的考察方法就是给出一个马尔科夫链，要求计算熵率 (这类题目的易错点在于，马尔科夫链的熵率计算公式要求在平稳分布下计算，所以一般题目给出的初始状态都是迷惑性的条件，并不能直接使用).

这一章主要需要掌握

5.1.1　平稳过程

对于任意的序列长度 n 和“偏移”l，恒有

$$\begin{aligned}&p\left(X_1=x_1, X_2=x_2, \ldots, X_n=x_n\right)\\&\quad=p\left(X_{1+l}=x_1, X_{2+l}=x_2, \ldots, X_{n+l}=x_n\right)\end{aligned}$$

5.1.2　马尔科夫链

1. 定义

$$\begin{aligned}&p\left(X_{n+1}=x_{n+1} \mid X_n=x_n, X_{n-1}=x_{n-1}, \ldots, X_1=x_1\right)\\&\quad=p\left(X_{n+1}=x_{n+1} \mid X_n=x_n\right)\end{aligned}$$

2. 表示方法

一个时间不变的马尔可夫链完全由其初始状态和概率转移矩阵 $P=\left(P_{ij}\right)$ 表征，其中

$$P_{ij}=p\left(X_{n+1}=j \mid X_n=i\right)$$

3. 不可约

若马尔可夫链可以从**任意状态**经过有限步转移到**另一任意状态**，且转移概率为正，则称此马尔可夫链是不可约的.

4. 非周期

   如果从一个状态转移到它自身的不同路径长度的最大公因子为 1，则称该马尔可夫链是非周期的.

5. 时间不变

   对于任意的 $a,b \in \mathcal{X}$(字母表)，有

$$p\left(X_{n+1}=b \mid X_n=a\right)=p\left(X_2=b \mid X_1=a\right)$$

   说明，平稳过程一定是时间不变的，反之不一定成立.

6. 平稳分布

   若在 $n+1$ 时刻状态空间的分布与 n 时刻的分布相同，则称此分布为平稳分布.

7. 平稳马尔可夫过程

   若马尔可夫链的初始状态服从平稳分布，则该马尔可夫链为平稳过程.

8. 平稳分布存在性

   若**有限状态**马尔可夫链是**不可约的和非周期**的，则它的**平稳分布惟一**，从**任意**的初始分布出发，当 n 趋向于无穷时，X_n 的分布必趋向于此平稳分布.

9. 马尔科夫链性质的概率描述

   由条件概率定义式可得

$$\begin{aligned}
p\left(x_1,\ldots,x_1\right) &= p\left(x_1\right)\cdot p\left(x_2,\cdots,x_n \mid x_1\right)\\
&= p\left(x_1\right)\cdot p\left(x_2 \mid x_1\right)\cdot p\left(x_3,\ldots x_1 \mid x_1,x_2\right)\\
&= p\left(x_1\right)\cdot p\left(x_2 \mid x_1\right)\cdot p\left(x_3 \mid x_1,x_2\right)\cdot p\left(x_4,\ldots x_n \mid x_1,x_2,x_3\right)\\
&= \cdots\\
&= p\left(x_1\right)\cdot p\left(x_2 \mid x_1\right)\cdot p\left(x_3 \mid x_1,x_2\right)\cdots p\left(x_n \mid x_1,x_2,\cdots x_{n-1}\right)\\
&= \prod_{i=1}^{n} p\left(x_i \mid x_{i-1}\right)
\end{aligned}$$

   由马尔科夫链性质可知

$$p\left(x_3 \mid x_1,x_2\right)=p\left(x_3 \mid x_2\right)$$

   因为

$$\begin{aligned}
p\left(x_3,x_1 \mid x_2\right) &= p\left(x_3 \mid x_2\right)\cdot p\left(x_1 \mid x_2\right)\\
&= p\left(x_3 \mid x_1,x_2\right)\cdot p\left(x_1 \mid x_2\right)
\end{aligned}$$

其中上式第一行是由马尔科夫链条件独立性质得到，第二行是由条件概率定义式得到.

类似地，我们可得

$$p\left(x_n \mid x_1 \ldots x_{n-1}\right)=p\left(x_n \mid x_{n-1}\right)$$

于是有

$$p\left(x_1 \cdots x_n\right)=p\left(x_1\right) \cdot p\left(x_2 \mid x_1\right) \cdots p\left(x_n \mid x_{n-1}\right)$$

这就是马尔科夫链性质的一种常用的概率描述，建议读者熟练掌握.

5.1.3 熵率

当极限存在时，随机过程的熵率定义为

$$H(\mathscr{X})=\lim_{n\to\infty}\frac{1}{n}H\left(X_1, X_2, \ldots, X_n\right)$$

为了方便计算熵率，定义了另一个量

$$H'(\mathscr{X})=\lim_{n\to\infty}H\left(X_n \mid X_{n-1}, X_{n-2}, \ldots, X_1\right)$$

并且有，对于**平稳的随机过程 (实际上，要求存在平稳分布即可**，所以有限状态、不可约、非周期的马尔科夫链可以直接应用这个进行计算熵率)，有二者**存在且相等**

$$H(\mathscr{X})=H'(\mathscr{X})$$

这门课在这一章节主要讨论有限状态、不可约、非周期的马尔科夫链，其熵率计算方法为

$$H(\mathscr{X})=H'(\mathscr{X})=\lim H\left(X_n \mid X_{n-1}, \ldots, X_1\right)=\lim H\left(X_n \mid X_{n-1}\right)$$

即，记马尔科夫链的**平稳分布**为 μ，转移概率矩阵为 P，有

$$H(\mathscr{X})=-\sum \mu_i P_{ij} \log P_{ij}$$

这里特别注意，**一定要用平稳分布计算**. 我们说初始状态和转移概率矩阵可以直接表征马尔科夫链，所以题目经常会给出初始状态和转移概率矩阵. 但是在计算熵率时，要求在平稳分布时进行计算，所以题目中的初始状态往往有迷惑性，导致易错.

5.1.4 加权图上随机移动的熵率

这个问题在作业中出现，需要直接利用结论.

简要描述如下

1. 节点集合：$\{1, 2, \ldots, m\}$
2. 边权重：$W_{ij} \geqslant 0$，且有 $W_{ij} = W_{ji}$. 如果节点之间没有边，$W_{ij} = 0$
3. 随机过程描述：$\{X_n\}$，其中 $X_n \in \{1, 2, \ldots, m\}$
4. 转移概率：$P_{ij} = \frac{W_{ij}}{\sum_k W_{ik}}$
5. **平稳分布** (记住结论，可以从直观上理解其含义)：

$$\mu_i = \frac{W_i}{2W}$$

   其中 $W_i = \sum_j W_{ij}$，$W = \sum_{i,j:j\geqslant i} W_{ij}$

6. **熵率**：

$$H\left(\cdots, \frac{W_{ij}}{2W}, \cdots\right) - H\left(\cdots, \frac{W_i}{2W}, \cdots\right)$$

5.1.5 把孤立系统看作马尔科夫链

这里提出主要是通过证明课件上提出的几个结论，强调一些证明信息不等式的方法. 另外这种把孤立系统看作马尔科夫链的想法，在处理判断题的时候可以提供一些思路.

1. 两个分布之间的相对熵随时间 n 递减

   由于是同一个孤立系统，两个不同的分布受到的影响相同. 形式化描述为 (对于任意的 p, q)

$$p(x_{n+1} \mid x_n) = q(x_{n+1} \mid x_n)$$

   于是有

$$\begin{aligned} D\left(p\left(x_n, x_{n+1}\right) \| q\left(x_n, x_{n+1}\right)\right) &= D\left(p(x_n) \| q(x_n)\right) + D\left(p(x_{n+1} \mid x_n) \| q(x_{n+1} \mid x_n)\right) \\ &= D\left(p(x_n) \| q(x_n)\right) \\ &= D\left(p(x_{n+1}) \| q(x_{n+1})\right) + D\left(p(x_n \mid x_{n+1}) \| q(x_n \mid x_{n+1})\right) \end{aligned}$$

   这样由于相对熵非负 $D\left(p(x_n \mid x_{n+1}) \| q(x_n \mid x_{n+1})\right) \geqslant 0$，就可以得到

$$D\left(p(x_n) \| q(x_n)\right) \geqslant D\left(p(x_{n+1}) \| q(x_{n+1})\right)$$

   这里的证明方法的思想仍然是，**利用同一个量的不同分解**完成证明.

2. 平稳马尔可夫过程，条件熵 $H(X_n \mid X_1)$ 随 n 增加

$$\begin{aligned} H(X_n \mid X_1) &\geqslant H(X_n \mid X_2, X_1) \\ &= H(X_n \mid X_2) \\ &= H(X_{n-1} \mid X_1) \end{aligned}$$

这种条件熵递增、递减问题，往往是通过“**增加条件**”使得熵减少来完成证明.

3. 洗牌使得熵增加 $H(TX) \geqslant H(X)$

注意到，这里所说的洗牌是指一个可逆变换.

$$\begin{aligned} H(TX) &\geqslant H(TX \mid T) \\ &= H(T^{-1}TX \mid T) \\ &= H(X \mid T) \\ &= H(X) \end{aligned}$$

这里同样是通过“**增加条件**”使得熵减少来完成证明.

第 5 章 数据压缩

6.1 主要内容

这一章是课程的重点之一. 这一章主要讨论离散情形的信源编码理论，介绍了即时码存在的判定准则——Kraft 不等式，即时码最优码长的大小，哈夫曼编码、香农码、Shannon-Fano-Elias 编码等. 考察时候往往会出现，利用 Kraft 不等式判定给定的码长序列能否构成即时码，给定概率分布进行哈夫曼编码，基于哈夫曼编码的码元个数特性解题等.

这一章主要需要掌握

6.1.1 信源编码的期望长度

$$L(C)=\sum_{x\in\mathcal{X}}p(x)l(x)$$

6.1.2 前缀码等几个基本概念

1. 非奇异编码

$$x\neq x'\Rightarrow C(x)\neq C\left(x'\right)$$

2. 扩展编码

$$C\left(x_1x_2\cdots x_n\right)=C\left(x_1\right)C\left(x_2\right)\cdots C\left(x_n\right)$$

3. 唯一可译码：扩展编码非奇异

4. 前缀码：码字空间中，没有任何码字是其他码字的前缀，也叫即时码 (指的是，可以即时译码，因为看到“前缀”即可唯一确定码字). **这一章主要讨论即时码** (实用价值更高)

6.1.3 Kraft 不等式

对于 D 元字母表上的即时码，码字长度 $l_1, l_2, \ldots, l_n$ 必须满足

$$\sum_i D^{-l_i} \leqslant 1$$

这个不等式是即时码存在的充要条件，也是即时码的必要条件. 在考试中，往往会出现一道题目，给出码长序列要求判定是否可以构成即时码，即直接应用 Kraft 不等式即可.

6.1.4 最优码长的界

给定信源分布 p 和一个 D 元字母表，最优码长的期望满足

$$H_D(X) \leqslant L^* < H_D(X) + 1$$

6.1.5 香农码

编码方案：计算码长

$$l(x) = \left\lceil \log \frac{1}{p(x)} \right\rceil$$

基于此构造码树，取叶子节点对应的路径上的比特串作为编码.

6.1.6 哈夫曼编码——一种最优编码

1. 编码方案：(首先检查是否需要增加零概率项) 对所有码字的出现概率进行排序，每次取最小的两个节点添加父节点 (父节点权重为两个子节点权重和) 然后重新排序，直到所有节点加入码树. 根据码树确定编码.

2. 特性：码元个数满足

$$|\mathcal{X}| = 1 + k(D-1)$$

3. **注意事项**：哈夫曼编码前，首先检查码字个数是否满足第 2 条的要求. 如果不满足，需要增加零概率项，直到满足后再正常进行编码.

6.1.7 Shannon-Fano-Elias 编码

1. 编码方案：首先计算修正的累计分布函数

$$\bar{F}(x) = \sum_{a<x} p(a) + \frac{1}{2} p(x)$$

然后取 $\bar{F}(x)$ 的前 $l(x)$ 位 $\lfloor \bar{F}(x) \rfloor_{l(x)}$ 作为 x 的编码，其中

$$l(x) = \left\lceil \log \frac{1}{p(x)} \right\rceil + 1$$

2. 特点：是前缀码，期望长度 (比香农码大 1)

$$L < H(X) + 2$$

3. 证明是前缀码 (重要)：修正累积分布函数相当于区间中点，而

$$\bar{F}(x) - \lfloor \bar{F}(x) \rfloor_{l(x)} < \frac{1}{2^{l(x)}} \leqslant \frac{p(x)}{2}$$

说明 $\lfloor \bar{F}(x) \rfloor_{l(x)}$ 落在区间 $[F(x-1), F(x)]$ 内，也就是取前 $l(x)$ 位后减小的量不超过区间长度一半，这样每个区间中只有一个元素被编码，所以是前缀码.

6.1.8 三种编码比较

1. Shannon-Fano-Elias 编码：不需要构造码树，实现效率较高；不需要排序；期望码长较长
2. 香农码：需要构造码树；不需要排序；期望码长中等
3. 哈夫曼编码：需要构造码树；需要排序；期望码长最短

第 7 章　信道容量

7.1　主要内容

这一章是这门课程的重点. 这一章主要介绍信道编码理论，重点考察信道容量计算. 这一章主要需要掌握

7.1.1　离散无记忆信道

由输入字母表 $\mathcal{X}$、输出字母表 $\mathcal{Y}$ 和概率转移矩阵 $p(y \mid x)$ 构成. 任意时刻的输出仅取决于对应的输入，与历史输入无关.(类比组合逻辑电路)

7.1.2　离散无记忆信道的信道容量

$$C = \max_{p(x)} I(X;Y)$$

7.1.3　信道容量计算

信道容量计算是这一章的重要内容，也是考试中的重要考点. 一般会有一道大题专门计算离散无记忆信道的信道容量，而在判断、填空中也可能会出现.

信道容量计算步骤为

1. 互信息分解 (写出熵和条件熵做差形式)
2. 根据信道本身确定条件熵的取值
3. 分析熵的最大值 (注意要求能够取到，这个是**易错点**)
4. 给出熵取到最大值的条件 (输入概率分布 $p(x)$)(这个是**容易忽略**的)
5. 写上单位——比特/信道使用 (**易错**)

下面我们详细地分析信道容量求解的过程.

1. 互信息分解

   信道容量计算是基于互信息的分解

$$I(X;Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X)$$

   在实际解题中应该选择哪种分解呢? 我们的答案是，如果信道满足给定输出 Y 后 X 确定 $(H(X \mid Y) = 0)$，则选择分解

$$I(X;Y) = H(X) - H(X \mid Y)$$

   否则，我们选择分解 (**最常用的一种分解**)

$$I(X;Y) = H(Y) - H(Y \mid X)$$

2. 第一种分解方案 (少数情况会使用)

   选择使用分解

$$I(X;Y) = H(X) - H(X \mid Y)$$

   并且注意到使用这种分解的前提是 $H(X \mid Y) = 0$. 所以此时互信息最大值求解相当于 $H(X)$ 最大值的求解. 利用熵的极值性即可得

$$C = \max_{p(x)} I(X;Y) = \max_{p(x)} H(X) = \log |\mathcal{X}|$$

   此时取得最大值对应输入的分布就是均匀分布

$$p(x) = \frac{1}{|\mathcal{X}|}$$

3. 第二种分解方案 (最常用)

   选择使用分解

$$I(X;Y) = H(Y) - H(Y \mid X)$$

   条件熵由信道决定. 求解条件熵的技巧是将其写作

$$H(Y \mid X) = \sum_{x \in \mathcal{X}} p(x) H(Y \mid X = x)$$

   根据信道，观察每个输入对应的输出概率分布，即可求解得到条件熵. 接下来的重要任务是求解 $H(Y)$ 的最大值.

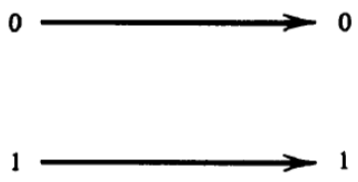
   我们知道熵的极值性，当服从均匀分布时熵取得最大值. 但这里要注意的问题是，

输出能否服从均匀分布. 所以，一般情况下，**我们的选择是：**

如果信道有“对称性”(信道概率转移矩阵的每一行都是另一行的一个置换，在不考虑取值顺序的情况下是相同的分布)，即当输入服从均匀分布时输出也服从均匀分布，则可以确定当输入服从均匀分布时，达到信道容量；如果信道不具有“对称性”，则需要先对输入的概率分布作一般的假设 $p(x)$，然后表达输出的概率分布，进而计算 $H(Y)$(是输入 $p(x)$ 的函数)，则求导可得最大值.

综上所述，信道容量求解的题目主要分为三类. 我们接下来以三个例子分别展示每种情况的信道容量计算过程.

7.1.4 信道容量计算第一种类型——无噪声二元信道



使用第一种分解

$$I(X;Y) = H(X) - H(X \mid Y)$$

由于给定输出时输入是确定的，所以 $H(X \mid Y) = 0$. 则

$$I(X;Y) = H(X)$$

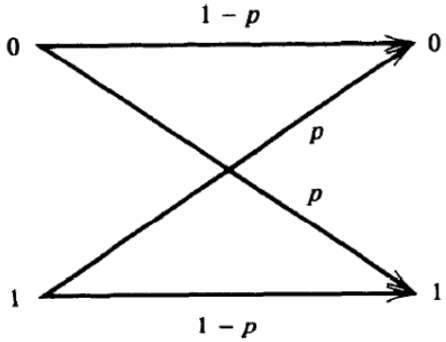
由熵的极值性，输入服从均匀分布时，熵 $H(X)$ 取得最大值. 所以

$$C = \max_{p(x)} I(X;Y) = \max_{p(x)} H(X) = \log|\mathcal{X}| = 1 \quad \text{比特/信道使用}$$

达到信道容量时的输入概率分布为

$$p(X=0) = p(X=1) = \frac{1}{2}$$

7.1.5 信道容量计算第二种类型——二元对称信道



由于输出给定时并不能唯一确定输入 (绝大多数情况都是这样)，所以使用第二种分解

$$I(X;Y) = H(Y) - H(Y \mid X)$$

根据信道计算条件熵. 首先将条件熵写作

$$H(Y \mid X) = \sum_{x \in \mathcal{X}} p(x) H(Y \mid X = x)$$

然后根据输入字母表种每个取值对应的输出概率分布，即可计算条件熵

$$\begin{aligned} H(Y \mid X) &= \sum_{x \in \mathcal{X}} p(x) H(Y \mid X = x) \\ &= p(X = 0) H(Y \mid X = 0) + p(X = 1) H(Y \mid X = 1) \\ &= p(X = 0) H(p) + p(X = 1) H(p) \\ &= H(p) \end{aligned}$$

接着计算熵 $H(Y)$ 的最大值. 对于这个信道，由于其具有对称性 (转移概率矩阵每一行都是相同的分布——不考虑取值顺序)，所以输入服从均匀分布时输出服从均匀分布. 这样就可以确定，输入服从均匀分布时，熵 $H(Y)$ 取得最大值

$$\max_{p(x)} H(Y) = \log |\mathcal{Y}| = \log 2 = 1 \quad \text{比特}$$

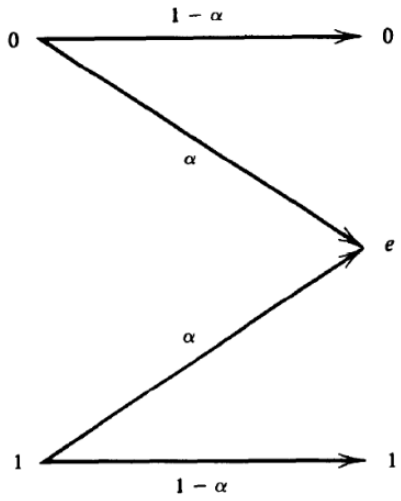
所以，信道容量为

$$C = \max_{p(x)} I(X;Y) = \max_{p(x)} H(Y) - H(Y \mid X) = 1 - H(p) \quad \text{比特/信道使用}$$

达到信道容量时的输入概率分布为

$$p(X=0)=p(X=1)=\frac{1}{2}$$

7.1.6 信道容量计算第三种类型——二元擦除信道



由于输出给定时并不能唯一确定输入 (绝大多数情况都是这样)，所以使用第二种分解

$$I(X;Y)=H(Y)-H(Y\mid X)$$

根据信道计算条件熵. 首先将条件熵写作

$$H(Y\mid X)=\sum_{x\in\mathcal{X}}p(x)H(Y\mid X=x)$$

然后根据输入字母表种每个取值对应的输出概率分布，即可计算条件熵

$$\begin{aligned}H(Y\mid X)&=\sum_{x\in\mathcal{X}}p(x)H(Y\mid X=x)\\&=p(X=0)H(Y\mid X=0)+p(X=1)H(Y\mid X=1)\\&=p(X=0)H(\alpha)+p(X=1)H(\alpha)\\&=H(\alpha)\end{aligned}$$

接着计算熵 $H(Y)$ 的最大值. 对于这个信道，由于其非对称，所以需要先对输入的概率分布作一般的假设 $p(x)$

$$p(X=0)=p,\quad p(X=1)=1-p$$

则输出的概率分布为

$$\begin{aligned}
p(Y=0) &= p(X=0)(1-\alpha) = p(1-\alpha) \\
p(Y=e) &= p(X=0)\alpha + p(X=1)\alpha = \alpha \\
p(Y=1) &= p(X=1)(1-\alpha) = (1-p)(1-\alpha)
\end{aligned}$$

进而可以计算输出分布的熵

$$\begin{aligned}
H(Y) &= \sum_{y\in\mathcal{Y}} -p(y)\log p(y) \\
&= -p(1-\alpha)\log\left[p(1-\alpha)\right] - \alpha\log\alpha - (1-p)(1-\alpha)\log\left[(1-p)(1-\alpha)\right] \\
&= -p(1-\alpha)\log p - p(1-\alpha)\log(1-\alpha) - \alpha\log\alpha - (1-p)(1-\alpha)\log(1-p) \\
&\quad - (1-p)(1-\alpha)\log(1-\alpha) \\
&= (1-\alpha)H(p) + H(\alpha)
\end{aligned}$$

所以，可得信道容量

$$C = \max_{p(x)} I(X;Y) = \max_{p(x)} H(Y) - H(Y\mid X) = \max_{p(x)} (1-\alpha)H(p) = (1-\alpha) \quad \text{比特/信道使用}$$

达到信道容量时的输入概率分布为

$$p(X=0) = p(X=1) = \frac{1}{2}$$

以上三个小节介绍的三种情形包括了这一部分所有可能出现的题目类型.

7.1.7 对称信道

信道转移概率矩阵 $p(y\mid x)$ 的任意两行互相置换，任意两列互相置换.(我们主要讨论弱对称信道，当然强对称信道也满足弱对称信道的要求，所以可以用相同的公式进行信道容量计算)

7.1.8 弱对称信道

信道转移概率矩阵 $p(y\mid x)$ 的任意两行互相置换，且所有列的元素和 $\sum_x p(y\mid x)$ 相等.

7.1.9 弱对称信道的信道容量计算

对于信道进行分析，由于输出给定时并不能唯一确定输入 (绝大多数情况都是这样)，所以使用第二种分解

$$I(X;Y) = H(Y) - H(Y \mid X) \leqslant \log(\mathcal{Y}) - H(\text{转移概率矩阵的行})$$

这里是假设输出均匀分布能够取到的，我们需要验证这个假设是否成立.

$$p(y) = \sum_{x\in\mathcal{X}} p(x)p(y \mid x)$$

当 X 服从均匀分布，即 $p(x) = \frac{1}{|\mathcal{X}|}$ 时，记转移概率矩阵列的元素和为 c(常数)，有

$$\begin{aligned} p(y) &= \sum_{x\in\mathcal{X}} p(x)p(y \mid x) \\ &= \frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}} p(y \mid x) \\ &= \frac{c}{|\mathcal{X}|} \end{aligned}$$

即 Y 服从均匀分布，熵的最大值 $\log(\mathcal{Y})$ 能够达到.

综上所述，当输入分布为均匀分布 $p(x) = \frac{1}{|\mathcal{X}|}$ 时，达到信道容量，为

$$C = \log(\mathcal{Y}) - H(\text{转移概率矩阵的行}) \quad \text{比特/信道使用}$$

7.1.10 并联信道的容量——重要结论

并联信道的信道容量计算一方面可以作为重要结论应用于解题，另一方面其求解过程中使用的重要技巧建议掌握，在其他问题求解中也会有应用.

我们首先讨论两个信道并联的问题. 第一个信道的输入字母表为 $\mathcal{X}_1$，输出字母表为 $\mathcal{Y}_1$；第二个信道的输入字母表为 $\mathcal{X}_2$，输出字母表为 $\mathcal{Y}_2$；其中 $\mathcal{X}_1$ 和 $\mathcal{X}_2$ 交集为空，$\mathcal{Y}_1$ 和 $\mathcal{Y}_2$ 交集也为空 (就是两个“完全独立”的信道按照某一概率分布将输入组合在一起). 记第一个信道的信道容量为 C_1，第二个信道的信道容量为 C_2，需要求解并联后的信道容量 C.

两个信道组合方式可以描述为

$$X = \begin{cases} X_1, & \text{概率}\alpha \\ X_2, & \text{概率}(1-\alpha) \end{cases}$$

对于这种输入的随机变量按照一定概率分布组合在一起的问题，一般的处理技巧是引入

示性变量. 记

$$\theta(X)=\begin{cases}1, & X=X_1\\ 2, & X=X_2\end{cases}$$

引入示性变量的目的是通过构造互信息的分解，简化 $I(X;Y)$ 的表达，从而完成最大值求解.

$$\begin{aligned}I(X;Y,\theta)&=I(X;\theta)+I(X;Y\mid\theta)\\&=I(X;Y)+I(X;\theta\mid Y)\end{aligned}$$

由于给定了输入取值 ($X=X_1$ 或 $X=X_2$) 后输出 $Y=Y_1$ 还是 $Y=Y_2$ 是确定的，即 $\theta=g(Y)$. 所以构成马尔科夫链

$$X\to Y\to\theta$$

于是有 $I(X;\theta\mid Y)$=0. 则可以得到互信息的表达

$$\begin{aligned}I(X;Y)&=I(X;\theta)+I(X;Y\mid\theta)\\&=H(\theta)-H(\theta\mid X)+\alpha I\left(X_1;Y_1\right)+(1-\alpha)I\left(X_2;Y_2\right)\\&=H(\alpha)+\alpha I\left(X_1;Y_1\right)+(1-\alpha)I\left(X_2;Y_2\right)\end{aligned}$$

对上式进行求导可得

$$H'(\alpha)+C_1-C_2=0$$

其中

$$\begin{aligned}H'(\alpha)&=-\log\alpha-\alpha\frac{1}{\alpha\ln 2}+\log(1-\alpha)+(1-\alpha)\frac{1}{(1-\alpha)\ln 2}\\&=\log\frac{1-\alpha}{\alpha}\end{aligned}$$

则可得

$$\log\frac{1-\alpha}{\alpha}=C_2-C_1$$

进而可得

$$\alpha=\frac{2^{C_1}}{2^{C_1}+2^{C_2}}$$

所以可得信道容量

$$\begin{aligned}C&=H\left(\frac{2^{C_1}}{2^{C_1}+2^{C_2}}\right)+\frac{2^{C_1}}{2^{C_1}+2^{C_2}}C_1+\frac{2^{C_2}}{2^{C_1}+2^{C_2}}C_2\\&=-\frac{2^{C_1}}{2^{C_1}+2^{C_2}}\log\frac{2^{C_1}}{2^{C_1}+2^{C_2}}-\frac{2^{C_2}}{2^{C_1}+2^{C_2}}\log\frac{2^{C_2}}{2^{C_1}+2^{C_2}}+\frac{2^{C_1}}{2^{C_1}+2^{C_2}}C_1+\frac{2^{C_2}}{2^{C_1}+2^{C_2}}C_2\\&=\frac{1}{2^{C_1}+2^{C_2}}[-2^{C_1}\log 2^{C_1}+2^{C_1}\log(2^{C_1}+2^{C_2})-2^{C_2}\log 2^{C_2}+2^{C_2}\log(2^{C_1}+2^{C_2})\\&\quad+2^{C_1}C_1+2^{C_2}C_2]\\&=\log\left(2^{C_1}+2^{C_2}\right)\end{aligned}$$

达到信道容量时的输入概率分布为

$$X=\begin{cases} X_1 & \text{概率}\frac{2^{C_1}}{2^{C_1}+2^{C_2}} \\ X_2 & \text{概率}\frac{2^{C_2}}{2^{C_1}+2^{C_2}} \end{cases}$$

而其中 X_1 和 X_2 的概率分布为使得其对应信道达到信道容量的概率分布.
也可以做一个简单的变形得到一种容易记忆的形式

$$2^C=2^{C_1}+2^{C_2}$$

我们可以发现，这样的形式可以简单地进行推广，即对于若干个满足上述要求的“独立”信道并联，并联后的信道容量满足

$$2^C=\sum_{i=1}^{n}2^{C_i}$$

这个关于并联信道的信道容量的结论建议掌握.

7.1.11 码率

1. 定义：(M,n) 码的码率为

$$R=\frac{\log M}{n}\quad \text{比特/传输}$$

2. 可达：存在 $(\lceil 2^{nR}\rceil,n)$ 码序列，满足当 $n\to\infty$ 时，最大误差概率 $\lambda^n\to 0$

3. 码率和信道容量关系：**信道容量是所有可达码率的上界**

7.1.12 反馈对于离散无记忆信道的影响

反馈容量满足

$$C_{FB}=C=\max_{p(x)}I(X;Y)$$

即，反馈不改变信道容量.(不过这个结论对于连续情形就不再成立)

第 8 章　微分熵

8.1　主要内容

这一章是这门课程的重点. 这一章主要介绍了连续随机变量的信息度量——微分熵，以及相应的联合微分熵、条件微分熵、相对熵、互信息的概念、性质、链式法则等. 一般的考察形式就是给出一个连续分布，进行相关量的计算. **在这一章的学习中，建议和离散熵部分比较学习，关注二者在概念、性质、链式法则等方面的异同**
这一章主要需要掌握

8.1.1　微分熵

1. 定义：连续随机变量的概率密度函数记为 $f(x)$，微分熵为

$$h(X) = -\int_S f(x)\log f(x)dx$$

注意，微分熵的表示用小写的 $h(X)$，而离散的熵的表达用 $H(X)$. 这里也是**易错点**.

2. 常用的微分熵：

一维正态分布 (熵的计算不受均值影响)

$$X \sim \mathcal{N}\left(0, \sigma^2\right), \quad \phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$

其微分熵为

$$\begin{aligned} h(\phi) &= -\int_S \phi \ln \phi dx \\ &= -\int_S \phi \left(-\frac{x^2}{2\sigma^2} - \ln \sqrt{2\pi\sigma^2} \right) dx \\ &= \frac{E[X^2]}{2\sigma^2} + \frac{1}{2} \log 2\pi\sigma^2 \\ &= \frac{1}{2} \ln 2\pi e\sigma^2 \quad \text{奈特} \\ &= \frac{1}{2} \log 2\pi e\sigma^2 \quad \text{比特} \end{aligned}$$

这里在求解过程中，应用了一些技巧.

首先处理正态分布的微分熵的时候，对数函数外边的概率密度函数保持不变 (之后会把积分表示成期望，方便求解)，并且我们习惯上先写成 $x \ln x$ 形式，即以奈特为单位，这样做是为了简化对数运算 (因为正态分布的密度函数中有 e 指数). 注意到这里得到以奈特为单位的结果后只需要把对数函数换成以 2 为底即可，不需要做其他改变.

另一个技巧是，利用概率论的知识. 记 X 的概率密度函数为 $f(x)$，则 $g(X)$ 的概率密度函数为

$$E[g(X)] = -\int_S g(x) f(x) dx$$

在应用的时候，这个结论往往是反着应用的，即在求解微分熵的时候把其中一部分符合等号右边的用等号左边进行替换，进而可以利用随机变量的期望、方差等简洁地得到结果.

3. 微分熵和离散熵

$$H\left(X^{\Delta}\right) + \log \Delta \to h(f) = h(X), \quad \text{当} \quad \Delta \to 0$$

和离散熵不同，微分熵的取值不代表信源 X 的不确定度，不具有直接表示信息度量的能力.

8.1.2 微分熵求解的小技巧

1. 不要急着把概率密度函数代入，先就写成 $f(x)$ 形式，然后对后面做对数运算的部分先计算，概率密度函数和后面对数运算的结果 $g(x)$ 的积分可以表达成期望的形式

$$\int_{-\infty}^{+\infty} f(x)g(x)\ \mathrm{dx} = E[g(X)]$$

2. 对于概率密度函数中含有 e 指数项的，建议计算微分熵的时候先用奈特单位，再

化回比特 (结果中的对数的底由 e 换成 2)，即计算微分熵的时候首先

$$h(X) = -\int f \ln f$$

8.1.3 联合微分熵

1. 定义：服从联合概率密度函数 $f(x_1, x_2, \ldots, x_n)$ 的随机变量的联合微分熵定义为

$$h\left(X_1, X_2, \cdots, X_n\right) = -\int f\left(x^n\right) \log f\left(x^n\right) dx^n$$

2. 常用的联合微分熵：

多元正态分布 $\mathcal{N}_n(\mu, K)$ 的概率密度函数为

$$f\left(x_1, \ldots, x_n\right) = \frac{1}{\sqrt{(2\pi)^n |K|}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}} K^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

微分熵为

$$\begin{aligned}
h\left(X_1, X_2, \cdots, X_n\right) &= -\int f\left(x^n\right) \ln f\left(x^n\right) dx^n \\
&= -\int f\left(x^n\right)\left[-\frac{1}{2}(x^n-\boldsymbol{\mu})^{\mathrm{T}} K^{-1}\left(x^n-\boldsymbol{\mu}\right)) - \ln\left(\sqrt{2\pi}\right)^n |K|^{\frac{1}{2}}\right] dx \\
&= E\left[\frac{1}{2}(x^n-\boldsymbol{\mu})^{\mathrm{T}} K^{-1}\left(x^n-\boldsymbol{\mu}\right))\right] + \frac{1}{2}\ln(2\pi)^n|K| \\
&= \frac{n}{2} + \frac{1}{2}\ln(2\pi)^n|K| \\
&= \frac{1}{2}\ln(2\pi e)^n|K| \quad \text{奈特} \\
&= \frac{1}{2}\log(2\pi e)^n|K| \quad \text{比特}
\end{aligned}$$

多元正态分布的微分熵是最常用、最常考的. 任意非负定矩阵都是某一多元正态分布的协方差矩阵，基于这一事实，有一个**常用技巧**是，遇到处理矩阵不等式，将其转化为多元正态的微分熵问题，利用微分熵的性质处理问题.

8.1.4 条件微分熵

$$h(X \mid Y) = -\int f(x, y) \log f(x \mid y) dx dy$$

8.1.5 相对熵

$$D(f\|g)=\int f\log\frac{f}{g}$$

8.1.6 互信息

$$I(X;Y)=\int f(x,y)\log\frac{f(x,y)}{f(x)f(y)}dxdy$$

满足

$$I(X;Y)=h(X)-h(X\mid Y)=h(Y)-h(Y\mid X)=h(X)+h(Y)-h(X,Y)=D(f(x,y)\|f(x)f(y))$$

8.1.7 微分熵、相对熵、互信息的性质 (关注和离散熵的不同)

1. 相对熵非负 (和离散一样)：$D(f\|g)\geqslant 0$，当且仅当 $f=g$ 时等号成立.(和离散情形类似，用 Jensen 不等式证明，Jensen 不等式在连续情形也适用)
2. 互信息非负 (和离散一样)：$I(X;Y)\geqslant 0$，当且仅当 X、Y 独立时等号成立.(和离散情形类似，由相对熵非负可得)
3. 条件使得相对熵减少 (因为互信息非负，和离散一样)：$h(X\mid Y)\leqslant h(X)$，当且仅当 X、Y 独立时等号成立.
4. 微分熵的链式法则 (和离散一样)：

$$h\left(X_1,X_2,\cdots,X_n\right)=\sum_{i=1}^{n}h\left(X_i\mid X_1,X_2,\cdots,X_{i-1}\right)$$

5. 微分熵的独立界 (和离散一样)：

$$h\left(X_1,X_2,\cdots,X_n\right)\leqslant\sum_{i=1}^{n}h(X_i)$$

6. 互信息的链式法则 (因为是基于熵/微分熵的链式法则，和离散一样)：

$$I\left(X_1,X_2,\cdots,X_n;Y\right)=\sum_{i=1}^{n}I\left(X_i;Y\mid X_{i-1},\cdots,X_1\right)$$

7. 平移变换不改变微分熵 (和离散一样)：

$$h(X+c)=h(X)$$

所以在研究正态分布的微分熵的时候，我们就直接研究零均值情形.

8. **尺度变换对于微分熵的影响 (和离散不同)：**

$$h(aX)=h(X)+\log|a|$$

和

$$h(A\boldsymbol{X})=h(\boldsymbol{X})+\log|\det A|$$

这里 A 表示线性变换的矩阵，$\boldsymbol{X}$ 是随机向量. **这一个性质在考试题中也常会出现!**

9. 关于条件微分熵的一些问题 (和离散不同)：
首先举一个例子

$$h(2X\mid X)\neq 0$$

按照我们对于条件分布的理解，给定了 X 以后，$2X$ 显然是确定的. 如果是离散熵的话，确定量的离散熵应该为 0. 但是对于微分熵来说，由于微分熵的取值和信息度量并不直接关联 (不精确地讲，典型集元素个数可以说明不确定性，即用来度量信息，按照这个想法，典型集元素个数趋于 0 对应微分熵 $h(2X\mid X)\to-\infty$)，所以无法断定这个条件微分熵为 0. 这是**需要特别注意**的!

另一类问题是，若 X、Y **独立**，则有

$$h(X+Y\mid X)=h(Y\mid X)$$

这一条性质在加性噪声信道 (输出是输入和随机噪声的线性叠加，例如高斯信道) 的信道容量计算中常常用到.

10. 协方差矩阵给定的时候，正态分布取得最大熵

记 g 为任意分布，其均值为 0，协方差矩阵给定为 $K=E[XX^t]$，ϕ_k 表示均值为

0、协方差矩阵为 $K = E[XX^t]$ 的正态分布. 利用相对熵的非负性，有

$$\begin{aligned}
0 &\leqslant D(g\|\phi_k) \\
&= \int g \ln \frac{g}{\phi_k} dx \\
&= -h(g) - \int g \left(c - \frac{x^2}{2\sigma^2} \right) dx \\
&= -h(g) - \left(\int (g \cdot c) dx - \frac{E[X^2]}{2\sigma^2} \right) \\
&= -h(g) - \left(\int (\phi_k \cdot c) dx - \frac{E[X^2]}{2\sigma^2} \right) \\
&= -h(g) + h(\phi_k)
\end{aligned}$$

11. 均值受限（非负），指数分布取得最大微分熵

记 g 为任意分布，其均值为 λ，e_λ 表示均值为 λ 的指数分布. 利用相对熵的非负性，有

$$\begin{aligned}
0 &\leqslant D(g\|e_\lambda) \\
&= \int g \ln \frac{g}{e_\lambda} dx \\
&= -h(g) - \int g \left(-\ln \lambda - \frac{x}{\lambda} \right) dx \\
&= -h(g) - \left(\int (g \cdot c) dx - \frac{E[X]}{\lambda} \right) \\
&= -h(g) - \left(\int (e_\lambda \cdot c) dx - \frac{E[X]}{\lambda} \right) \\
&= -h(g) + h(e_\lambda)
\end{aligned}$$

12. 取值范围受限，均匀分布取得最大微分熵

记 g 为任意分布，其取值范围为 $[a, b]$，u_{ab} 表示取值范围为 $[a, b]$ 的均匀分布. 利用相对熵的非负性，有

$$\begin{aligned}
0 &\leqslant D(g\|u_{ab}) \\
&= \int g \ln \frac{g}{u_{ab}} dx \\
&= -h(g) - \int g \left(\frac{1}{b-a} \right) dx \\
&= -h(g) - \int \frac{1}{b-a} \left(\frac{1}{b-a} \right) dx \\
&= -h(g) + h(u_{ab})
\end{aligned}$$

第 9 章　高斯信道

9.1　主要内容

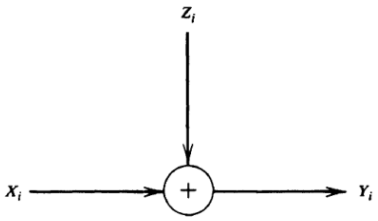
这一章是这门课程中的重要内容. 这一章主要介绍连续情形的信道编码理论，从考察角度讲，主要考察高斯信道的信道容量计算、基于对高斯信道信道容量计算方法的理解求解其他限制下的信道容量、带宽有限的高斯信道的信道容量计算、并联高斯信道的注水法、高斯彩色噪声信道的信道容量计算、反馈对于高斯信道信道容量的影响.

这一章主要需要掌握

9.1.1　高斯信道

$$Y_i = X_i + Z_i, \quad Z_i \sim \mathcal{N}(0, N)$$

模型为



其中对于输入信号的限制为

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2 \leqslant P$$

即输入信号平均功率受限.

9.1.2　高斯信道的信道容量

1. 定义：功率限制为 P 的高斯信道的信道容量定义为

$$C = \max_{f(x):E[X^2]\leqslant P} I(X;Y)$$

2. 计算：

$$\begin{aligned} I(X;Y) &= h(Y) - h(Y \mid X) \\ &= h(Y) - h(X+Z \mid X) \\ &= h(Y) - h(Z \mid X) \\ &= h(Y) - h(Z) \\ &= h(Y) - \frac{1}{2}\log 2\pi eN \end{aligned}$$

接下来要讨论 $h(Y)$ 的最大值. 我们前面有说明，对于微分熵来说，只有确定了限制条件才能求解最大值. 由于高斯信道对于输入平均功率有限制，那么我们猜测输出的平均功率也有限制. 于是我们验证

$$\begin{aligned} E[Y^2] &= E[(X+Z)^2] \\ &= E[X^2] + 2E[X]E[Z] + E[Z^2] \\ &\leqslant P + 2\cdot 0 + N \\ &= P + N \end{aligned}$$

进一步可得

$$Var[Y] = E[Y^2] - (E[Y])^2 \leqslant E[Y^2] \leqslant P + N$$

所以，当输入服从均值为 0，协方差为 P 正态分布时，达到信道容量

$$C = \max_{f(x):E[X^2]\leqslant P} I(X;Y) = \frac{1}{2}\log\left(1+\frac{P}{N}\right) \quad \text{比特/信道使用}$$

高斯信道的信道容量求解过程是需要**熟练掌握**的. 从直观上看，求解过程和离散无记忆信道的信道容量求解是类似的. 这里的一些小的**不同在于**：互信息分解方式固定；加性噪声信道的条件微分熵处理——变成噪声的微分熵；微分熵的最大值求解，需要首先根据输入的限制确定输出的限制，再根据给定限制条件确定何种分布可以使得微分熵达到最大.

9.1.3 带宽有限信道

1. 信道描述：带宽 W 为频带上截止频率和下截止频率之差，信噪比 P/N_0W 如果单位是分贝需要转换，这种定义的计算方式是

$$10\lg\frac{P}{N_0W}(dB)$$

所以由 10 倍对数值即可确定 P/N_0W 实际大小 (**注意分贝单位的计算中对数是以 10 为底**)

2. 关于信噪比的定义: 在前年考试题目中出现了一个和课件上不同的定义，在课程群里很多同学和老师反映这个问题，老师的意思是就按照给出的字符表示来处理，比如那个题目说信噪比 P/N_0 等于多少，就不要管“信噪比”三个字，直接把公式中的相应字符替换掉即可.

3. 计算公式：

$$C = W\log\left(1+\frac{P}{N_0W}\right) \text{ 比特/ 秒}$$

当 $W\to\infty$ 时，由

$$x\to 0\text{时}:\ln(1+x)\to x$$

以及换底公式

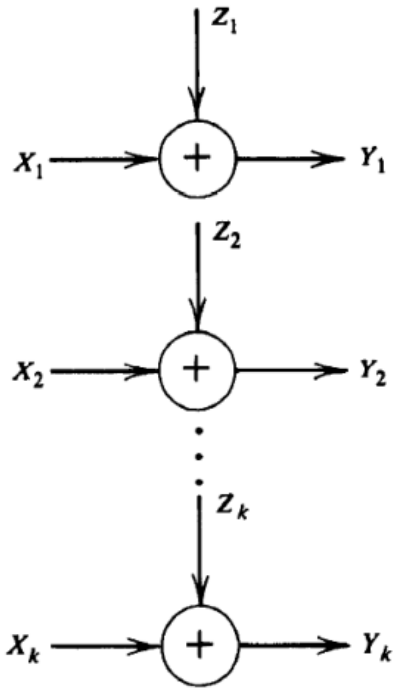
$$\log_a x=\frac{\log_b x}{\log_b a}$$

可得

$$C\to W\cdot\frac{P}{N_0W}\cdot\log_2 e=\frac{P}{N_0}\log_2 e \text{ 比特/ 秒}$$

9.1.4 并联高斯信道

1. 信道模型：



2. 信道描述：

$$Y_i=X_i+Z_i, i=1,2\cdots,k$$

其中

$$Z_i \sim \mathcal{N}(0, N_i)$$

对于输入的限制为总的平均功率限制

$$E\sum_{i=1}^{k} X_i^2 \leqslant P$$

3. 总的信道容量

$$C = \max I\left(X_1, X_2, \cdots, X_k; Y_1, Y_2, \cdots, Y_k\right)$$

和高斯信道的信道容量计算类似

$$\begin{aligned} I(X^n; Y^n) &= h(Y^n) - h(Y^n \mid X^n) \\ &= h(Y^n) - h(X^n + Z^n \mid X^n) \\ &= h(Y^n) - h(Z^n) \\ &= h(Y^n) - \sum_{i=1}^{n} h(Z^i) \\ &\leqslant \sum_{i=1}^{n} h(Y^i) - \sum_{i=1}^{n} h(Z^i) \\ &\leqslant \sum_{i=1}^{n} \frac{1}{2} \log\left(1 + \frac{P_i}{N}\right) \end{aligned}$$

则可知，当输入服从如下分布时，达到信道容量.

$$(X_1, X_2, \cdots, X_k) \sim \mathcal{N}\left(0, \begin{bmatrix} P_1 & 0 & \cdots & 0 \\ 0 & P_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & P_k \end{bmatrix}\right)$$

现在要解决的问题是功率如何分配. 这是一个优化问题，使用拉格朗日乘子法 (这一块的详细推导过程我觉得不需要掌握——指的是从现在开始到注水法的结论之间，即只需要掌握注水法结论即可).

构造函数

$$J\left(P_1, \cdots, P_k\right) = \sum \frac{1}{2} \log\left(1 + \frac{P_i}{N_i}\right) + \lambda\left(\sum P_i\right)$$

对 P_i 求导可得

$$\frac{1}{2} \frac{1}{P_i + N_i} + \lambda = 0$$

解得

$$P_i = v - N_i$$

其中 v 为常数. 但是考虑到 P_i 的非负性，所以修正得到

$$P_i = (v - N_i)^+$$

其中 v 满足

$$\sum (v - N_i)^+ = P$$

这里的 $(x)^+$ 表示

$$(x)^+ = \begin{cases} x, & \text{若} x \geqslant 0 \\ 0, & \text{若} x < 0 \end{cases}$$

基于此分析，我们得到了一种**求解并联高斯信道的信道容量的方法——注水法**. 首先求解 v 使得

$$\sum (v - N_i) = P$$

然后基于此计算 P_i

$$P_i = (v - N_i)$$

如果计算得到的 P_i 都是非负的，则说明得到了正确结果. 否则，需要另噪声方差最大的信道对应的功率分配为 0，然后重复上述操作，直到得到的 P_i 都是非负的.

9.1.5 高斯彩色噪声信道

1. 信道描述：上面提到的并联高斯信道指的是其中的加性噪声是高斯白噪声，即并联信道中每个信道的噪声是独立的

$$Z_i \sim \mathcal{N}(0, N_i)$$

而这里要讨论的是高斯彩色噪声信道，即**每个信道的噪声不独立，是互相关的**. 数学表述就是，噪声协方差矩阵 K_Z 不是对角矩阵.

现在我们给出高斯彩色噪声信道的形式化描述：

$$Y_i = X_i + Z_i, i = 1, 2 \cdots, k$$

其中 Z_i 之间不独立，$Z = (Z_1, Z_2, \ldots, Z_k) \sim \mathcal{N}(0, K_Z)$，$K_Z$ 非对角矩阵. 高斯彩色噪声信道对于输入的功率限制也是总的平均功率限制 (和并联高斯信道一样)

$$\frac{1}{n} \sum_i E[X_i^2] \leqslant P$$

在这里为了后面的讨论方便，给出这个限制的一种等价表达

$$\frac{1}{n}\operatorname{tr}(K_X) \leqslant P$$

2. 总的信道容量

$$C = \max I\left(X_1, X_2, \cdots, X_k; Y_1, Y_2, \cdots, Y_k\right)$$

和高斯信道的信道容量求解过程类似. 首先进行互信息分解.

$$\begin{aligned} I(X^n; Y^n) &= h(Y^n) - h(Y^n \mid X^n) \\ &= h(Y^n) - h(X^n + Z^n \mid X^n) \\ &= h(Y^n) - h(Z^n) \\ &= h(Y^n) - \frac{1}{2}\log(2\pi e)^n |K_Z| \\ &\leqslant \frac{1}{2}\log(2\pi e)^n |K_X + K_Z| - \frac{1}{2}\log(2\pi e)^n |K_Z| \\ &= \frac{1}{2}\log\frac{|K_X + K_Z|}{|K_Z|} \end{aligned}$$

和并联高斯信道相比，这里的求解遇到的**困难在于**：噪声不独立，需要求解协方差矩阵行列式的极值 (**并联高斯信道的互信息也可以这样表达**，只不过那里由于噪声独立，可以把联合熵分解为微分熵的和来处理. 另外注意，分母是一个常量，由噪声决定，不需要我们考虑；我们**只需要求解分子中协方差矩阵行列式的最大值**). 那么我们需要了解如何求解协方差矩阵行列式的最大值. 这里我们需要用到一个**重要的不等式，阿达玛不等式**.

$$|K| \leqslant \sum_{i=1}^{n} K_{ii}$$

关于这个不等式的证明，我们在这里详细说明. 因为这种矩阵行列式问题的**证明技巧**在这门课程中是通用的，而且是重要的.

技巧：任何非负定的矩阵都可以是某个多元正态分布的协方差矩阵，因此我们利用多元正态分布的微分熵进行处理.

记 $X = (X_1, X_2, \ldots, X_n) \sim \mathcal{N}(0, K)$，则有

$$h(X) = \frac{1}{2}\log(2\pi e)^n |K|$$

考虑到 $X_i \sim \mathcal{N}(0, K_{ii})$，有

$$h(X_i) = \frac{1}{2}\log 2\pi e K_{ii}$$

由于微分熵的独立界，可得

$$h(X)=\frac{1}{2}\log(2\pi e)^n|K|\leqslant\sum_{i=1}^{n}h(X_i)=\sum_{i=1}^{n}\frac{1}{2}\log 2\pi eK_{ii}=\frac{1}{2}\log(2\pi e)^n\prod_{i=1}^{n}K_{ii}$$

所以可得

$$|K|\leqslant\sum_{i=1}^{n}K_{ii}$$

有了阿达玛不等式，我们就可以讨论前面我们计算得到的互信息表达式中分子的协方差矩阵行列式的最大值.

阿达玛不等式告诉我们，对于对角元存在限制的情况下，当矩阵为对角矩阵时行列式达到最大值. 在并联高斯信道的信道容量讨论中，K_Z 是对角矩阵，所以只需要 K_X 是对角矩阵，然后结合限制即可求解最大值. 但是在这里，由于 K_Z 不是对角矩阵，所以直接给出让 K_X+K_Z 是对角矩阵并且行列式取得最大值是困难的. 所以我们首先对 K_Z 进行对角化.

$$K_Z=Q\Lambda Q^T$$

然后我们讨论

$$\begin{aligned}|K_X+K_Z|&=|K_X+Q\Lambda Q^T|\\&=|Q|\cdot|Q^TK_XQ+\Lambda|\cdot|Q^T|\qquad(\text{记}A=Q^TK_XQ)\\&=|A+\Lambda|\\&\leqslant\prod_{i=1}^{n}(A_{ii}+\lambda_i)\end{aligned}$$

类似并联高斯信道注水法部分的讨论，可得高斯彩色噪声信道达到信道容量的条件是

$$A_{ii}=(v-\lambda_i)^+$$

所以，**高斯彩色噪声信道的信道容量求解过程可以描述为：**首先对噪声协方差矩阵进行对角化

$$K_Z=Q\Lambda Q^T$$

然后基于对角矩阵 Λ 的对角元，按照如下公式进行注水法求解.

$$A_{ii}=(v-\lambda_i)^+$$

然后由 A_{ii} 得到对角矩阵 A. 根据

$$A = Q^T K_X Q$$

即可得到 K_X.

作业题目和考试题目中往往是考察并联高斯信道的信道容量求解，即注水法. 对于高斯彩色噪声信道，并没有见到考察题目，但根据我们的分析，其实也并不复杂，并且其中核心操作也包括注水法. 这里需要用到实对称矩阵对角化的技术，下一小节我们简要回顾一下.

9.1.6 实对称矩阵对角化——数学基础复习

这里以一道例题说明**实对称矩阵的对角化方法.**

例：设

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{pmatrix}$$

求正交矩阵 T，使得 $T^{-1}AT$ 为对角矩阵.

解：

首先计算矩阵 A 的特征多项式

$$p_A(\lambda) = \det(\lambda I - A) = \begin{vmatrix} \lambda - 1 & -2 & -2 \\ -2 & \lambda - 1 & -2 \\ -2 & -2 & \lambda - 1 \end{vmatrix} = (\lambda - 5)(\lambda + 1)^2$$

所以，A 的特征值为 $\lambda_1 = 5$，$\lambda_2 = -1$(二重). 接下来求解对应的特征向量.
对于 $\lambda_1 = 5$，求解

$$(5I - A)\boldsymbol{x} = 0$$

解得

$$\boldsymbol{x}_1 = (1, 1, 1)^{\mathrm{T}}$$

进行单位化 (我们需要利用施密特标准正交化方法将特征向量构成的矩阵正交化)

$$\boldsymbol{e}_1 = \frac{1}{\sqrt{3}}(1, 1, 1)^{\mathrm{T}}$$

对于 $\lambda_2 = -1$，求解

$$(-I - A)\boldsymbol{x} = 0$$

解得

$$\boldsymbol{x}_2 = (-1,1,0)^{\mathrm{T}}, \boldsymbol{x}_3 = (-1,0,1)^{\mathrm{T}}$$

进行标准正交化得

$$\boldsymbol{e}_2 = \frac{1}{\sqrt{2}}(-1,1,0)^{\mathrm{T}}, \boldsymbol{e}_3 = \frac{1}{\sqrt{6}}(-1,-1,2)^{\mathrm{T}}$$

所以可得正交矩阵 T

$$T = (\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3) = \begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \end{pmatrix}$$

且有对角化后得到对角元为对应特征值的对角矩阵

$$T^{-1}AT = \begin{pmatrix} 5 & & \\ & -1 & \\ & & -1 \end{pmatrix}$$

这里我们总结求解正交矩阵对实对称矩阵进行对角化的步骤：首先求解特征多项式并得到特征值；求解每个特征值对应的特征向量；对特征向量进行施密特标准正交化；把标准正交化后的特征向量排列构成矩阵，即为我们需要的正交矩阵，而利用这个矩阵进行正交化后得到的是原矩阵对角元构成的对称矩阵.(在线性代数理论中证明了实对称矩阵一定可以对角化)

特征值和特征向量的求解在上面例题中有所体现，这里再简单回顾一下施密特正交化的操作过程 (线性代数就是工具，用到了就翻翻看).

设 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_n$ 是 n 维欧氏空间的一组基 (这里的正交化方法对于对称矩阵的特征向量正交化适用). 首先将 α_1 归一化 (对应上面例题中对于 $\lambda = 5$ 对应的特征向量单位化)，令

$$\boldsymbol{e}_1 = \frac{\boldsymbol{\alpha}_1}{|\boldsymbol{\alpha}_1|}$$

然后对 $\boldsymbol{\alpha}_2$ 进行操作. 为了实现正交，首先在 $\boldsymbol{\alpha}_2$ 上减去其在 $\boldsymbol{e}_1$ 上的投影

$$\boldsymbol{\beta}_2 = \boldsymbol{\alpha}_2 - (\boldsymbol{\alpha}_2, \boldsymbol{e}_1)\,\boldsymbol{e}_1$$

然后将其归一化

$$\boldsymbol{e}_2 = \frac{\boldsymbol{\beta}_2}{|\boldsymbol{\beta}_2|}$$

类似地对后面每一个向量进行正交化、归一化，即可得到标准正交化的结果.

$$\boldsymbol{\beta}_k = \boldsymbol{\alpha}_k - \sum_{i=1}^{k-1} (\boldsymbol{\alpha}_k, \boldsymbol{e}_i)\, \boldsymbol{e}_i$$

$$\boldsymbol{e}_k = \frac{\boldsymbol{\beta}_k}{|\boldsymbol{\beta}_k|}$$

9.1.7 带反馈的高斯信道的信道容量

1. 概述：

   对于无记忆 (噪声互相独立) 的高斯信道，反馈不增加信道容量；

   对于有记忆 (噪声相关) 的高斯信道，反馈会对信道容量产生影响.

2. 有记忆高斯信道的信道容量：

   无反馈情况下

   $$C_n = \max_{\frac{1}{n}\operatorname{tr}(K_X)\leqslant P} \frac{1}{2n} \log \frac{|K_X + K_Z|}{|K_Z|}$$

   有反馈情况下

   $$C_{n,FB} = \max_{\frac{1}{n}\operatorname{tr}(K_X)\leqslant P} \frac{1}{2n} \log \frac{|K_{X+Z}|}{|K_Z|}$$

   无论是否有反馈，信道容量都是可达码率的上界.

3. 反馈对于有记忆高斯信道的信道容量的影响：

   引入反馈后信道容量提升，并且受两个约束限制. 分别是

   $$C_{n,FB} \leqslant C_n + \frac{1}{2} \quad \text{比特/传输}$$

   和

   $$C_{n,FB} \leqslant 2C_n$$

4. 证明上述结论所需要的准备工作：

   这里做出详细说明的原因是，这些关于矩阵的等式和不等式的证明都可以用信息论的方法解决，老师在课堂上也经常强调这种思想和相关技巧. 所以在考察的时候有可能会用到.

   关于为什么会想到用这些结论，实际上是从目标出发的. 我们需要比较引入反馈前后信道容量，按照我们前面的分析，信道容量的比较可以转化为 $|K_{X+Z}|$ 和 $|K_X + K_Z|$ 的比较. 第一个结论就是要建立两个协方差矩阵的数量关系；第二个结论为了第三个结论服务，直接进行行列式的比较，得到其中一个约束条件；第四个结论则是要结合第一个结论，得到第二个约束条件.

- $K_{X+Z}+K_{X-Z}=2K_X+2K_Z$

$$\begin{aligned}K_{X+Z}&=E[(X+Z)(X+Z)^T]\\&=E[XX^T+ZX^T+XZ^T+ZZ^T]\\&=E[XX^T]+E[ZX^T]+E[XZ^T]+E[ZZ^T]\end{aligned}$$

$$\begin{aligned}K_{X-Z}&=E[(X-Z)(X-Z)^T]\\&=E[XX^T-ZX^T-XZ^T+ZZ^T]\\&=E[XX^T]-E[ZX^T]-E[XZ^T]+E[ZZ^T]\end{aligned}$$

所以有

$$K_{X+Z}+K_{X-Z}=2\left(E[XX^T]+E[ZZ^T]\right)=2K_X+2K_Z$$

- 对于非负定矩阵 A、B，若 $A-B$ 是非负定的，则有 $|A|\geqslant|B|$
这种行列式不等式问题的解决技巧是，利用正态分布的微分熵.
记 $C=A-B$,$X_1\sim\mathcal{N}(0,B)$，$X_2\sim\mathcal{N}(0,C)$，则有

$$Y=X_1+X_2\sim\mathcal{N}(0,A)$$

要证明 $|A|\geqslant|B|$，可以转化为证明 $h(Y)\geqslant h(X_1)$，同时**注意一般证明熵不等式的时候需要用到熵的性质、链式法则等，比如条件使熵减少 (自己构造条件)、熵的独立界等.** 这里我们利用条件使熵减少

$$\begin{aligned}h(Y)&=\frac{1}{2}\log(2\pi e)^n|A|\\&\geqslant h(Y\mid X_2)\\&=h(X_1\mid X_2)\\&=h(X_1)\\&=\frac{1}{2}\log(2\pi e)^n|B|\end{aligned}$$

- $|K_{X+Z}|\leqslant 2^n\,|K_X+K_Z|$
前面我们证明了一个结论

$$K_{X+Z}+K_{X-Z}=2K_X+2K_Z$$

记 $A=2(K_X+K_Z)$，$B=K_{X+Z}$，$C=K_{X-Z}$. 于是，利用上一条结论有

$$|K_{X+Z}|\leqslant 2^n\,|K_X+K_Z|$$

- 对于非负定矩阵 A、B，和 $0 \leqslant \lambda \leqslant 1$，有 $|\lambda A + (1-\lambda)B| \geqslant |A|^{\lambda}|B|^{1-\lambda}$

  这个结论的证明首先需要**两个重要的想法**. 一个是看到**非负定矩阵**，想到可以是某个多元正态分布的协方差矩阵，进而利用微分熵来解决；一个是看到**线性组合 (这里是看到行列式中有协方差矩阵的线性组合——对应随机变量的线性组合)，可以看成另一个随机变量依概率取这两个随机变量中的一个**，这种情况我们往往要引入示性变量，**并且引入的示性变量一般用于构成条件熵**.

  记 $X \sim \mathcal{N}(0, A)$，$Y \sim \mathcal{N}(0, B)$

  $$\theta = \begin{cases} 1, & \text{概率} : \lambda \\ 2, & \text{概率} : 1-\lambda \end{cases}$$

  以及

  $$Z = \begin{cases} X, & \text{若} \theta = 1 \\ Y, & \text{若} \theta = 2 \end{cases}$$

  于是可得

  $$K_Z = \lambda A + (1-\lambda)B$$

  于是，有

  $$\begin{aligned} \frac{1}{2}\log(2\pi e)^n|\lambda A + (1-\lambda)B| &\geqslant h(Z) \\ &\geqslant h(Z \mid \theta) \\ &= \sum p(\theta = \alpha_i)h(Z \mid \theta = \alpha_i) \\ &= p(\theta = 1)h(Z \mid \theta = 1) + p(\theta = 2)h(Z \mid \theta = 2) \\ &= \lambda h(Z \mid \theta = 1) + (1-\lambda)h(Z \mid \theta = 2) \\ &= \lambda h(X \mid \theta = 1) + (1-\lambda)h(Y \mid \theta = 2) \\ &= \lambda h(X) + (1-\lambda)h(Y) \\ &= \lambda \cdot \frac{1}{2}\log(2\pi e)^n|A| + (1-\lambda) \cdot \frac{1}{2}\log(2\pi e)^n|B| \\ &= \frac{1}{2}\log(2\pi e)^n|A|^{\lambda}|B|^{1-\lambda} \end{aligned}$$

- $|K_{X-Z}| \geqslant |K_Z|$

  矩阵行列式不等式，仍然采用信息论方法证明. 按照协方差矩阵构造随机变

量，有

$$\begin{aligned}
h(X^n - Z^n) &= \sum_{i=1}^{n} h(X_i - Z_i \mid X^{i-1} - Z^{i-1}) \\
&\geqslant \sum_{i=1}^{n} h(X_i - Z_i \mid X^{i-1}, Z^{i-1}, X_i) \\
&= \sum_{i=1}^{n} h(-Z_i \mid X^{i-1}, Z^{i-1}, X_i) \\
&= \sum_{i=1}^{n} h(Z_i \mid X^{i-1}, Z^{i-1}, X_i) \quad (\text{由于} h(-Z) = h(Z) + \log|1| = h(Z)) \\
&= \sum_{i=1}^{n} h(Z_i \mid Z^{i-1}) \\
&= h(Z^n)
\end{aligned}$$

5. 结论证明：

- $C_{n,FB} \leqslant C_n + \frac{1}{2}$ 比特/传输

$$\begin{aligned}
C_{n,FB} &= \max_{\frac{1}{n}\operatorname{tr}(K_X)\leqslant P} \frac{1}{2n} \log \frac{|K_{X+Z}|}{|K_Z|} \\
&\leqslant \max_{\frac{1}{n}\operatorname{tr}(K_X)\leqslant P} \frac{1}{2n} \log \frac{2^n |K_X + K_Z|}{|K_Z|} \\
&= C_n + \frac{1}{2} \quad \text{比特/传输}
\end{aligned}$$

- $C_{n,FB} \leqslant 2C_n$

$$\begin{aligned}
C_n &= \max_{\frac{1}{n}\operatorname{tr}(K_X)\leqslant P} \frac{1}{2n} \log \frac{|K_X + K_Z|}{|K_Z|} \\
&= \max_{\frac{1}{n}\operatorname{tr}(K_X)\leqslant P} \frac{1}{2n} \log \frac{\left|\frac{1}{2}K_{X+Z} + \frac{1}{2}K_{X-Z}\right|}{|K_Z|} \\
&\geqslant \max_{\frac{1}{n}\operatorname{tr}(K_X)\leqslant P} \frac{1}{2n} \log \frac{|K_{X+Z}|^{\frac{1}{2}} |K_{X-Z}|^{\frac{1}{2}}}{|K_Z|} \\
&\geqslant \max_{\frac{1}{n}\operatorname{tr}(K_X)\leqslant P} \frac{1}{2n} \log \frac{|K_{X+Z}|^{\frac{1}{2}}}{|K_Z|^{\frac{1}{2}}} \\
&= \frac{1}{2} C_{n,FB}
\end{aligned}$$

第 10 章　率失真理论

10.1　主要内容

这一章是这门课程中的难点所在，按照老师的说法每年都会出一道大题考察率失真函数的计算，也是失分比较严重的一道题. 这里根据题目练习的经验，我们总结一下求解率失真函数的完整过程，并指出易错点所在.

这一章主要需要掌握的内容包括

10.1.1　率失真理论

首先说明一下这一章在干啥. 这一章本质上分析的是允许失真情况下的最优信源编码.

对于离散情形，我们可以做到无失真的编码；对于连续情形，我们知道无论我们用多少比特都无法精确描述一个连续量，也就是说我们无法做到无失真编码.

从另一个角度看，其实无失真编码也不一定是最好的. 直观上看，如果允许失真，那么我们可以进行更大程度的数据压缩，即信源编码可以更短，而我们信源编码就是希望更大程度的压缩从而提高传输效率. 因此，允许失真对于信源编码是有利的. 但是失真如果过大的话会影响接收方对于信息的恢复，甚至可能使得信息传输无效 (接收方无法得到正确的信息). 所以说，合适的才是最好的. **我们需要根据失真和压缩率的要求，给出最合适的信源编码方案**. 这一章要研究的就是这个问题.

简言之，这一章研究的率失真函数就是：讨论在不同的允许失真情况下码率的下界，即传输信息的最小值，也就代表了信源编码的最大数据压缩率.

10.1.2　失真度量

既然我们要量化地研究不同允许失真情况下的码率下界，那么我们首先要对失真进行量化. 常用的失真量化方式有两种

1. 离散情形：汉明失真 (误差概率失真)

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

并且有

$$\mathrm{E}\, d(X,\hat{X}) = p(X \neq \hat{X})$$

这类失真一般用于离散分布，如伯努利分布.

2. 连续情形：平方误差失真

$$d(x,\hat{x}) = (x-\hat{x})^2$$

这类失真一般用于连续分布，如正态分布.

要注意的是，我们研究问题首先要考虑问题有意义. 所以我们对于失真度量提出的一个要求是——有界. 即

$$d_{\max} \triangleq \max_{x\in\mathcal{X},\hat{x}\in\hat{\mathcal{X}}} d(x,\hat{x}) < \infty$$

具体采用哪种量化方式，一般题目会给出. 注意到实际上采用哪种失真函数作为失真度量没有固定要求，是人为规定的，是人们结合经验、根据实际问题的特点和我们的需求来确定的.

10.1.3 率失真码及其失真定义

$\left(2^{nR}, n\right)$ 率失真码包括

1. 编码函数

$$f_n : \mathcal{X} \to \left\{1, 2, \cdots, 2^{nR}\right\}$$

2. 译码函数

$$g_n : \left\{1, 2, \cdots, 2^{nR}\right\} \to \hat{\mathcal{X}}^n$$

其失真定义为

$$D = \mathrm{E}\, d\left(X^n, g_n\left(f_n\left(X^n\right)\right)\right) = \sum_{x^n} p\left(x^n\right) d\left(x^n, g_n\left(f_n\left(x^n\right)\right)\right)$$

10.1.4 率失真函数

称率失真对 (R,D) 是可达的，若存在一个 $\left(2^{nR}, n\right)$ 率失真码，使得

$$\lim_{n\to\infty} Ed\left(X^n, g_n\left(f_n\left(X^n\right)\right)\right) \leqslant D$$

全体可达率失真对 (R,D) 构成了率失真区域. 对于给定的失真 D，率失真区域内的码率 R 的下确界称为率失真函数.

这个是率失真函数的定义，**不过在求解的时候，我们用的不是这个定义，而是信息率失真函数**. 有定理保证，对于独立同分布信源，失真函数有界情况下，率失真函数和

对应的信息率失真函数相等. 下面我们介绍信息率失真函数.

10.1.5 信息率失真函数

信息率失真函数的定义是

$$R^{(I)}(D)=\min_{p(\hat{x}|x):\sum_{x,\hat{x}}p(x)p(\hat{x}|x)d(x,\hat{x})\leqslant D} I(X;\hat{X})$$

这个定义的含义是：在不同信源编码方案下，满足失真限制时为了恢复信源所需要的最小互信息. 如果对于这个信息率失真函数中为什么是求在不同的条件概率 (转移概率) 下的最小互信息，请看下一小节.

10.1.6 信息率失真函数和信道容量的比较

信道容量的定义是

$$C=\max_{p(x)} I(X;Y)$$

它表示信道的最大传输能力，**反映的是信道本身的特性，应该与信源无关**. 但平均互信息量与信源的特性有关，**为排除信源特性对信道容量的影响，在所有的信源中以那个能够使平均互信息量达到最大的信源为参考**，从而使信道容量仅与信道特性有关.

信息率失真函数的定义是

$$R^{(I)}(D)=\min_{p(\hat{x}|x):\sum_{x,\hat{x}}p(x)p(\hat{x}|x)d(x,\hat{x})\leqslant D} I(X;\hat{X})$$

信息率失真函数本质上是描述信源的，它与信道无关. **为此在所有信道中以能够使平均互信息量达到最小的信道为参考**，从而使信息率失真函数仅与信源特性有关.

所以，在求信道容量时我们是求不同 $p(x)$ 下的互信息最大值，而信息率失真函数是不同 $p(\hat{x}\mid x)$ 下的互信息最小值.

在求解方法层面，求解信道容量的最一般、最普适的方法是：假设信道输入的概率分布，进而得到互信息的形式化表达，将其看作输入概率的函数，求解多元函数的极大值；求解信息率失真函数的最一般、最普适的方法是：假设转移概率分布，进而得到互信息的形式化表达，将其看作转移概率的函数，求解多元函数的条件极值，其中约束条件即为失真约束（需要用到拉格朗日乘子法）.

10.1.7 信息率失真函数求解过程

这一小节先形式化地描述信息率失真函数求解的过程，也是经过练习一些题目后的总结. 首先，在上一小节我们解释了求解信息率失真函数的最一般的方法，这一小节介绍一种形式化的处理流程，在三类常见问题上可以使得求解得到简化.

1. 求解 $I(X;\hat{X})_{min}$

   - 互信息分解：$I(X;\hat{X}) = H(X) - H(X \mid \hat{X})$

     在离散无记忆信道的信道容量求解部分，我们讨论了两种不同的分解方式. 在这里，分解方式固定，因为我们的信源是已知的，即 $H(X)$ 已知.

   - **引入失真** (这里我们介绍最常用的情形)：

     离散情形：

     对于二元 (例如：二元伯努利信源) 分布，引入汉明失真的方法是

     $$H(X \mid \hat{X}) = H(X \oplus \hat{X} \mid \hat{X}) \leqslant H(X \oplus \hat{X})$$

     因为有

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

     对于多元 (例如：离散均匀分布信源) 分布，引入汉明失真的方式是利用费诺不等式

     $$H(X \mid \hat{X}) \leqslant H(P_e) + P_e \log|\mathcal{X} - 1|$$

     由于 $P_e = p(x \neq \hat{x}) = Ed(x,\hat{x}) = D$，所以在这里应用的形式是

     $$H(X \mid \hat{X}) \leqslant H(D) + D \log|\mathcal{X} - 1|$$

     连续情形：引入平方误差失真的方法是

     $$h(X \mid \hat{X}) = h(X - \hat{X} \mid \hat{X}) \leqslant h(X - \hat{X})$$

   - 求解 $H(X \mid \hat{X})$ 的最大值

     在引入失真后，实际上我们需要求解的是 $H(X \oplus \hat{X})$ 或 $h(X - \hat{X})$ 的最大值. 即要求解的是最大离散熵或者给定限制下的最大微分熵.

2. 确定等号成立的条件

   实际上要做的就是按照等号成立时的 $H(X \mid \hat{X})$，写出对应的虚拟信道.

3. 求解临界值

   这个是**易错点**，很容易忽略这一问题. 实际上我们知道，我们要求解的率失真函数是给定失真限制下的可达码率的下确界，而码率是非负值，所以率失真函数是非负的. 而按照我们上面求解过程得到的函数，当失真 D 足够大时得到的函数值会变成负数，这是由于前面的分析中没有考虑到——当允许失真大于临界值时，不需要任何信息传输即可满足要求，即率失真函数应该为 0.

临界值的求解其实比较简单，就令前面求解得到的率失真函数为 0，得到的 D 就是临界的 D.

接下来的三个小节，我们用三个例子分别展示率失真函数求解中最常见的三种引入失真的方法.

10.1.8 信息率失真函数求解例子——离散情形 (汉明失真)，利用异或

这一小节，我们分析二元伯努利信源在汉明失真度量下的率失真函数求解，其中引入失真的方法为异或.

首先我们对互信息进行分解

$$I(X;\hat{X}) = H(X) - H(X \mid \hat{X})$$

然后引入失真. 对于二元信源，采用汉明失真，引入失真的方法为

$$H(X \mid \hat{X}) = H(X \oplus \hat{X} \mid \hat{X}) \leqslant H(X \oplus \hat{X})$$

接着求解条件熵的最大值即可得互信息最小值.

$$\begin{aligned} I(X;\hat{X}) &= H(X) - H(X \mid \hat{X}) \\ &= H(p) - H(X \oplus \hat{X} \mid \hat{X}) \\ &\geqslant H(p) - H(X \oplus \hat{X}) \\ &\geqslant H(p) - H(D) \end{aligned}$$

这里 $H(X \oplus \hat{X}) \leqslant H(D)$ 是由于伯努利分布的熵的单调性得到的. 由于

$$0 \leqslant p(X \oplus \hat{X} = 1) = E[d(x,\hat{x})] \leqslant D \leqslant \min\{p, 1-p\} \leqslant \frac{1}{2}$$

以及伯努利分布 (参数为 p) 的熵在 $0 \leqslant p \leqslant \frac{1}{2}$ 时随着 p 单调递增，所以有

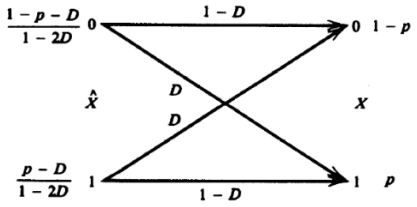
$$H(X \oplus \hat{X}) \leqslant H(D)$$

这里简单补充说明 $D \leqslant \min\{p, 1-p\}$ 的理由，事实上我们考虑当 $R(D) = 0$ 即没有任何信息时，那么 X 的最小误差估计应该为 $\hat{X} = 1$ 若 $p \leqslant 1-p$，$\hat{X} = 0$ 若 $p > 1-p$，按照定义计算此时失真即可得 $D \leqslant \min\{p, 1-p\}$。

这样我们完成了第一步. 接着我们确定等号成立的条件. 由前面分析可知，等号成立即

$$H(X \oplus \hat{X}) = H(D)$$

即 $p(X \neq \hat{X}) = D$. 所以可得虚拟信道



最后一步是确定临界值. 这一步比较简单 (但是容易忽略)，直接令前面求解的率失真函数为 0，得到对应的 D. 令

$$H(p) - H(D) = 0$$

得

$$D = \min\{p, 1-p\}$$

综上所述，我们得到了伯努利信源在汉明失真度量下的率失真函数

$$R(D) = \begin{cases} H(p) - H(D), & 0 \leqslant D \leqslant \min\{p, 1-p\}, \\ 0, & D > \min\{p, 1-p\}. \end{cases}$$

10.1.9 信息率失真函数求解例子——离散情形 (汉明失真)，利用费诺不等式

这一小节，我们讨论离散均匀分布信源在汉明失真下的率失真函数求解，引入失真的方法为费诺不等式.

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.

解：

首先我们对互信息进行分解

$$I(X; \hat{X}) = H(X) - H(X \mid \hat{X})$$

然后引入失真. 对于二元信源，采用汉明失真，引入失真的方法为费诺不等式

$$H(P_e) + P_e \log(|\mathcal{X}| - 1) \geqslant H(X \mid \hat{X})$$

由题意知，$P_e = p(x \neq \hat{x}) = Ed(x, \hat{x}) = D$. 由费诺不等式有

$$H(X \mid \hat{X}) \leqslant H(D) + D\log(m-1)$$

接着求解条件熵的最大值即可得互信息最小值.

$$\begin{aligned} I(X;\hat{X}) &= H(X) - H(X \mid \hat{X}) \\ &\geqslant \log m - H(D) - D\log(m-1) \end{aligned}$$

下面分析等号成立条件. 根据费诺不等式的等号成立条件可知

$$p(\hat{x} \mid x) = \begin{cases} 1-D, & \text{如果} \hat{x} = x \\ \frac{D}{m-1}, & \text{如果} \hat{x} \neq x \end{cases}$$

最后一步是确定临界值. 这一步比较简单 (但是容易忽略)，直接令前面求解的率失真函数为 0，得到对应的 D. 令率失真函数为 0，可得 $D = 1 - 1/m$. 综上所述，率失真函数为

$$R(D) = \begin{cases} \log m - H(D) - D\log(m-1), & 0 \leqslant D \leqslant 1 - \frac{1}{m}, \\ 0, & D > 1 - \frac{1}{m}. \end{cases}$$

这里我们简要总结一下. 对于离散情形，一般会引入汉明失真 (但是注意，失真函数的选择是任意的，取决于实际需要，所以题目可能给出各种形式的失真函数，实际解题需要关注失真函数具体是什么. 不过从老师讲的内容和教材上看，离散情形常用汉明失真). **如果信源是二元的，则可以用异或来引入失真；如果信源是多元的，则可以用费诺不等式引入失真.**

10.1.10 信息率失真函数求解例子——连续情形 (平方误差失真)

这一小节我们选取的例子是，求解高斯信源 $\mathcal{N}(0, \sigma^2)$ 在平方误差失真度量下的率失真函数.

首先我们对互信息进行分解

$$I(X;\hat{X}) = H(X) - H(X \mid \hat{X})$$

然后引入失真. 对于平方误差失真，引入失真的方法为

$$h(X \mid \hat{X}) = h(X - \hat{X} \mid \hat{X}) \leqslant h(X - \hat{X})$$

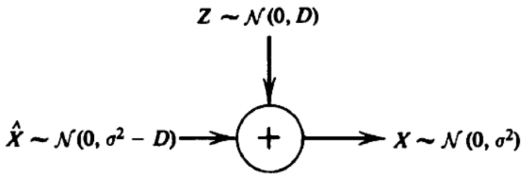
接着求解条件熵的最大值即可得互信息最小值.

$$\begin{aligned} I(X;\hat{X}) &= h(X) - h(X \mid \hat{X}) \\ &= \frac{1}{2}\log 2\pi e\sigma^2 - h(X-\hat{X} \mid \hat{X}) \\ &\geqslant \frac{1}{2}\log 2\pi e\sigma^2 - h(X-\hat{X}) \\ &\geqslant \frac{1}{2}\log 2\pi e\sigma^2 - h(\mathcal{N}(0, E[(X-\hat{X})^2])) \\ &\geqslant \frac{1}{2}\log\frac{\sigma^2}{D} \end{aligned}$$

这样我们完成了第一步. 接着我们确定等号成立的条件. 由前面分析可知，等号成立即

$$h(X-\hat{X}) = h(\mathcal{N}(0, E[(X-\hat{X})^2])) = \frac{1}{2}\log 2\pi eD$$

即虚拟信道为高斯信道，加性噪声 $Z \sim \mathcal{N}(0, D)$.



最后一步是确定临界值. 这一步比较简单 (但是容易忽略)，直接令前面求解的率失真函数为 0，得到对应的 D. 令

$$\frac{1}{2}\log\frac{\sigma^2}{D} = 0$$

得

$$D = \sigma^2$$

综上所述，我们得到了高斯信源在平方误差失真度量下的率失真函数

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

10.1.11 最一般的信息率失真求解方法

如果遇到了用上述方法不方便求解的问题，可以考虑采用最基本的方法求解率失真. 在前文介绍过求解的过程，这里通过一个例子说明.

信源 $X \sim$ Bernoulli $\left(\frac{1}{2}\right)$，失真度量由矩阵给出

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

计算该信源的率失真函数.
解析：由于失真有限，所以可得

$$p(x=0,\hat{x}=1)=p(x=1,\hat{x}=0)=0$$

这类问题处理方法和信道容量求解第三类模型类似，那里是假设输入的概率分布求解互信息极大值，这里是假设虚拟信道概率分布求解有约束条件下互信息极小值. 具体地，假设

$$p(\hat{x}\mid x)=\begin{pmatrix} p_0 & 1-p_0 & 0 \\ 0 & 1-p_1 & p_1 \end{pmatrix}$$

可得 $\hat{X}$ 的概率分布为

$$p(\hat{x})=p(x)\cdot p(\hat{x}\mid x)=\left(\frac{1}{2}p_0,1-\frac{1}{2}p_0-\frac{1}{2}p_1,\frac{1}{2}p_1\right)$$

进而可得

$$\begin{aligned}
&I(X;\hat{X})=H(\hat{X})-H(\hat{X}\mid X)=H\left(\frac{1}{2}p_0,1-\frac{1}{2}p_0-\frac{1}{2}p_1,\frac{1}{2}p_1\right)-\frac{1}{2}H\left(p_0\right)-\frac{1}{2}H\left(p_1\right)\\
&=-\frac{1}{2}p_0\log\frac{1}{2}p_0-\frac{1}{2}p_1\log\frac{1}{2}p_1-\left(1-\frac{1}{2}p_0-\frac{1}{2}p_1\right)\log\left(1-\frac{1}{2}p_0-\frac{1}{2}p_1\right)\\
&\quad+\frac{1}{2}p_0\log p_0+\frac{1}{2}\left(1-p_0\right)\log\left(1-p_0\right)+\frac{1}{2}p_1\log p_1+\frac{1}{2}\left(1-p_1\right)\log\left(1-p_1\right)\\
&=\frac{1}{2}p_0+\frac{1}{2}p_1+\frac{1}{2}\left(1-p_0\right)\log\left(1-p_0\right)+\frac{1}{2}\left(1-p_1\right)\log\left(1-p_1\right)-\frac{1}{2}\left(2-p_0-p_1\right)\log\frac{1}{2}\left(2-p_0-p_1\right)
\end{aligned}$$

同时有约束

$$Ed(x,\hat{x})=\frac{1}{2}\left(1-p_0\right)+\frac{1}{2}\left(1-p_1\right)\leqslant D$$

于是，可以使用拉格朗日乘子法求解这一有约束的条件极值问题. 设

$$\begin{aligned}
f(x,y,\lambda)=&\frac{1}{2}x+\frac{1}{2}y+\frac{1}{2}(1-x)\log(1-x)+\frac{1}{2}(1-y)\log(1-y)\\
&-\frac{1}{2}(2-x-y)\log\frac{1}{2}(2-x-y)+\lambda\cdot\left[1-D-\frac{1}{2}(x+y)\right]
\end{aligned}$$

令

$$\begin{aligned}
\frac{\partial f}{\partial x}=&\frac{1}{2}+\frac{1}{2}(-1)\cdot\log(1-x)+\frac{1}{2}(1-x)\frac{(-1)}{1-x}\cdot\frac{1}{\ln 2}\\
&-\frac{1}{2}(-1)\cdot\log\frac{1}{2}(2-x-y)-\frac{1}{2}(2-x-y)\frac{-\frac{1}{2}}{\frac{1}{2}(2-x-y)}\frac{1}{\ln 2}-\frac{1}{2}\lambda=0
\end{aligned}$$

可得

$$\log\frac{2-x-y}{1-x}=\lambda$$

令

$$\begin{aligned}\frac{\partial f}{\partial y}=&\frac{1}{2}+\frac{1}{2}(-1)\cdot\log(1-y)+\frac{1}{2}(1-y)\frac{(-1)}{1-y}\cdot\frac{1}{\ln 2}\\&-\frac{1}{2}(-1)\cdot\log\frac{1}{2}(2-x-y)-\frac{1}{2}(2-x-y)\frac{-\frac{1}{2}}{\frac{1}{2}(2-x-y)}\frac{1}{\ln 2}-\frac{1}{2}\lambda=0\end{aligned}$$

可得

$$\log\frac{2-x-y}{1-y}=\lambda$$

令

$$\frac{\partial f}{\partial\lambda}=1-D-\frac{1}{2}(x+y)=0$$

可得

$$x+y=2D$$

由 $\dfrac{\partial f}{\partial x}=0$ 和 $\dfrac{\partial f}{\partial y}=0$ 可得 $x=y$. 结合 $\dfrac{\partial f}{\partial\lambda}=0$ 可得 $x=y=D$. 于是，互信息达到最小值时的条件概率分布为

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

最后确定临界值. 令率失真函数为 0，得 $D=1$. 综上所述，可得率失真函数

$$R(D)=\begin{cases}1-D, & 0\leqslant D\leqslant 1\\ 0, & D>1\end{cases}$$

10.1.12 并联高斯信源的率失真函数——反注水法

并联高斯信源的率失真函数求解是这一部分的另一个重点，在考试题目中经常出现，需要熟练掌握.

考虑有 m 个独立的高斯信源 $X_i\sim\mathcal{N}(0,\sigma_i^2)$，取失真度量

$$d\left(X^m,\hat{X}^m\right)=\sum_{i=1}^{m}(x_i-\hat{x}_i)^2$$

并定义率失真函数

$$R(D)=\min_{f(\hat{x}^m|x^m):Ed\left(X^m,\hat{X}^m\right)\leqslant D}I\left(X^m;\hat{X}^m\right)$$

类似前面的求解过程，首先对互信息进行分解. 这里利用条件使熵减少，将其转化为若干高斯高斯信源互信息的叠加，进而利用高斯信源率失真函数的求解得到结果.

$$\begin{aligned} I\left(X^m;\hat{X}^m\right) &= h\left(X^m\right) - h\left(X^m \mid \hat{X}^m\right) \\ &= \sum_{i=1}^{m} h\left(X_i\right) - \sum_{i=1}^{m} h\left(X_i \mid X^{i-1}, \hat{X}^m\right) \\ &\geqslant \sum_{i=1}^{m} h\left(X_i\right) - \sum_{i=1}^{m} h\left(X_i \mid \hat{X}_i\right) \\ &= \sum_{i=1}^{m} I\left(X_i;\hat{X}_i\right) \\ &\geqslant \sum_{i=1}^{m} R\left(D_i\right) \\ &= \sum_{i=1}^{m} \left(\frac{1}{2}\log\frac{\sigma_i^2}{D_i}\right)^+ \end{aligned}$$

和并联高斯信道的信道容量求解类似，这里也需要求解优化问题. 利用拉格朗日乘子法，可得

$$R(D) = \begin{cases} \sum_{i=1}^{m} \frac{1}{2}\log\frac{\sigma_i^2}{D_i}, & 0 \leqslant D \leqslant \sum_{i=1}^{n}\sigma_i^2, \\ 0, & D > \sum_{i=1}^{n}\sigma_i^2. \end{cases}$$

其中

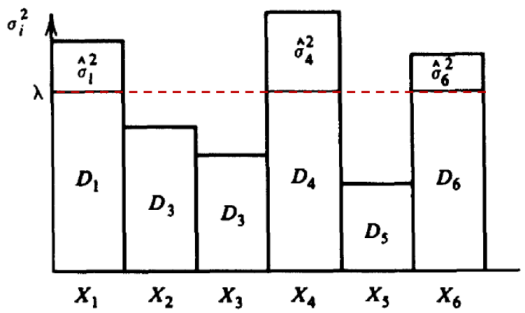
$$D_i = \begin{cases} \lambda, & \text{如果} \lambda < \sigma_i^2 \\ \sigma_i^2, & \text{如果} \lambda \geqslant \sigma_i^2 \end{cases}$$

λ 的选取使得

$$\sum_{i=1}^{n} D_i = D$$

这里我们其实已经完成了率失真函数求解的三步. 其中互信息最小值求解和临界值求解已经显式给出，至于等号成立条件，则是和高斯信源一样，对于独立高斯信源中的每一个 $X_i \sim \mathcal{N}(0,\sigma_i^2)$，其虚拟信道为高斯信道，加性噪声 $Z_i\mathcal{N}(0,D_i)$.

这种方法也叫做反注水法. 如图形象地描述了这个过程.



我们总结一下求解并联高斯信源的率失真函数的反注水法的操作过程.

1. 对所有高斯信源的方差由小到大排序，记为 $\sigma_i'^2(i=1,2,\ldots,n)$

2. D 的第一个区间上界 $D_{m1}=\sigma_1'^2\cdot n$. 当 $0<D\leqslant D_{m1}$ 时，失真分配方案为平均分配，即 $D_i=D/n$

3. D 的第二个区间上界 $D_{m2}=\sigma_i'^2+\sigma_2'^2\cdot(n-1)$. 当 $D_{m1}<D\leqslant D_{m2}$ 时，失真分配方案为：方差最小的信源分配失真为其方差，然后其他信源平均分配失真. 记方差为 $\sigma_1'^2$ 的信源编号为 j，则失真分配为 $D_{j1}=\sigma_1'^2$，对于 $k\neq j$ 有 $D_k=(D-\sigma_1'^2)/(n-1)$

4. 按照上述操作，逐步确定每个区间 $(D_{m(i-1)},D_{mi}]$. 并基于此，分配失真. 分配方案是：对于方差最小的前 $(i-1)$ 个信源分配失真和其方差相同 $D_{ji}=\sigma_i'^2$；然后对于其余方差较大的信源，平均分配剩余的失真

5. 当分配方案使得所有信源都分配了和其自身方差相等的失真，即 $D=\sum_{i=1}^n\sigma_i^2$ 时，率失真函数变成 0. 并且再增加失真，率失真函数不变，始终保持 $0(D=\sum_{i=1}^n\sigma_i^2$ 即是临界值)

10.1.13 对失真函数进行线性变换对于率失真函数的影响

这一小节我们介绍一个重要的结论——失真函数进行线性变换后对应的率失真函数.

1. $\tilde{d}(x,\hat{x})=d(x,\hat{x})+a$，其中 $a>0$

$$\begin{aligned}\tilde{R}(D)&=\min_{p(\hat{x}|x):E(\tilde{d}(x,\hat{x}))\leqslant D}I(X;\hat{X})\\&=\min_{p(\hat{x}|x):E(d(x,\hat{x}))+a\leqslant D}I(X;\hat{X})\\&=\min_{p(\hat{x}|x):E(d(x,\hat{x}))\leqslant D-a}I(X;\hat{X})\\&=R(D-a)\end{aligned}$$

2. $d^*(x,\hat{x})=bd(x,\hat{x})$，其中 $b>0$

$$\begin{aligned}R^*(D)&=\min_{p(\hat{x}|x):E(d^*(x,\hat{x}))\leqslant D}I(X;\hat{X})\\&=\min_{p(\hat{x}|x):E(bd(x,\hat{x}))\leqslant D}I(X;\hat{X})\\&=\min_{p(\hat{x}|x):E(d(x,\hat{x}))\leqslant \frac{D}{b}}I(X;\hat{X})\\&=R\left(\frac{D}{b}\right)\end{aligned}$$

这里介绍的结论建议记住. 这个源自于一道作业题目，并且老师说过曾经有一道考试题直接考察这一结论的应用. 不妨举一个例子 (作业题目的最后一问，考试时候没有给这两个结论的提示直接考察).

对于高斯信源 $X \sim \mathcal{N}(0, \sigma^2)$，失真函数为 $d(x, \hat{x}) = 5(x - \hat{x})^2 + 3$，求解率失真函数.

解：

首先求得高斯信源在平方误差失真下的率失真函数

$$R_0(D) = \begin{cases} \frac{1}{2}\log\frac{\sigma^2}{D}, & 0 \leqslant D \leqslant \sigma^2, \\ 0, & D > \sigma^2. \end{cases}$$

利用上面的结论可知，在此失真度量情况下的率失真函数和平方误差失真下的率失真函数关系为

$$R(D) = R_0\left(\frac{D-3}{5}\right)$$

于是有

$$R(D) = \begin{cases} \frac{1}{2}\log\frac{5\sigma^2}{D-3}, & 3 \leqslant D \leqslant 5\sigma^2 + 3, \\ 0, & D > 5\sigma^2 + 3. \end{cases}$$

在前面部分我们讨论了伯努利信源在汉明失真下的率失真函数、高斯信源在平方误差失真下的率失真函数等，我们讨论了给定的失真函数如何引入互信息最小值的求解. 而**实际考试题目中，可能给出一个不同的失真函数**. 这时，我们有两种方案：

1. 转化：将问题转化为熟悉的失真函数下率失真函数求解的问题. 对于给出的失真函数是我们熟悉的失真函数的线性变换场景时，可以直接应用上面的结论进行转化.

2. 借鉴：根据给定的失真函数分析如何将其引入互信息最小值的求解，并按照本章中总结的率失真函数求解流程，模仿给出的两个例子的求解过程，完成率失真函数求解.

10.1.14 从信息的价值理解率失真

考虑课堂上讨论的概率模型

$$\begin{bmatrix} \boldsymbol{X} \\ p(x_i) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \\ 0.99 & 0.01 \end{bmatrix}$$

失真度量为

$$\begin{bmatrix} 0 & 1 \\ 100 & 0 \end{bmatrix}$$

1. 产品未经检验全部出厂：损失 $100 \times 0.01 = 1$ 元；

2. 产品未经检验全部报废：损失 $1 \times 0.99 = 0.99$ 元；

3. 检验完全正确：需要信息

$$I(X;\hat{X}) = H(X) - H(X \mid \hat{X}) = H(0.99) = 0.081\text{bit}$$

避免损失 0.99 元。由此可以计算信息的价值为 $0.99/0.081 = 12.2$ 元/bit；

4. 检验有一定误差 (设错判概率为 0.1)：首先计算 $\hat{X}$ 的概率分布

$$p(\hat{X}) = (0, 99, 0.01) \cdot \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix} = (0.892, 0.108)$$

因此可得，需要信息

$$I(X;\hat{X}) = H(\hat{X}) - H(\hat{X} \mid X) = H(0.892) - H(0.9) = 0.025\text{bit}$$

可能造成的损失为

$$D = 0.99 * 0.1 * 1 + 0.01 * 0.1 * 100 = 0.199$$

因此可以得到，这种策略避免了 $0.99 - 0.199 = 0.791$ 元的损失. 由此可以计算信息的价值为 $0.791/0.025 = 31.6$ 元/bit。

通过这个例子，我们不难发现，在允许失真的情况下，信息的价值是可能得到提高的.

这个例子也可以帮助我们更好地理解研究率失真理论的意义. 在本书开篇，我们讲到第十章的内容是关于连续信源的编码理论，实际上更严谨地说应该是关于允许失真条件下的信源编码理论. 对于连续信源来说，编码是一定存在失真的；对于离散信源来说，可以实现无失真编码，但是在实际应用场景下，研究在允许失真条件下的离散信源编码也是有价值的.

末章　信息论和信息安全

11.1　主要内容

这一章主要需要掌握的内容包括

11.1.1　Equivocation

Equivocation 代表收到密文后对明文或密钥还存在的不确定度，和序列长度有关，包括 $H(M \mid E)$ 和 $H(K \mid E)$. 这里我们主要需要掌握 Equivocation 的性质.

1. $H(M \mid E) \leqslant H(K \mid E)$

$$\begin{aligned} H(M \mid E) &\leqslant H(M \mid E) + H(K \mid M, E) \\ &= H(K, M \mid E) \\ &= H(K \mid E) + H(M \mid K, E) \\ &= H(K \mid E) \end{aligned}$$

2. $H(K \mid E) = H(M) + H(K) - H(E)$

$$\begin{aligned} H(K \mid E) &= H(K, E) - H(E) \\ &= H(K, E) + H(M \mid K, E) - H(E) \\ &= H(M, K, E) - H(E) \\ &= H(M, K) + H(E \mid M, K) - H(E) \\ &= H(M) + H(K) - H(E) \end{aligned}$$

11.1.2　冗余度

冗余度的定义为

$$D_N = \log G - H(M)$$

其中 G 为所有可能的 N 长序列明文个数.

平均冗余度的定义为

$$D = \frac{D_N}{N}$$

一个结论：N 长序列密文所提供的关于密钥的信息量小于冗余度.

$$\begin{aligned} H(K \mid E) &= H(K) + H(M) - H(E) \\ &\geqslant H(K) + H(M) - \log G \\ &= H(K) - (\log G - H(M)) \\ &= H(K) - D_N \end{aligned}$$

所以有

$$H(K) - H(K \mid E) = I(K; E) \leqslant D_N$$

11.1.3 唯一解距离

唯一解距离的定义为：将密钥惟一确定所平均需要的密文长度，即使得 $H(K \mid E) = 0$ 的 N.

$$\begin{aligned} 0 &= H(K \mid E) \\ &= H(K) + H(M) - H(E) \\ &= H(K) + N \log G - ND - N \log G \\ &= H(K) - ND \end{aligned}$$

所以可得

$$N = \frac{H(K)}{D}$$

11.1.4 完善保密性

完善保密性的充要条件为以下条件之一 (彼此等价)

1. $I(M; E) = 0$

   即明文密文独立，从密文中不能获取任何明文的信息.

2. $H(M \mid E) = H(M)$

   直接用互信息的展开式，可以得到这个结论和第一条结论等价.

3. $I(K; E) = H(E) - H(M \mid E)$

   由前面得到的结论

$$H(K \mid E) = H(M) + H(K) - H(E)$$

可得

$$I(K;E) = H(E) - H(M)$$

又

$$I(K;E) = H(E) - H(M \mid E)$$

所以有

$$I(M;E) = H(M) - H(M \mid E) = 0$$

关于完善保密性的必要条件 $H(K) \geqslant H(E)$，证明过程如下.

$$\begin{aligned} H(K) &= H(K \mid M) \\ &= H(E \mid M, K) + H(K \mid M) \\ &= H(E, K \mid M) \\ &= H(E \mid M) + H(K \mid E, M) \\ &\geqslant H(E \mid M) \\ &= H(E) \end{aligned}$$

自主测试试卷

1. (每题 4 分，共 8 分) 选择题 (多选题——指的是不定项选择)

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

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

   (b) 以下 D 元字母表上的码字长度符合即时码要求的是 (　　)

   $$(A) D = 2, l_i = 1,2,3,3,4$$
   $$(B) D = 2, l_i = 1,3,3,3,4,5,5$$
   $$(C) D = 4, l_i = 1,1,1,2,2,3,3,3,4$$
   $$(D) D = 5, l_i = 1,1,1,1,1,3,4$$

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) 存在两个不同的概率分布 $p_1 \geqslant p_2 \geqslant \cdots \geqslant p_n > 0$ 和 $q_1 \geqslant q_2 \geqslant \cdots \geqslant q_m > 0$ 使得

   $$H(p_1, p_2, \cdots, p_n) = H(q_1, q_2, \cdots, q_m)$$

   (b) 设一马尔可夫过程共有 $N \geqslant 2$ 个状态，且每个状态到所有 N 个状态的转移概率均不为 0，则当这些概率均相等时，该马尔可夫过程的熵率最大.

3. (每题 4 分，共 16 分) 填空题

   (a) 有三个二元离散随机变量 X, Y, Z，若要使得 $I(X;Y) = 1$ bit，$I(X;Y \mid Z) = 0$ bit，则 X, Y, Z 的联合概率分布为_____.

   (b) 有一连续信源 X，取值 x 处在 a_1 和 $a_2\,(a_1 < a_2)$ 之间. 该信源连接到某一信道, 信道输出 Y 的取值处在 b_1 和 $b_2\,(b_1 < b_2)$ 之间. 已知随机变量 X 和 Y 的联合概率密度函数为 $f(X,Y) = \frac{1}{(a_2-a_1)(b_2-b_1)}$，则互信息 $I(X;Y) =$_____.

   (c) 设一均匀分布的离散信源有 N 个符号，若 $N = 2^i + 1$，i 为正整数，则该信源的二元哈夫曼编码的平均码长为_____.

   (d) 设一有线电话信道，带宽限制在 $300 \sim 3400Hz$ 信噪比为 $P/N_0 = 1000$，则该信道的信道容量为_____.

4. (10 分) 设有两个离散信道，其输入分别为 X_1, X_2，输出分别为 Y_1, Y_2，对应这两个信道的转移概率分别为 $P_1(y \mid x), P_2(y \mid x)$.$X_1, X_2$ 的概率分布分别为 $Q_1(x), Q_2(x)$，X_1 和 X_2 的取值符号集的交集为空集，Y_1 和 Y_2 的取值符号集的交集也为空集.

   (a) 由这两个信道组成一个新的信道，新信道的输入符号 X 由 X_1 和 X_2 组成，输出符号 Y 由 Y_1 和 Y_2 组成. 新信道的输入符号是这样选择的：首先以 $\lambda(1 \geqslant \lambda \geqslant 0)$ 概率选取 X_1 或以 $1-\lambda$ 概率选取 X_2；然后以 $Q_1(x)$ 或 $Q_2(x)$ 概率分布选取相应符号集中的符号. 试求：$H(X)$ (用 $H(X_1), H(X_2), \lambda$ 表示).

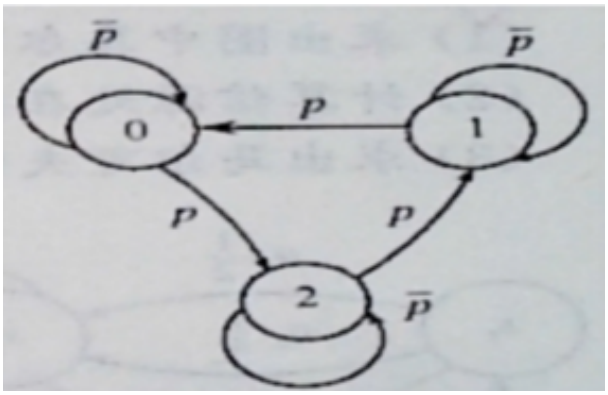
   (b) 新信道的转移概率满足

$$P(y \mid x) = \begin{cases} P_1(y \mid x) & x \in X_1, y \in Y_1 \\ P_2(y \mid x) & x \in X_2, y \in Y_2 \\ 0 & \text{其他} \end{cases}$$

   试求：$H(X \mid Y)$ (用 $H(X_1 \mid Y_1), H(X_2 \mid Y_2), \lambda$ 表示).

   (c) 试求：$I(X;Y)$ (用 $I(X_1;Y_1), I(X_2;Y_2), \lambda$ 表示).

5. (8 分) 设 $\{X_i\}$ 为时间不变的马尔可夫链，初始状态概率分布为 $P(X_1=0)=0.5$，$P(X_1=1)=0.25$，$P(X_1=2)=0.25$. 状态转移概率如图所示 ($\bar{p}=1-p$). 请计算该马尔可夫链的熵率.



6. (10 分) 有一离散信源 X，取值空间大小为 K，熵为 $H(X)$ 比特. 若该信源有一个三元前缀码满足平均码长

$$\bar{L}=\frac{H(X)}{\log 3}=H_3(X)$$

证明：

(a) 信源 X 的每个取值的概率分布均为 3^{-i} 的形式，其中 i 为正整数；

(b) K 一定为奇数.

7.（12 分）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).

8. (10 分) 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 $h(U)$，$h(V)$，$I(U;V)$.

9. (8 分) 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.

10. (10 分) 一个三维独立并联高斯信源 (X_1, X_2, X_3)，其中 X_1，X_2，X_3 均值都为零，方差分别是 2、8 和 4. 采用平方误差失真度量，$D = \sum_{i=1}^{3} D_i = \sum_{i=1}^{3} (x_i - \hat{x}_i)^2$，求该信源的信息率失真函数 $R(D)$.

自主测试试卷解析

1. (每题 4 分，共 8 分) 选择题 (多选题——指的是不定项选择)

(a) 设 X, Y, Z 均为离散随机变量，则以下等式或不等式正确的是 (AD)

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

解析：

(A) 正确. 根据等式左右两边的内容，确定需要使用熵的分解

$$H(X,Y,Z) = H(X,Z) + H(Y \mid X,Z)$$

结合互信息的分解

$$I(Z;Y \mid X) = H(Y \mid X) - H(Y \mid X,Z)$$

可得结论正确.

(B) 错误. 反例：X, Y 独立时左右两边相等.(熵的独立界公式是 $H(X+Y) \leqslant H(X) + H(Y)$，当且仅当独立时等号成立).

(C) 错误. 这个看起来就不正确 orz，不过为了严谨我们简单证明一下.

$$H(X \mid Y) = H(X,Y) - H(Y)$$

$$H(Y \mid X) = H(X,Y) - H(X)$$

显然，当 $H(X) \neq H(Y)$ 时等式不成立.

(D) 正确. 对于左边

$$H(X,Y,Z) - H(X,Y) = H(Z \mid X,Y)$$

对于右边

$$H(X,Z)-H(X)=H(Z\mid X)$$

由条件使得熵减少可知

$$H(Z\mid X,Y)\leqslant H(Z\mid X)$$

所以结论成立.

选择题很可能还会出这样的等式、不等式正误判断. 这个主要考察：概念、性质、链式法则以及重要不等式等. 一种简单粗暴的方式是左边减掉右边然后和 0 比较，不过很多时候，等式左右边是一组的可以一起进行运算，所以另一种方式就是，对左右两边分别化简 (对照着另一边，确定需要怎样化简，比如 (A) 中，左边是联合熵 $H(X,Y,Z)$，右边有 $H(X,Z)$，所以直接应用 $H(X,Y,Z)=H(X,Z)+H(Y\mid X,Z)$).

(b) 以下 D 元字母表上的码字长度符合即时码要求的是 (BC)

$$(A)D=2,l_i=1,2,3,3,4$$
$$(B)D=2,l_i=1,3,3,3,4,5,5$$
$$(C)D=4,l_i=1,1,1,2,2,3,3,3,4$$
$$(D)D=5,l_i=1,1,1,1,1,3,4$$

解析：

Kraft 不等式的考察一般也就是这样来考察了. 所以很可能还会考类似的题目. 难度较小，只要掌握 Kraft 不等式，计算即可.

(A)

$$2^{-1}+2^{-2}+2^{-3}\times 2+2^{-4}=2^{-1}+2^{-2}+2^{-2}+2^{-4}>1$$

(B)

$$2^{-1}+2^{-3}\times 3+2^{-4}+2^{-5}\times 2=2^{-1}+2^{-3}\times 3+2^{-4}+2^{-4}=2^{-1}+2^{-3}\times 4=1$$

(C)

$$4^{-1}\times 3+4^{-2}\times 2+4^{-3}\times 3+4^{-4}<4^{-1}\times 3+4^{-2}\times 2+4^{-3}\times 3+4^{-3}=4^{-1}\times 3+4^{-2}\times 3<4^{-1}\times 3+4^{-1}=$$

(D)

$$5^{-1}\times 5+5^{-3}+5^{-4}>1$$

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) 存在两个不同的概率分布 $p_1 \geqslant p_2 \geqslant \cdots \geqslant p_n > 0$ 和 $q_1 \geqslant q_2 \geqslant \cdots \geqslant q_m > 0$ 使得

$$H(p_1, p_2, \cdots, p_n) = H(q_1, q_2, \cdots, q_m)$$

(b) 设一马尔可夫过程共有 $N \geqslant 2$ 个状态，且每个状态到所有 N 个状态的转移概率均不为 0，则当这些概率均相等时，该马尔可夫过程的熵率最大.

(a) 对.

证明：

命题是存在两个不同概率分布满足熵相等，则只需要找到两个不同概率分布即可. 这里要特别注意，两个分布的取值个数不要求相等. 那么我们考虑一种简单的情形.

我们最熟悉的离散熵就是二元伯努利分布的熵 $H(p)$，我们知道 $0 \leqslant H(p) \leqslant 1$，当且仅当 $p = 1/2$ 时取得最大值，并且对于 [0,1] 之间的任意值 a，存在 $p \in [0,1]$ 使得 $H(p) = a$. 那么，我们举例只需要构造一个三元分布，使得其熵 $H(p_1, p_2, p_3) \in [0,1]$，即可.

例如 $H(0.8, 0.1, 0.1) = 0.92$ 比特 <1 比特. 所以一定存在 $p \in [0,1]$ 使得 $H(p) = H(0.8, 0.1, 0.1) = 0.92$，即

$$H(p, 1-p) = H(0.8, 0.1, 0.1)$$

(b) 对.

证明：

由马尔可夫过程熵率计算公式

$$H(\mathscr{X}) = \sum \mu_i H(\text{平稳分布的转移概率矩阵第}i\text{行})$$

而我们知道，对于离散熵，当均匀分布时熵最大. 于是有

$$H(\mathscr{X}) = \sum \mu_i H(\text{平稳分布的转移概率矩阵第}i\text{行}) \leqslant \sum \mu_i \log|\mathcal{X}| = \log|\mathcal{X}|$$

马尔可夫过程共有 $N \geqslant 2$ 个状态，且每个状态到所有 N 个状态的转移概率均不为 0，则当这些概率均相等时，平稳分布为均匀分布，恰好使得上述等式成立. 所以结论正确.

3. (每题 4 分，共 16 分) 填空题

(a) 有三个二元离散随机变量 X,Y,Z，若要使得 $I(X;Y)=1$ bit，$I(X;Y\mid Z)=0$ bit，则 X,Y,Z 的联合概率分布为_____.

(b) 有一连续信源 X，取值 x 处在 a_1 和 $a_2\,(a_1<a_2)$ 之间. 该信源连接到某一信道, 信道输出 Y 的取值处在 b_1 和 $b_2\,(b_1<b_2)$ 之间. 已知随机变量 X 和 Y 的联合概率密度函数为 $f(X,Y)=\frac{1}{(a_2-a_1)(b_2-b_1)}$，则互信息 $I(X;Y)=$_____.

(c) 设一均匀分布的离散信源有 N 个符号，若 $N=2^i+1$，i 为正整数，则该信源的二元哈夫曼编码的平均码长为_____.

(d) 设一有线电话信道，带宽限制在 $300\sim3400Hz$ 信噪比为 $P/N_0=1000$，则该信道的信道容量为_____.

答案：(a)$p(X=i,Y=j,Z=k)=p(X=1-i,Y=1-j,Z=1-k)=\frac{1}{2}$.(b)0. (c)$i+\frac{2}{2^i+1}$.(d)413 比特/秒.

解析：

(a) 我们考试的题目中也出现了一道和这个非常类似的题目，建议认真理解这种题目的考察点和解题方法. 首先处理第一个互信息等式，由于

$$I(X;Y)=H(X)-H(X\mid Y)\leqslant H(X)\leqslant\log|\mathcal{X}|=1\text{比特}$$

所以可知，等号取到，则有

$$H(X)=1\text{比特}\quad H(X\mid Y)=0$$

即 X 为均匀分布 $(p(X=0)=p(X=1)=\frac{1}{2})$，且给定 Y 时 X 确定，对于二元分布即 $Y=X$ 或者 $Y=-X$. 第二个互信息等式，有互信息为 0 的意义可知，给定 Z 的时候 X,Y 独立. 但是我们由第一个等式分析得到，二者给定一个另一个就唯一确定，那么独立只可能是一种情形，即给定 Z 的时候，X,Y 都为确定量 $(p(x,y\mid z)=p(x\mid z)p(y\mid z)=1)$.

综上所述，可得满足题意的概率分布为

$$p(X=i,Y=j,Z=k)=p(X=1-i,Y=1-j,Z=1-k)=\frac{1}{2}$$

(b) 由均匀分布的概率密度函数可知，其联合概率密度函数为边缘概率密度函数的乘积，即独立. 所以互信息为 0.

(c) 二元哈夫曼编码的码树为二叉树，由二叉树性质可知，该信源的哈夫曼码树前 i 层为满二叉树 (，其中 2^i-1 个节点为叶子节点，第 $i+1$ 层有 2 个节点.

所以可得

$$l(c)=\left(2^{i-1}\times i+2\times(i+1)\right)/\left(2^{i}+1\right)=i+\frac{2}{2^{i}+1}$$

有限带宽信道容量计算公式

$$C=W\log\left(1+\frac{P}{N_0W}\right)\text{ 比特/ 秒}$$

由题意知，$W=3400-300=3100Hz$，$P/N_0=1000$. 代入可得

$$C=3100\log\left(1+\frac{1000}{3100}\right)=413\text{ 比特/ 秒}$$

有限带宽信道容量计算的考察感觉也就是这种填空题形式了吧. 只要记住公式，不要管题目中的信噪比定义，计算即可.

4. (10 分) 设有两个离散信道，其输入分别为 X_1,X_2，输出分别为 Y_1,Y_2，对应这两个信道的转移概率分别为 $P_1(y\mid x),P_2(y\mid x)$.$X_1,X_2$ 的概率分布分别为 $Q_1(x),Q_2(x)$，X_1 和 X_2 的取值符号集的交集为空集，Y_1 和 Y_2 的取值符号集的交集也为空集.

   (a) 由这两个信道组成一个新的信道，新信道的输入符号 X 由 X_1 和 X_2 组成，输出符号 Y 由 Y_1 和 Y_2 组成. 新信道的输入符号是这样选择的：首先以 $\lambda(1\geqslant\lambda\geqslant0)$ 概率选取 X_1 或以 $1-\lambda$ 概率选取 X_2；然后以 $Q_1(x)$ 或 $Q_2(x)$ 概率分布选取相应符号集中的符号. 试求：$H(X)$ (用 $H\left(X_1\right),H\left(X_2\right),\lambda$ 表示).

   (b) 新信道的转移概率满足

$$P(y\mid x)=\begin{cases}P_1(y\mid x), & x\in X_1,y\in Y_1\\ P_2(y\mid x), & x\in X_2,y\in Y_2\\ 0, & \text{其他}\end{cases}$$

   试求：$H(X\mid Y)$ (用 $H\left(X_1\mid Y_1\right),H\left(X_2\mid Y_2\right),\lambda$ 表示).

   (c) 试求：$I(X;Y)$ (用 $I\left(X_1;Y_1\right),I\left(X_2;Y_2\right),\lambda$ 表示).

   解析：

   (a) 概率组合型信源. 这类题目的处理方法就是：引入示性变量. **重要方法！**

$$\theta=\begin{cases}1, & X=X_1\\ 2, & X=X_2\end{cases}$$

然后就可以利用联合熵的分解可得

$$\begin{aligned}
H(X) &= H(X) + H(\theta \mid X) \\
&= H(X, \theta) \\
&= H(\theta) + H(X \mid \theta) \\
&= H(\lambda) + \lambda H(X \mid \theta = 1) + (1-\lambda) H(X \mid \theta = 2) \\
&= H(\lambda) + \lambda H(X_1 \mid \theta = 1) + (1-\lambda) H(X_2 \mid \theta = 2) \\
&= H(\lambda) + \lambda H(X_1) + (1-\lambda) H(X_2)
\end{aligned}$$

(b) 类似的做法.

$$\theta = \begin{cases} 1, & X = X_1 \\ 2, & X = X_2 \end{cases}$$

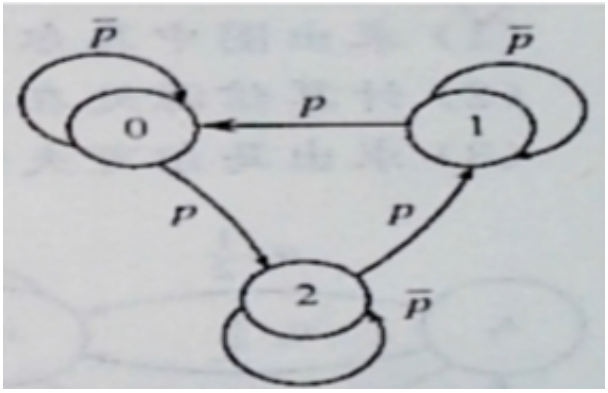
然后可得

$$\begin{aligned}
H(X \mid Y) &= H(X \mid Y) + H(\theta \mid X, Y) \\
&= H(X, \theta \mid Y) \\
&= H(\theta \mid Y) + H(X \mid Y, \theta) \quad \text{给定}Y, \theta\text{已知} \\
&= 0 + \lambda H(X \mid Y, \theta = 1) + (1-\lambda) H(X \mid Y, \theta = 2) \\
&= \lambda H(X_1 \mid Y_1, \theta = 1) + (1-\lambda) H(X_2 \mid Y_2, \theta = 2) \\
&= \lambda H(X_1 \mid Y_1) + (1-\lambda) H(X_2 \mid Y_2)
\end{aligned}$$

(c)

$$I(X;Y) = H(X) - H(X \mid Y) = H(\lambda) + \lambda I(X_1; Y_1) + (1-\lambda) I(X_2; Y_2)$$

5. (8 分) 设 $\{X_i\}$ 为时间不变的马尔可夫链，初始状态概率分布为 $P(X_1 = 0) = 0.5$，$P(X_1 = 1) = 0.25$，$P(X_1 = 2) = 0.25$. 状态转移概率如图所示 ($\bar{p} = 1 - p$). 请计算该马尔可夫链的熵率.



解析：

马尔科夫链熵率计算过程：计算平稳分布；代入熵率公式. 注意易错点在于，公式中需要用平稳分布，初始分布往往是干扰项，并不能直接使用.

由图示可知，转移概率矩阵为

$$P=\begin{pmatrix}1-p & 0 & p\\ p & 1-p & 0\\ 0 & p & 1-p\end{pmatrix}$$

设平稳分布为 $(p(X=0),p(X=1),p(X=2))=(\mu_1,\mu_2,\mu_3)$. 结合 $\mu_1+\mu_2+\mu_3=1$ 求解方程

$$(\mu_1,\mu_2,\mu_3)\cdot P=(\mu_1,\mu_2,\mu_3)$$

可得平稳分布为 (实际上根据对称性可以直接写出结果，不过过程上这么写严谨一些，不然就是根据对称性写出结果，代入上面的方程验证一下，也可以)

$$(\mu_1,\mu_2,\mu_3)=(\frac{1}{3},\frac{1}{3},\frac{1}{3})$$

代入马尔科夫链熵率公式

$$H(\mathscr{X})=\sum \mu_i H(\text{平稳分布的转移概率矩阵第}i\text{行})$$

可得

$$H(\mathscr{X})=\sum \mu_i H(0,p,1-p)=H(p)$$

6. (10 分) 有一离散信源 X，取值空间大小为 K，熵为 $H(X)$ 比特. 若该信源有一个三元前缀码满足平均码长

$$\bar{L}=\frac{H(X)}{\log 3}=H_3(X)$$

证明：

(a) 信源 X 的每个取值的概率分布均为 3^{-i} 的形式，其中 i 为正整数；

(b) K 一定为奇数.

证明：

(a) 由定理可知，对于 D 元即时码，有 $L\geqslant H_D(X)$，等号成立当且仅当 $D_{-l_i}=p_i$. 所以结论成立.

(b) 由 $L\geqslant H_D(X)$ 取等号可知，Kraft 不等式取等号，进而可知码树为满三叉树. 所以 K 一定为奇数.

或者考虑哈夫曼编码. 我们已知哈夫曼编码是最优编码，而由于已知存在一种编码可以使得码长达到下界——熵，那么 3 元哈夫曼编码的码长同样也是达到熵. 而这说明在哈夫曼编码时没有增加零概率项. 则由哈夫曼编码性质可知

$$K = 1 + k(D-1) = 1 + 2k$$

所以 K 一定为奇数.

7.（12 分）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).

解析：

这道题目考察离散无记忆信道的信道容量求解. 和我们熟悉的题目不同之处在于，这里没有给出信道的形式化描述，需要我们基于理解进行建模. 建模的方式就是：找信道的基本元素 (输入、输出、转移概率).

输入 X：字母表为 $\mathcal{X} = \{0, 1, \ldots, N\}$.

输出 Y：字母表为 $\mathcal{Y} = \{0, 1, \ldots, N\}$.

转移概率：$p(0 \mid 0) = 1$，对于 $i \in \{1, \ldots, N\}$ 有 $p(0 \mid i) = 1 - p$ 和 $p(i \mid i) = p$. 转移概率矩阵中其余元素均为 0.

完成了以上的建模工作，接下来要求解的就是我们熟悉的题目.

由于打电话和不打电话时的转移概率有所不同，可以看成两个输入按照概率的组合，也就是我们熟悉的概率组合信源问题 (第 4 题就是这种问题). 我们仍然引入示性变量

$$\theta = \begin{cases} 0, & X = 0 \\ 1, & 1 \leqslant X \leqslant N \end{cases}$$

然后按照求解信道容量的步骤正常求解. 首先分解互信息

$$I(X;Y) = H(Y) - H(Y \mid X)$$

引入示性变量以后，条件熵可以表达为

$$\begin{aligned}
H(Y \mid X) &= H(Y \mid X, \theta) + I(Y; \theta \mid X) \\
&= H(Y \mid X, \theta) \\
&= p(\theta = 0) H(Y \mid X, \theta = 0) + p(\theta = 1) H(Y \mid X, \theta = 1) \\
&= (1 - q) H(Y \mid X, \theta = 0) + q H(Y \mid X, \theta = 1) \\
&= (1 - q) \cdot 0 + q \sum_{x \in \mathcal{X}} p(x) H(Y \mid X = x) \\
&= q H(p)
\end{aligned}$$

其中用到了互信息 $I(Y; \theta \mid X)$ 为 0. 这是因为给定了 X 以后，θ 为确定量，所以 θ 和 Y 独立.

接下来求解输出的熵的最大值. 首先我们已知 $p(X = 0) = 1 - q$，则可知

$$p(Y = 0) = 1 - q + q(1 - p) = 1 - pq$$

所以可知，达到最大熵时 Y 的分布为

$$p(Y = 0) = 1 - pq, p(Y = i) = \frac{pq}{N}, \text{其中} i \in \{1, \ldots, N\}$$

此时对应的输入 X 的分布为

$$p(X = 0) = 1 - q, p(X = i) = \frac{q}{N}, \text{其中} i \in \{1, \ldots, N\}$$

综上所述，信道容量为

$$\begin{aligned}
C &= \max_{p(x)} I(X; Y) \\
&= \max_{p(x)} H(Y) - H(Y \mid X) \\
&= \max_{p(x)} H(Y) - q H(p) \\
&= H(1 - pq, \frac{pq}{N}, \frac{pq}{N}, \ldots, \frac{pq}{N}) - q H(p) \text{比特/信道使用}
\end{aligned}$$

8. (10 分) 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 $h(U)$，$h(V)$，$I(U; V)$.

解析：

猜测一定会有一道的微分熵大题. 这类题目主要利用概念、性质、链式法则等，尤其注意，线性变换对于微分熵的影响的这个性质的应用.

首先根据联合概率密度函数可以分析得到 $X \sim \mathcal{N}(m_x, \sigma_x^2)$，$Y \sim \mathcal{N}(m_y, \sigma_y^2)$，$X, Y$ 独立. 则可知 $U = X+Y \sim \mathcal{N}(m_x+m_y, \sigma_x^2+\sigma_y^2)$，$V = X-Y \sim \mathcal{N}(m_x-m_y, \sigma_x^2+\sigma_y^2)$. 特别要注意的是，正态分布的差的分布中，均值为二者均值之差，方差为二者之和.

于是可得

$$h(U) = h(V) = \frac{1}{2}\left[\log 2\pi e(\sigma_x^2 + \sigma_y^2)\right] \text{比特}$$

利用互信息的分解式

$$I(U;V) = h(U) + h(V) - h(U,V) = \left[\log 2\pi e(\sigma_x^2 + \sigma_y^2)\right] - h(U,V)$$

这时我们需要用到线性变换对于微分熵的影响这一性质. 我们在前面有介绍过

$$h(A\boldsymbol{X}) = h(\boldsymbol{X}) + \log|A|$$

对于本题，有

$$\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}$$

其中

$$A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

其行列式为 $|A| = 2$. 进而可得

$$h(U,V) = h(X,Y)+1 = h(X)+h(Y)+1 = \frac{1}{2}\left[\log 2\pi e(\sigma_x^2)\right]+\frac{1}{2}\left[\log 2\pi e(\sigma_y^2)\right]+1 \text{比特}$$

所以可得，互信息为

$$\begin{aligned} I(U;V) &= h(U) + h(V) - h(U,V) \\ &= \left[\log 2\pi e(\sigma_x^2 + \sigma_y^2)\right] - \left[\frac{1}{2}\left[\log 2\pi e(\sigma_x^2)\right] + \frac{1}{2}\left[\log 2\pi e(\sigma_y^2)\right] + 1\right] \\ &= \log\frac{\sigma_x^2 + \sigma_y^2}{\sigma_x\sigma_y} - 1 \quad \text{比特} \end{aligned}$$

要注意的是，如果没有想到利用这个性质来计算联合微分熵，很可能会说服自己 U, V 独立，但实际上这个是不正确的.

9. (8 分) 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.

解析：

猜测一定会有一道的连续信道的信道容量计算大题. 信道噪声和输入限制可能会和我们熟悉的高斯信道不同，不过求解方法是类似的.

首先分解互信息.

$$\begin{aligned} I(X;Y) &= h(Y) - h(Y \mid X) \\ &= h(Y) - h(X + Z \mid X) \\ &= h(Y) - h(Z) \\ &= h(Y) - \log e\mu \quad \text{比特} \end{aligned}$$

由题意知，$E[Y] = E[X + Z] = E[X] + E[Z] \leqslant \lambda + \mu$. 我们知道均值受限时，指数分布时取得最大微分熵. 所以有

$$\begin{aligned} C &= \max I(X;Y) \\ &= \max h(Y) - \log e\mu \\ &= \log e(\lambda + \mu) - \log e\mu \\ &= \log\left(1 + \frac{\lambda}{\mu}\right) \text{比特} \end{aligned}$$

由 $Y = X + Z$，以及 Y, Z 服从指数分布，可以计算得到取得信道容量时的 X 的分布.

引入辅助变量 $W = Z$ 构成可逆变换

$$\begin{cases} Y = X + W \\ Z = W \end{cases}$$

计算雅可比行列式的绝对值 $||J|| = 1$. 于是可得

$$\begin{aligned} l(x) &= \int_0^{+\infty} f(x+w, w)dw \\ &= \int_0^{+\infty} f_Y(x+w)f_Z(w)dw \\ &= \int_0^{+\infty} \frac{1}{\lambda+\mu} e^{-\frac{x+w}{\lambda+\mu}} \frac{1}{\mu} e^{-\frac{w}{\mu}} dw \\ &= \frac{1}{\lambda+2\mu} e^{\frac{x}{\lambda+\mu}} \end{aligned}$$

即为达到信道容量时 X 的概率分布. **注意指数分布和正态分布不同，独立指数分布的和不是指数分布.**

10. (10 分) 一个三维独立并联高斯信源 (X_1, X_2, X_3)，其中 X_1，X_2，X_3 均值都为零，方差分别是 2、8 和 4. 采用平方误差失真度量，$D = \sum_{i=1}^{3} D_i = \sum_{i=1}^{3} (x_i - \hat{x}_i)^2$，求该信源的信息率失真函数 $R(D)$.

    解析：

    考察并联高斯信源的反注水法. 老师说每年考试最后一道题都是率失真这一章的题目，往往和作业题目类似甚至是原题. 一般来讲，会考的题目可能包括：离散信源 (二元、多元) 的率失真函数计算、高斯信源的率失真函数计算、并联高斯信源的反注水法、失真函数存在无穷的类型、失真函数线性变换问题.

    直接由反注水法可得

    - $0 < D < 2 \times 3 = 6$ 时：
      失真分配为平均分配 $(\frac{D}{3}, \frac{D}{3}, \frac{D}{3})$. 此时率失真函数为

      $$R(D) = \sum_{i=1}^{3} \frac{1}{2} \log \frac{\sigma_i^2}{D_i} = \frac{1}{2} \log \frac{1728}{D^3}$$

    - $6 \leqslant D < 2 + 4 \times 2 = 10$ 时：
      失真分配为 $(2, \frac{D}{2}, \frac{D}{2})$. 此时率失真函数为

      $$R(D) = \sum_{i=1}^{3} \frac{1}{2} \log \frac{\sigma_i^2}{D_i} = \frac{1}{2} \log \frac{64}{D^2} = \frac{8}{D}$$

    - $10 \leqslant D < 2 + 4 + 8 = 14$ 时：

失真分配为 $(2,4,D)$. 此时率失真函数为

$$R(D)=\sum_{i=1}^{3}\frac{1}{2}\log\frac{\sigma_i^2}{D_i}=\frac{1}{2}\log\frac{8}{D}=$$

- $D\geqslant 14$ 时：
  失真分配为 $(2,4,D)$. 此时率失真函数为

$$R(D)=0$$

综上所述，率失真函数为

$$R(D)=\begin{cases}\frac{1}{2}\log\frac{1728}{D^3}, & 0<D<6,\\ \log\frac{8}{D}, & 6\leqslant D<10,\\ \frac{1}{2}\log\frac{8}{D}, & 10\leqslant D<14,\\ 0, & D>14.\end{cases}$$

这种题目就是考察反注水法的使用. 其中易错点在于，容易和注水法搞混，这里当失真平均值大于信源方差时，并不是不分配失真，而是分配的失真等于信源的方差. 不过在计算时候应该也可以反应过来，毕竟如果按照不分配失真 $D_i=0$ 计算，则这个信源的率失真函数趋于无穷.

后记

终于还是决定写了这本《信息论学习指导》，也算是给两次信息论助教经历的一个总结和交待. 信息论课程是信息安全专业/网络空间安全学院难度较高的课程之一，希望这本学习指导能够帮助到学习这门课程的学弟学妹，可以顺利完成课程学习.

想说什么却又不知道从何说起，就分享一首歌作为结尾吧.

《海底》
原唱：一支榴莲
散落的月光穿过了云
躲着人群
铺成大海的鳞
海浪打湿白裙
试图推你回去
海浪清洗血迹
妄想温暖你
往海的深处听
谁的哀鸣在指引
灵魂没入寂静
无人将你吵醒
你喜欢海风咸咸的气息
踩着湿湿的沙砾
你说人们的骨灰应该撒进海里
你问我死后会去哪里
有没有人爱你
世界能否不再
总爱对凉薄的人扯着笑脸
岸上人们脸上都挂着无关
人间毫无留恋
一切散为烟

散落的月光穿过了云
躲着人群
溜进海底
海浪清洗血迹
妄想温暖你
灵魂没入寂静
无人将你吵醒
你喜欢海风咸咸的气息
踩着湿湿的沙砾
你说人们的骨灰应该撒进海里
你问我死后会去哪里
有没有人爱你
世界已然将你抛弃
总爱对凉薄的人扯着笑脸
岸上人们脸上都挂着无关
人间毫无留恋
一切散为烟
来不及来不及
你曾笑着哭泣
来不及来不及
你颤抖的手臂
来不及来不及
无人将你打捞起
来不及来不及
你明明讨厌窒息