信息论

信号传输与处理的理论基础

大数定律、典型集的渐进均分性质

教程 第3章

预备知识：大数定律（1）

* Markov不等式（习题3.1.a）
* X是取值非负的随机变量，E[X]是其均值（数学期望），t是任意的正数，则恒有不等式 $P[X \geq t] \leq E[X]/t$。

【注】⊙该不等式给出X偏离某个数值t的概率大小的估计。

⊙以上不等式适合于任何概率分布，只要X取值非负。

⊙以下的证明是针对离散型的X，请将其修改推广到连续型随机变量。

* 证：$E[X] = \sum_x xP(x) \geq \sum_{x \geq t} xP(x) \geq t\sum_{x \geq t} P(x) = tP[X \geq t]$.

【思考】该命题为什么实质性地依赖于X取值非负这一条件？

*

预备知识：大数定律（2）

* Chebyshev不等式（习题3.1.b）
* X是随机变量，μ是其均值，σ^2是其方差，t是任意的正数，则恒有不等式 $P[|X-\mu| \geq t] \leqq \sigma^2/t^2$。

【注】⊙该不等式给出X偏离其均值程度为t的概率大小的估计。
⊙以上不等式适合于任何概率分布，无须X取值非负。

* 证：考虑随机变量$Y=|X-\mu|^2$，注意到$E[Y]=\sigma^2$, 因此应用Markov不等式得到
* $P[|X-\mu| \geq t] = P[|X-\mu|^2 \geq t^2] = P[Y \geq t^2] \leqq \sigma^2/t^2$。
*
*

预备知识：大数定律（3）

* 弱大数定律（习题3.1.c）
* 设$X_1,\ldots,X_n$是两两独立、同分布的随机变量，分布的
* 均值为μ、方差为σ^2，定义统计均值$\overline{X_n} = \frac{1}{n}\sum_{j=1}^{n} X_j$，则对
* 任何正数t>0有 $P[|\overline{X_n} - \mu| > t] \leqq \sigma^2/nt^2$。
* 证明 第一步：计算$\bar{X}$ 的方差Δ_n:
* $\Delta_n = E[|\bar{X}|^2] - E[\bar{X}]^2$，其中$E[\bar{X}] = \frac{1}{n}\sum_{j=1}^{n} E[X_j] = \frac{1}{n}\sum_{j=1}^{n}\mu = \mu$，
* $E[|\bar{X}|^2] = \frac{1}{n^2}\sum_{j=1}^{n} E[X_j{}^2] + \frac{1}{n^2}\sum_{i,j=1,i\neq j}^{n} E[X_i X_j]$
* $= \frac{1}{n^2}\sum_{j=1}^{n} E[X_j{}^2] + \frac{1}{n^2}\sum_{i,j=1,i\neq j}^{n} E[X_i]E[X_j]$
* $= \frac{1}{n^2}\sum_{j=1}^{n}(\sigma^2 + \mu^2) + \frac{1}{n^2}\sum_{i,j=1,i\neq j}^{n} \mu^2 = \sigma^2/n + \mu^2/n$ 【请完成最后的代数运算】
* 因此$\Delta_n = \sigma^2/n$。
* 第二步：对计算$\bar{X}$ 应用Chebeyshev不等式【请自行完成】。证毕。

预备知识：大数定律（4）

* 弱大数定律的一个极限形式：
* 设$X_1,\dots,X_n$是两两独立、同分布的随机变量，分布的
* 均值为μ、方差为σ^2，则统计均值$\overline{X_n}=\frac{1}{n}\sum_{j=1}^{n}X_j$满足概率收敛性质 $\lim_{n\to\infty} P[|\overline{X_n}-\mu|>t]=0$。

【注】以上性质称为$\overline{X_n}$依概率收敛到μ，记为$\overline{X_n}\xrightarrow{P}\mu$；

等价地，$\lim_{n\to\infty} P[|\overline{X_n}-\mu|<t]=1$对任何t>0成立。

* 弱大数定律的应用：定理3.1.1
* 设$X_1,\dots,X_n$是两两独立、同分布的随机变量，概率分布
* 函数为$P(x)$, 则$n\to\infty$时有 $-\frac{1}{n}\log P(X_1,\dots,X_n)\xrightarrow{P}H[X]$.

【习题】应用弱大数定律证明该结论。

预备知识：大数定律（5）

* 强大数定律（补充知识，今后暂不应用）
* 设$X_1,\ldots,X_n$是两两独立、同分布的随机变量，分布的
* 均值为μ、方差为σ^2，则统计均值$\overline{X_n}=\frac{1}{n}\sum_{j=1}^{n}X_j$的概率分布
* 渐进为Gauss分布$N(\mu,\sigma^2/n)$。

* 弱大数定律仅断言$\overline{X_n}$依概率收敛到均值μ，未能对随机
* 变量$\overline{X_n}$的概率分布做出任何结论。强大数定律更进一步断
* 言，无论X_j原来具有何种概率分布，只要n足够大，$\overline{X_n}$就
* 接近于Gauss分布！
* 这表明足够多的独立样本被平均后，本性的概率特征被彼此抵消，而越来越表现出Gauss分布的特征。这也表明Gauss为什么在通信系统中具有很强的普适性，特别是在大量独立随机噪声叠加的情况下。

典型集合及其渐进均分性质（1）

- 1 典型集合的概念
- 设$X_1,\ldots,X_n$是两两独立、同分布的随机变量，概率分布
- 函数为$P(x)$，$\varepsilon>0$是一个参量。与ε和n相关的典型集合
- $A(n,\varepsilon) \equiv \{(x_1,\ldots,x_n): 2^{-n(H(X)+\varepsilon)} \leq p(x_1,\ldots,x_n) \leq 2^{-n(H(X)-\varepsilon)}\}$
- 等价地
- $A(n,\varepsilon) = \{(x_1,\ldots,x_n): -\varepsilon \leq -\frac{1}{n}\log_2 p(x_1,\ldots,x_n) - H[X] \leq \varepsilon\}$

典型集$A(n,\varepsilon)$ 是根据熵$H[X]$在样本空间中所界定的一个特殊的子集。

- 前述的大数定律已表明 $-\frac{1}{n}\log_2 p(x_1,\ldots,x_n) \xrightarrow{P} H[X]$，因此不难预期典型集合
- 具有很特殊的性质。下面的渐进均分性质正是如此。

典型集合及其渐进均分性质（2）

* 2 渐进均分性质 定理3.1.2

* (1) 对典型集合$A(n,\varepsilon)$中的每个样本$(x_1,\ldots,x_n)$，有
* $$\left| -\frac{1}{n}\log_2 p(x_1,\ldots,x_n) - H[X] \right| \le \varepsilon$$
* (2) 对充分大的n，有$P[A(n,\varepsilon)] > 1-\delta$.
* (3) 典型集合的大小有上界$|A(n,\varepsilon)| \le 2^{n(H(X)+\varepsilon)}$.
* (4) 典型集合的大小数有下界$|A(n,\varepsilon)| \ge (1-\delta)2^{n(H(X)-\varepsilon)}$,
* δ是属于区间$(0,1)$并随ε减小而减小的正数.

典型集合及其渐进均分性质（3）

* 3 渐进均分性质的证明

* (1) 是$A(n,\varepsilon)$的定义的等价形式。
* (2) 是大数定律 $-\frac{1}{n}\log_2 p(x_1,\ldots,x_n) \xrightarrow{P} H[X]$的等价形式：
* $$\lim_{n\to\infty} P[(x_1,\ldots,x_n)\text{属于}A(n,\varepsilon)] = \lim_{n\to\infty} P[|-\frac{1}{n}\log_2 p(x_1,\ldots,x_n) - H[X]| \le \varepsilon] = 1.$$
* (3) 的证明概要：
* $$1 = \sum_{(x1,\ldots,xn)} P(x_1,\ldots,x_n) \ge \sum_{(x1,\ldots,xn)\text{属于}A(n,\varepsilon)} P(x_1,\ldots,x_n)$$
* $$\ge \sum_{(x1,\ldots,xn)\text{属于}A(n,\varepsilon)} 2^{-n(H(X)+\varepsilon)} = 2^{-n(H(X)+\varepsilon)}|A(n,\varepsilon)|$$

* (4) 的证明概要：由(2)知当n充分大时有
* $$1-\delta \le P[A(n,\varepsilon)] = \sum_{(x1,\ldots,xn)\text{属于}A(n,\varepsilon)} P(x_1,\ldots,x_n)$$
* $\le \sum_{(x1,\ldots,xn)\text{属于}A(n,\varepsilon)} 2^{-n(H(X)-\varepsilon)} = 2^{-n(H(X)-\varepsilon)}|A(n,\varepsilon)|$ 。证毕。

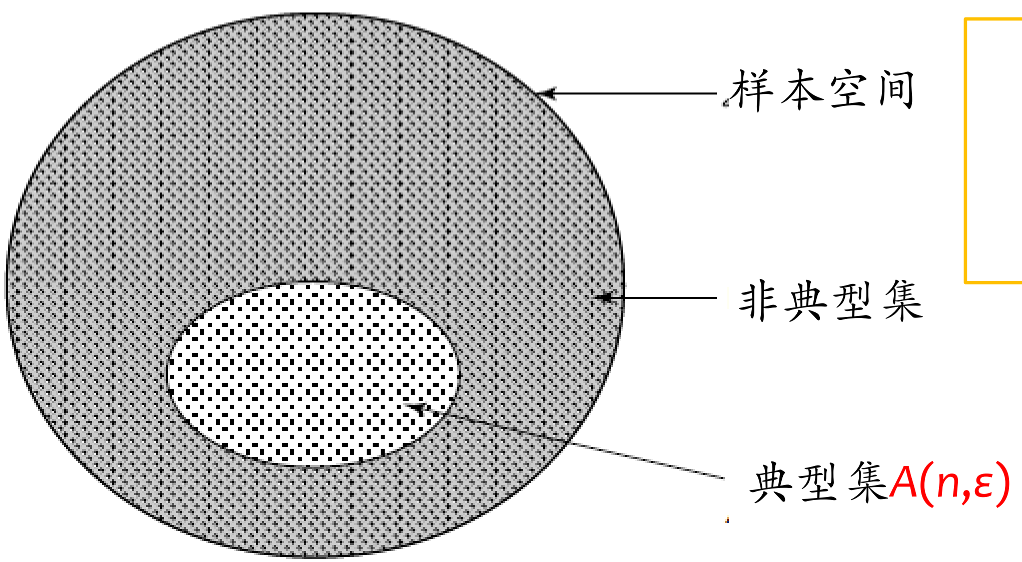
典型集合及其渐进均分性质（3A）

* 3 渐进均分性质的涵义

*

* 渐进均分性质表明典型集合具有一个不寻常的特性特性：
* 随着观测样本数n越来越大，$A(n,\varepsilon)$一方面是一个相对大小越来越小的子集，
* 另一方面，$A(n,\varepsilon)$所承载的概率却越来越大，即：
* 随着n越来越大，越来越频繁地观测到落在$A(n,\varepsilon)$中的序列样本$(x_1,\ldots,x_n)$！

*



$$\lim_{n\to\infty} |A(n,\epsilon)|/2^n = 0$$

$$\lim_{n\to\infty} P[A(n,\epsilon)] = 1$$

典型集合及其渐进均分性质（4）

4 联合典型集及其渐进均分性质(7.6节定理7.6.1)

* 设$X_1,\dots,X_n$是两两独立、同分布的随机变量，概率分布函数为$P(x)$；$Y_1,\dots,Y_n$也是两两独立、同分布的随机变量，概率分布函数为$P(y)$；每对(X_i,Y_i)具有联合的概率分布$P(x,y)$；$\varepsilon>0$是一个参量。与参数ε和n相关的联合典型集合
* $A(n,\varepsilon)\equiv\{(x_1,y_1,\dots,x_n,y_n)$: $2^{-n(H(X)+\varepsilon)}\le p(x_1,\dots,x_n)\le 2^{-n(H(X)-\varepsilon)}$,
* $2^{-n(H(Y)+\varepsilon)}\le p(y_1,\dots,y_n)\le 2^{-n(H(Y)-\varepsilon)}$,
* $2^{-n(H(X,Y)+\varepsilon)}\le p(x_1,y_1,\dots,x_n,y_n)\le 2^{-n(H(X,Y)-\varepsilon)}\}$
* 等价地
* $A(n,\varepsilon)=\{(x_1,\dots,x_n)$: $-\varepsilon\le -\frac{1}{n}\log_2 p(x_1,\dots,x_n)-H[X]\le\varepsilon$,
* $-\varepsilon\le -\frac{1}{n}\log_2 p(y_1,\dots,y_n)-H[Y]\le\varepsilon$,
* $-\varepsilon\le -\frac{1}{n}\log_2 p(x_1,y_1\dots,x_n,y_n)-H[X,Y]\le\varepsilon\}$

典型集合及其渐进均分性质（5）

* 4 联合典型集合及其渐进均分性质(续)

* (1) $\lim\limits_{n\to\infty}$ $P[(x_1,y_1,\dots,x_n,y_n)\in A(n,\varepsilon)]=1$.
* (2) 联合典型集的大小有上界$|A(n,\varepsilon)|\le 2^{n(H(X,Y)+\varepsilon)}$.
* (3) 联合典型集的大小数有下界$|A(n,\varepsilon)|\ge (1-\delta)2^{n(H(X,Y)-\varepsilon)}$,
* δ是属于区间(0,1)并随ε减小而减小的正数.
* (4) 若$X_1^*,\dots X_n^*$是两两独立、同分布的随机变量，概率分布函数为$P(x)$； $Y_1^*,\dots,Y_n^*$也是两两独立、同分布的随机变量，概率分布函数为$P(y)$；每对(X_i^*,Y_i^*)具有联合的概率分布$\underline{P(x)P(y)}$,
* 则 $P[(x_1^*,y_1^*,\dots,x_n^*,y_n^*)\in A(n,\varepsilon)]\le 2^{-n(I(X,Y)-3\varepsilon)}$
* $P[(x_1^*,y_1^*,\dots,x_n^*,y_n^*)\in A(n,\varepsilon)]\ge (1-\delta)2^{-n(I(X,Y)+3\varepsilon)}$

典型集合及其渐进均分性质（6）

* 4 联合典型集的渐进均分性质(续)：证明
* (1) 是大数定律的等价形式：
* (2) 的证明概要：
* $1 = \sum_{(x1,y1\ldots,xn,yn)} P(x_1, y_1, \ldots, x_n, y_n)$
* $\geq \sum_{(x1,y1,\ldots,xn,yn)属于A(n,\varepsilon)} P(x_1, y_1 \ldots, x_n, y_n)$
* $\geq \sum_{(x1,y1,\ldots,xn,yn)属于A(n,\varepsilon)} 2^{-n(H(X,Y)+\varepsilon)} = 2^{-n(H(X,Y)+\varepsilon)}|A(n,\varepsilon)|$

* (3) 的证明概要：由(2)知当n充分大时有

$1-\delta \leq \mathrm{P}[A(n,\varepsilon)] = \sum_{(x1,y1\ldots,xn,yn)属于A(n,\varepsilon)} P(x_1, y_1 \ldots, x_n, y_n)$

$\leq \sum_{(x1,y1,\ldots,xn,yn)属于A(n,\varepsilon)} 2^{-n(H(X,Y)-\varepsilon)} = 2^{-n(H(X,Y)-\varepsilon)}|A(n,\varepsilon)|$ 。

典型集合及其渐进均分性质（7）

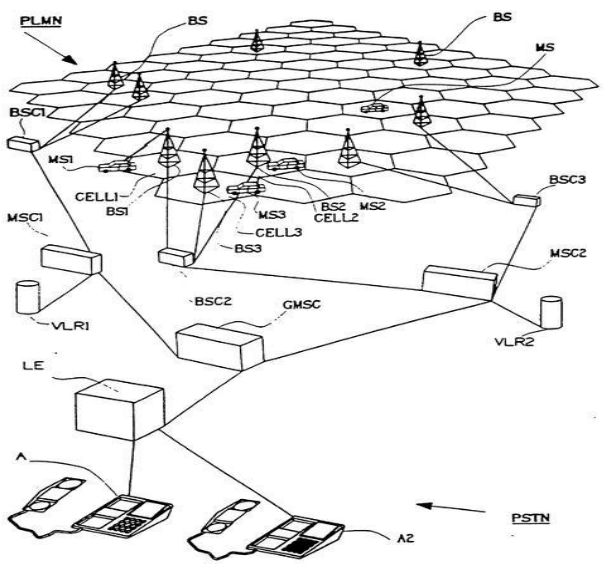
* 4 联合典型集的渐进均分性质(续)：证明
* (4) 的证明概要：
* $P[(x_1^*,y_1^*,\ldots,x_n^*,y_n^*)\in A(n,\varepsilon)]$
* $=\sum_{(x1*,y1*\cdots,xn^*,yn^*)属于A(n,\varepsilon)} P(x_1^*,y_1^*,\ldots,x_n^*,y_n^*)$
* $=\sum_{(x1*,y1*\cdots,xn^*,yn^*)属于A(n,\varepsilon)} P(x_1^*)P(y_1^*)\ldots P(x_n^*)P(y_n^*)$
* $=\sum_{(x1*,y1*\cdots,xn^*,yn^*)属于A(n,\varepsilon)} P(x_1^*,\ldots,x_n^*)P(y_1^*,\ldots,y_n^*)$
* $\leq\sum_{(x1*,y1*\cdots,xn^*,yn^*)属于A(n,\varepsilon)} 2^{-n(H(X)-\varepsilon)}\,2^{-n(H(Y)-\varepsilon)}$
*
* $=2^{-n(H(X)+H(Y)-2\varepsilon)}|A(n,\varepsilon)|$
* $\leq 2^{n(H(X,Y)+\varepsilon)}\,2^{-n(H(X)+H(Y)-2\varepsilon)}$
* $=2^{-n(H(X)+H(Y)-H(X,Y)-3\varepsilon)}=2^{-n(I(X;Y)-3\varepsilon)}$
* 另一个不等式仿定理3.2.1的(4)和上述分析证明【习题】。

习题集部分解答（1）

* 第三章习题
* 习题3.2、3.4、3.5、3.6、3.9、3.11（3.3节）、
* 习题7.37（又一个典型集合及其渐进均分性质）；
* 补充习题：结合本章结果和习题7.37，试定义一组m个随机变量的典型集合并猜测其渐进均分性质。你能证明你的猜测吗？

第二单元结束

* 下一讲开始学习单三单元：理论的应用（约3周）
* 教程第一版第七章 信道容量7.1-7.7、7.11（该节以补充的更完整的线性分组编码及其译码算法和性能分析为主）；
* 教程第一版第九章 Gauss信道 9.1-9.5。



Фиг.1a