信息论基础习题课

助教：刘志强

时间：2015-10-26

作业4

3.4 AEP

设X_i为$i.i.d. \sim p(x), x \in \{1,2,\ldots,m\}, \mu = EX$ 以及$H = -\sum p(x)\log p(x)$。

设$A^n = \{x^n \in \mathcal{X}^n : |-\frac{1}{n}\log p(x^n) - H| \le \varepsilon\}$,

$B^n = \{x^n \in \mathcal{X}^n : |\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| \le \varepsilon\}$。

(a)$\Pr\{X^n \in A^n\} \to 1$吗?

(b)$\Pr\{X^n \in A^n \cap B^n\} \to 1$吗?

(c)证明：对任意的n,$|A^n \cap B^n| \le 2^{n(H+\varepsilon)}$

(d)证明：当n充分大时,$|A^n \cap B^n| \ge (1/2)2^{n(H-\varepsilon)}$

3.4 AEP

(a) 由AEP的定理3. 1. 2 可显然得到

(b) 由容斥原理

$$\Pr(A^n \cap B^n) = \Pr(A^n) + \Pr(B^n) - \Pr(A^n \cup B^n)$$

由大数定理 $\Pr(B^n) \to 1$, 且 $\Pr(A^n) \to 1$, $\Pr(A^n \cup B^n) \to 1$

$$\therefore \Pr(A^n \cap B^n) \to 1$$

3.4 AEP

$$c)\ |A^n \cap B^n| \le |A^n| \le 2^{n(H+\varepsilon)}.$$

$d) \because \Pr(A^n \cap B^n) \to 1$, 且$A^n$为典型集，$\mathrm{p}(\mathrm{x}^n) \le 2^{-n(H-\varepsilon)}$

$\therefore$ n充分大时，$0.5 \le \Pr(A^n \cap B^n) \le \sum_{x \in A^n \cap B^n} 2^{-n(H-\varepsilon)}$

$= |A^n \cap B^n| \, 2^{-n(H-\varepsilon)}$,

$\therefore |A^n \cap B^n| \ge 0.5 * 2^{n(H-\varepsilon)} \ge (\frac{1}{2}) 2^{n(H-\varepsilon)}$

3.5 由概率定义的集合

3.5 由概率定义的集合

$C_n(t)=\{x^n \in \chi^n : p(x^n) \geq 2^{-nt}\}$

(a)证明$|C_n(t)| \leq 2^{nt}$

(b)当t为何值时$P(X^n \in C_n(t)) \to 1$

(a)反证法，若不满足，则概率和大于1

(b)由渐进均分性知，n无穷大时$p(x^n)$等效于均匀分布，故

故$2^{-nt} < 2^{-nH(X)}, t > H(X)$.(等号不可以取得)

3.9 AEP

AEP. Let $X_1, X_2, \ldots$ be independent identically distributed random variables drawn according to the probability mass function $p(x), x \in \{1, 2, \ldots, m\}$. Thus $p(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p(x_i)$. We know that $-\frac{1}{n}\log p(X_1, X_2, \ldots, X_n) \to H(X)$ in probability. Let $q(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} q(x_i)$, where q is another probability mass function on $\{1, 2, \ldots, m\}$

（a）计算 $\lim -\frac{1}{n}\log q(X_1, X_2, \ldots, X_n)$

解：

$$\begin{aligned}
\lim -\frac{1}{n}\log q(X_1, X_2, \ldots, X_n) &= \lim -\frac{1}{n}\sum \log q(X_i) \\
&= -E(\log q(X)) \text{ w.p. } 1 \\
&= -\sum p(x)\log q(x) \\
&= \sum p(x)\log\frac{p(x)}{q(x)} - \sum p(x)\log p(x) \\
&= D(\mathbf{p}||\mathbf{q}) + H(\mathbf{p}).
\end{aligned}$$

3.9 AEP

（b）说明当p为真实分布时，偏好分布q的优势将以指数衰减。

(b) Now evaluate the limit of the log likelihood ratio $\frac{1}{n}\log\frac{q(X_1,\ldots,X_n)}{p(X_1,\ldots,X_n)}$ when $X_1, X_2, \ldots$ are i.i.d. $\sim p(x)$. Thus, the odds favoring q are exponentially small when p is true.

计算得 $\frac{1}{n}\log\frac{q(X_1,X_2,\ldots,X_n)}{p(X_1,X_2,\ldots,X_n)} \to -D(p(x)\|q(x))$

可知该极限为负的相对熵。似然比会接近于$2^{-nD(p(x)\|q(x))}$，

是一个指数小的值（exponentially small），在似然比检验理论中，

应该接受p(x).

AEP的简单分析

- 大数定理

如果有n个独立同分布的随机变量X，它们有共同的均值m，那么这n个变量的平均数就会随着n的增大而趋近m。

辛钦大数定理：$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} a_i - \mu\right| < \varepsilon\right) = 1$

- 渐进均分性

n个独立分布的随即变量X，且分布函数为p(x)，则：$-\frac{1}{n}\log p(X_1, X_2, \ldots, X_n) \to H(X)$

- 对比：

大数定理是对n个独立分布的随机变量值求均值，而渐进均分性定理则是对n个独立分布的随机变量的概率的 对数值进行求均值。即渐进均分性是对随机变量的概率的概率性质。

典型集的简单分析

- 典型集：
- 特点：
  - 典型集的概率近似为1
  - 典型集中的所有元素几乎都等概率
  - 典型集中的元素个数近似为 2^{nH}
  - 总集合中的元素个数为$|\chi|^n$

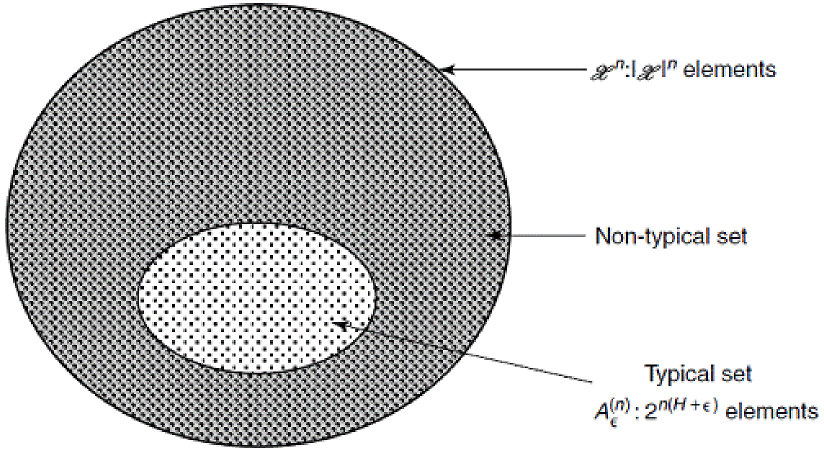


FIGURE 3.1. Typical sets and source coding.

- 分析
  - 为什么会存在非典型集？或这非典型集中都是什么序列。

答：由大数定理可知，随机变量X的取值x1出现的次数近似为np(x1), p(x1)!=1,则（x1，x1，……，x1）序列不是由p(X)产生的n个随机序列，所以（x1，x1，……，x1）不属于典型集。

所以那些不是由p(X)生成的序列不属于典型集。

作业5

4.1 双随机矩阵

- 双随机矩阵：任意行和列的和都为1，置换矩阵：n*n的双随机矩阵，且每行每列只含有一个值为1.

a) $H(\mathbf{b})-H(\mathbf{a})=-\sum_i\sum_j a_j p_{ji}\log\frac{b_i}{a_j}\geq-\log\sum_{i,j}\frac{b_i}{a_j}a_j p_{ji}=-\log 1=0$

主要用到的是函数的log函数的凹性。

另外，根据题目的提示，任意双随机矩阵均可以表示为

置换矩阵的凸组合，即$\mathbf{b}=\mathbf{a}(\sum\lambda_i\mathbf{P_i})=\sum\lambda_i\mathbf{aP}_i$，$\sum\lambda_i=1$,

熵是凹函数，则$\mathrm{H}(\mathbf{b})\geq\sum\lambda_i\mathrm{H}(\mathbf{aP}_i)=\mathrm{H}(\mathbf{a})$, 也能得到结果

b)假设$\mu_\mathrm{i}=\frac{1}{\mathrm{m}}$，很容易证明其满足$\mu=\mu\mathrm{P}$

c) 分布为均匀分布：$\mu_\mathrm{i}=\frac{1}{\mathrm{m}}=\sum_\mathrm{j}\mu_\mathrm{j}\mathrm{P_{ji}}=\frac{1}{\mathrm{m}}\sum_\mathrm{j}\mathrm{P_{ji}}$

可得到$\sum_\mathrm{j}\mathrm{P_{ji}}=1$,　而转移矩阵满足$\sum_\mathrm{i}\mathrm{P_{ji}}=1$，即证P是双随机的。

4.7 马尔科夫链的熵率

4.7 *Entropy rates of Markov chains*

(a) Find the entropy rate of the two-state Markov chain with transition matrix

$$P = \begin{bmatrix} 1-p_{01} & p_{01} \\ p_{10} & 1-p_{10} \end{bmatrix}.$$

(b) What values of p_{01}, p_{10} maximize the entropy rate?

(c) Find the entropy rate of the two-state Markov chain with transition matrix

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

(d) Find the maximum value of the entropy rate of the Markov chain of part (c). We expect that the maximizing value of p should be less than $\frac{1}{2}$, since the 0 state permits more information to be generated than the 1 state.

(e) Let $N(t)$ be the number of allowable state sequences of length t for the Markov chain of part (c). Find $N(t)$ and calculate

$$H_0 = \lim_{t\to\infty} \frac{1}{t} \log N(t).$$

[*Hint*: Find a linear recurrence that expresses $N(t)$ in terms of $N(t-1)$ and $N(t-2)$. Why is H_0 an upper bound on the entropy rate of the Markov chain? Compare H_0 with the maximum entropy found in part (d).]

两状态的马尔科夫链的熵率：$H(\chi) = \mathrm{H}(X_2 \mid X_1)$

(a) 平稳分布：　$\mu=\mu\mathrm{P}$

$$\left(\frac{p_{10}}{p_{01}+p_{10}}, \frac{p_{01}}{p_{01}+p_{10}}\right)$$

熵率$=p_0 H(p_{01}) + p_1 H(p_{10})$

$(b) 1/2$

$(c) H(p)/(1+p)$

(d)求导，得$\mathrm{p}=(3-\sqrt{5})/2$

4.9初始状态

4.9 *Initial conditions.* Show, for a Markov chain, that

$$H(X_0|X_n) \geq H(X_0|X_{n-1}).$$

Thus, initial conditions X_0 become more difficult to recover as the future X_n unfolds.

由$X_0 \to X_{n-1} \to X_n$

$I(X_0; X_{n-1}) \geq I(X_0; X_n)$ 数据处理不等式

$\Rightarrow H(X_0) - H(X_0 | X_{n-1}) \geq H(X_0) - H(X_0 | X_n)$

$\Rightarrow H(X_0 | X_n) \geq H(X_0 | X_{n-1})$

作业6

5.3 Slackness of Kraft inequality

5.3 *Slackness in the Kraft inequality*. An instantaneous code has word lengths $l_1, l_2, \ldots, l_m$, which satisfy the strict inequality

$$\sum_{i=1}^{m} D^{-l_i} < 1.$$

The code alphabet is $\mathcal{D} = \{0, 1, 2, \ldots, D-1\}$. Show that there exist arbitrarily long sequences of code symbols in $\mathcal{D}^*$ which cannot be decoded into sequences of codewords.

$l_{\max} = \max(l_1, l_2, \ldots, l_m)$,则总共有$D^{l_{\max}}$个长度为$l_{\max}$序列，

在这些序列中， 有$D^{l_{\max}-l_i}$个以第i个编码作为前缀，

因此，以各个编码为前缀的序列总共有

$\sum_i D^{l_{\max}-l_i} = D^{l_{\max}} \sum_i D^{-l_i} < D^{l_{\max}}$,说明这些序列中还有些

不能译为码字序列

5.32 坏葡萄酒

a)按照概率从大到小的顺序试酒时，平均次数最短（这个场景和重复试验不同）

$$EX = 1\times\frac{8}{23}+2\times\frac{6}{23}+3\times\frac{4}{23}+4\times\frac{2}{23}+5\times\frac{2}{23}+5\times\frac{1}{23}=\frac{55}{23}$$

问：$P(X=2)=\frac{6}{23}$?

答：设x_i表示第i瓶酒是否是坏酒，x_i=1表示是坏酒。

$$P(X=2)=P(x_0=0)\times P(x_1=1|x_0=0)=\frac{16}{23}\times\frac{6}{16}=\frac{6}{23}$$

b) 首先应该品尝通过观察所判定的最有可能为假酒的第一瓶

c) 用构造Huffman树的方法来确定如何混合。

使用Huffman树的方法可以使用最小的品尝次数（最小的码长）对有可能为坏酒的瓶子进行品尝

5.32 坏葡萄酒

c) 用构造Huffman树的方法来确定如何混合。

使用Huffman树的方法可以使用最小的品尝次数（最小的码长）对有可能为坏酒的瓶子进行品尝。

构造Huffman树得到的码字（11，10，00，010，0111，0110）

则期望尝试次数可表述为Huffman树的期望码长：

$$
\begin{aligned}
\sum_{i=1}^{6} p_i l_i &= 2\times\frac{8}{23}+2\times\frac{6}{23}+2\times\frac{4}{23}+3\times\frac{2}{23}+4\times\frac{2}{23}+4\times\frac{1}{23} \\
&= \frac{54}{23} \\
&= 2.35
\end{aligned}
$$

d) 该首先品尝那种混合形式。

由Huffman树构造过程可知，首先对坏酒概率最大的第一瓶喝第二瓶进行混合

5.37 码

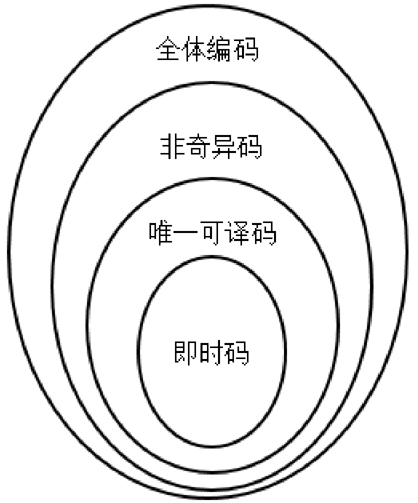
如下码：

$$\begin{aligned} C_1 &= \{00, 01, 0\} \\ C_2 &= \{00, 01, 100, 101, 11\} \\ C_3 &= \{0, 10, 110, 1110, \ldots\} \\ C_4 &= \{0, 00, 000, 0000\} \end{aligned}$$

唯一可译码：C2，C3

即时码：C2，C3

注：作业批改错误，将C1看作唯一可译码。



知识点：

非奇异（nonsingular）编码：

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

扩展（extension）编码：

$$C(x_1 x_2 \cdots x_n) = C(x_1)C(x_2)\cdots C(x_n)$$

唯一可译（uniquely decodable）编码：扩展编码是非奇异的

前缀码（prefix code）或即时码（instantaneous code）：码中无任何码字是其他码字的前缀。

唯一可译码的任一编码字符串只能源自唯一可能的信源字符串。

例子：唯一可译码但不是即时码

{10，00，11，110} 判断：11000 =>需要通观整个编码字符串

作业7

5.4 赫夫曼码

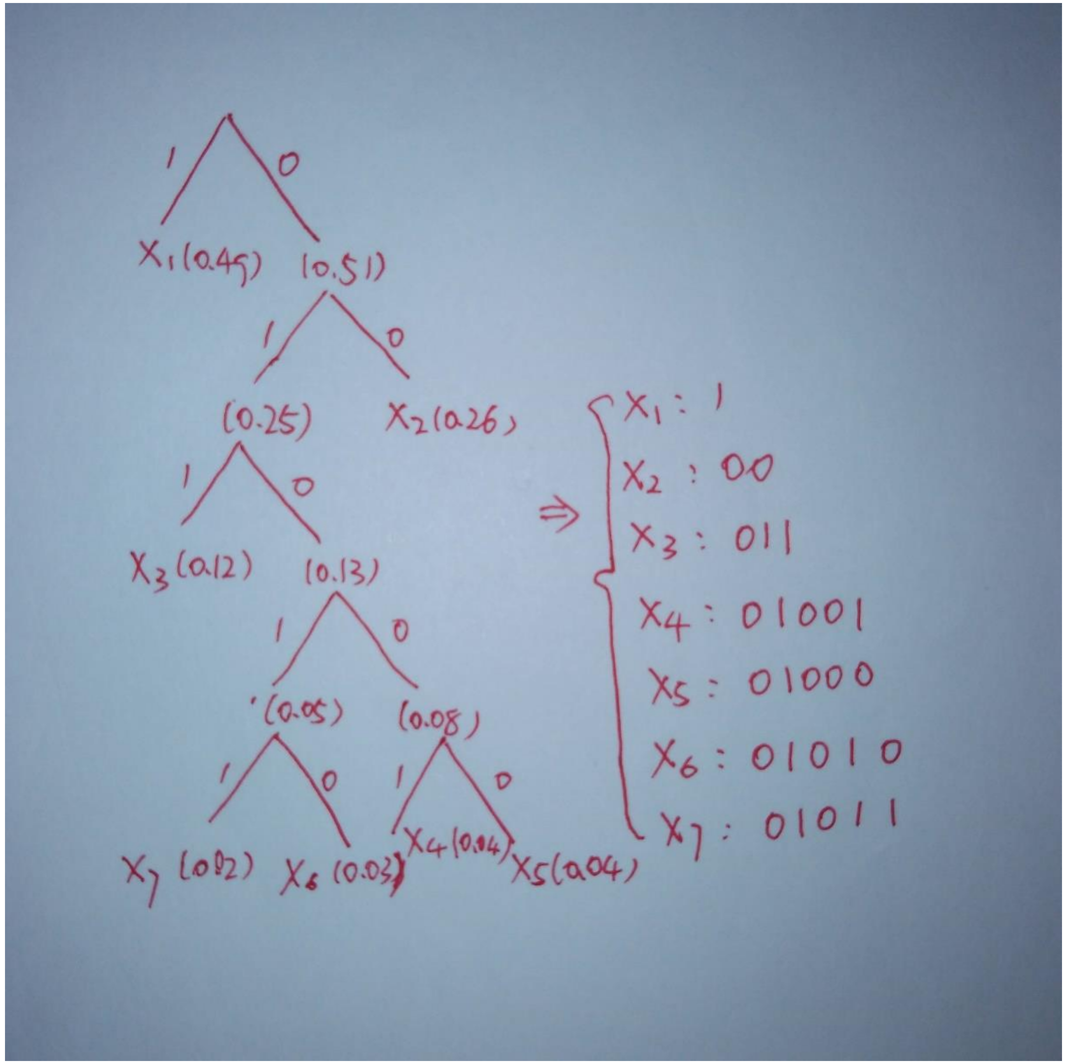
- a) 求X的二元赫夫曼码

Codeword								
1	x_1	0.49	0.49	0.49	0.49	0.49	0.51	1
00	x_2	0.26	0.26	0.26	0.26	0.26	0.49	
011	x_3	0.12	0.12	0.12	0.13	0.25		
01000	x_4	0.04	0.05	0.08	0.12			
01001	x_5	0.04	0.04	0.05				
01010	x_6	0.03	0.04					
01011	x_7	0.02						

构造二叉码树：

1：统一左子树的根节点的值小于（或大于）右子树的根节点的值

2：随机变量的值都在叶子节点上。



5.4 赫夫曼码

- b) 求改编码的期望码长
  - 期望码长：

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

  - 容易求得：L=2.02 bits，这里H(X)=2.01 bits
- c) 三元赫夫曼码

Codeword					
0	x_1	0.49	0.49	0.49	1.0
1	x_2	0.26	0.26	0.26	
20	x_3	0.12	0.12	0.25	
22	x_4	0.04	0.09		
210	x_5	0.04	0.04		
211	x_6	0.03			
212	x_7	0.02			

期望码长：L=1.34 铁特， H(X)=1.27 铁特

5.5 一码多用对的赫夫曼码

- 某信源的概率分布(1/3,1/5,1/5,2/15,2/15)，求其二元赫夫曼码。并讨论所得到的码对概率分布为(1/5,1/5,1/5,1/5,1/5)的信源也是最优的。
- 解：与5.4求解类似，可得(1/3,1/5,1/5,2/15,2/15)的 码字为{00，10，11，010，011}.

  要证明码字对(1/5,1/5,1/5,1/5,1/5)也是最优码，则要证明该码字具有最小的期望长度，并且该码字要满足条件：H(x)<=E(L)

  H(X)=log5=2.32 bits, 而该码字的期望长度为：E(L)=12/5=2.4 bits.

  而对于任意的编码的期望长度为：$E(L(\text{any code})) = \sum_{i=1}^{5} \frac{l_i}{5} = \frac{k}{5}$ bits

  这里k只能去整数值，而如果k<12,则不满足条件H(x)<=E(L)。即证。

5.16 赫夫曼码

- 考虑随机变量X，取6个值{A, B, C, D, E, F}，其概率以此是0.5, 0.25, 0.1, 0.05, 0.05, 0.05.

a) 构造二元赫夫曼码、期望长度？

Code	Source symbol	Prob.					
0	A	0.5	0.5	0.5	0.5	0.5	1.0
10	B	0.25	0.25	0.25	0.25	0.5	
1100	C	0.1	0.1	0.15	0.25		
1101	D	0.05	0.1	0.1			
1110	E	0.05	0.05				
1111	F	0.05					

期望长度：1 × 0.5+2× 0.25+4× (0.1+0.05+0.05+0.05) = 2 bits ， H(X)=1.98 bits

5.16 赫夫曼码

- b)构造四元赫夫曼码

解：

由于 每次简化过程中，字符数均减少D-1个，则要求字符的总数为（1+k(D-1)），这里D=4，而字符总数为7不满足，需要添加一个虚拟符。

期望长度：E(L)=1.15

Code	Symbol	Prob.		
a	A	0.5	0.5	1.0
b	B	0.25	0.25	
d	C	0.1	0.15	
ca	D	0.05	0.1	
cb	E	0.05		
cc	F	0.05		
cd	G	0.0		

5.16 赫夫曼码

- c)构造二元赫夫曼码(由四元进行映射)，求二元码的平均长度。

解：映射：a->00,b->01,c->10,d->11.

则结合(b)的四元赫夫曼码的结果映射得二元码：

A->00;

B->01;

C->11;

D->1000;

E->1001;

F->1010;

期望长度：E(L)=2*0.85+4*0.15=2.3 bits

5.16 赫夫曼码

- d)证明由四元映射所得的二元赫夫曼码平均长度 L_{QB} 满足

$$L_H \le L_{QB} < L_H + 2$$

- 解：

显然，$L_H \le L_{QB}$.

又，$H(X) \le L_H \le H(X)+1$,

$H_4(X) \le L_{H_4} \le H_4(X)+1$.

其中$H_4(X) = 1/2H(X)$.

所以$1/2H(X) \le L_{H_4} \le 1/2H(X)+1$.

即$H(X) \le 2L_{H_4} \le H(X)+2 \le L_H +2$.

又$L_{QB} = 2L_{H_4}$.

即证。

5.16 赫夫曼码

- e)该例子的下界是紧致的。举例说明最优四元赫夫曼码变换而来的编码也是最优二元码。
- 例子：变量X有四个可能值，并且每个可能值得概率均相同。则四元赫夫曼码对每个字符均采用一个四元字符，平均长度为1 quaternary symbol。 而四元映射到二元的平均长度 L_{QB} 为2 bits。使用二元赫夫曼码对每个字符使用2bits。所以 $L_H = L_{QB}$。

5.16 赫夫曼码

- f)证明较好的上界为 $L_{QB} \leq L_H + 1$
- 证明：

$L_4 = L_{QB} / 2.$

逆向映射：由二元映射到四元

对任意的四元编码$L_4^{'} \geq L_4.$

考虑一个二元编码变成一个四元编码，

只需要在所有非偶数的编码后添加0或1，

这样，$L_2 / 2 \leq L_4^{'} \leq (L_2 + 1) / 2.$

下界已经紧致，我们来看上界：

$L_4 \leq L_4^{'} \leq (L_2 + 1)/2.$

$L_4 = L_{QB} / 2 \leq (L_2 + 1)/2.$ 即证.