第3章 渐近均分性

大数定律：

独立同分布(i.i.d)随机变量，当n很大时，$\frac{1}{n}\sum_{i=1}^{n} X_i$ 近似于期望值EX。

渐近均分性：

$\frac{1}{n}\log\frac{1}{p(X_1,X_2,\cdots,X_n)}$ 近似于熵H，其中$X_1,X_2,\cdots,X_n$为独立同分布(i.i.d)随机变量，

$p(X_1,X_2,\cdots,X_n)$是观察序列$X_1,X_2,\cdots,X_n$出现的概率。

当n很大时，一个观察序列出现的概率$p(X_1,X_2,\cdots,X_n)$近似等于2^{-nH}

$$\frac{1}{n}\log\frac{1}{p(X_1,X_2,\cdots,X_n)}=H \Rightarrow p(X_1,X_2,\cdots,X_n)=2^{-nH}$$

典型集： 样本熵近似于真实熵

非典型集： 包含其余的序列.

"几乎一切事件都令人同等的意外"

当$X_1,X_2,\cdots,X_n$为i.i.d. $\sim p(x)$，则

$$Pr\{(X_1,X_2,\cdots,X_n): p(X_1,X_2,\cdots,X_n)=2^{-n(H\pm\varepsilon)}\}\approx 1$$

定义：随机变量的收敛.

给定一个随机变量序列$X_1,X_2,\cdots$. 序列$X_1,X_2,\cdots$收敛于随机变量X有如下三种情形：

(1) 如果对任意的$\varepsilon>0$，$Pr\{|X_n-X|>\varepsilon\}\to 0$，则称为依概率收敛

(2) 如果$E(X_n-X)^2\to 0$，则称为均方收敛.

(3) 如果$Pr\{\lim_{n\to\infty} X_n = X\}=1$，则称为以概率1（或称几乎处处）收敛。

3.1 渐近均分性定理.

定理3.1.1(AEP) 若$X_1,X_2,\cdots,X_n$为i.i.d $\sim p(x)$，则

$$-\frac{1}{n}\log p(X_1,X_2,\cdots,X_n)\to H(X) \text{ 依概率}$$

即对$\forall\varepsilon>0$，$Pr\{|-\frac{1}{n}\log p(X_1,X_2,\cdots,X_n)-H(X)|>\varepsilon\}\to 0$

证明：独立随机变量的函数依然是独立随机变量.

$$-\frac{1}{n}\log p(X_1,X_2,\cdots,X_n)=-\frac{1}{n}\sum_{i=1}^{n}\log p(X_i)$$

由弱大数定律

$$\to -E\log p(X) \text{ 依概率}$$

$$= H(X)$$

定义：关于$p(x)$的典型集$A_\varepsilon^{(n)}$(typical set)是序列$(x_1,x_2,\cdots,x_n)\in\mathcal{X}^n$的集合，且满足性质：

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

定理3.1.2

1. 如果$(x_1,x_2,\cdots,x_n)\in A_\varepsilon^{(n)}$，则$H(X)-\varepsilon\le-\frac{1}{n}\log p(x_1,x_2,\cdots,x_n)\le H(X)+\varepsilon$
2. 当n充分大时，$Pr\{A_\varepsilon^{(n)}\}>1-\varepsilon$
3. $|A_\varepsilon^{(n)}|\le 2^{n(H(X)+\varepsilon)}$，其中$|A|$表示集合A中的元素个数.
4. 当n充分大时，$|A_\varepsilon^{(n)}|\ge(1-\varepsilon)\,2^{n(H(X)-\varepsilon)}$

由此可知，典型集的概率近似为1，典型集中的所有元素几乎是等可能的，且典型集的元素个数近似等于2^{nH}.

?

证明：

(1). 由定义知 $2^{-n(H(X)+\varepsilon)}\le p(x_1,x_2,\cdots,x_n)\le 2^{-n(H(X)-\varepsilon)}$

$$\therefore H(X)-\varepsilon\le-\frac{1}{n}\log p(x_1,x_2,\cdots,x_n)\le H(X)+\varepsilon.$$

(2) 由于当$n\to\infty$时，事件$(X_1,X_2,\cdots,X_n)\in A_\varepsilon^{(n)}$的概率趋于1

$\therefore\forall\delta>0$，存在$n_0$，使$n\ge n_0$时，有

$$Pr\left\{\left|-\frac{1}{n}\log p(X_1,X_2,\cdots X_n)-H(X)\right|<\varepsilon\right\}>1-\delta$$

令$\delta=\varepsilon$.

(3)
$$\begin{aligned}1&=\sum_{\vec{x}\in\mathcal{X}}p(\vec{x})\\&\ge\sum_{\vec{x}\in A_\varepsilon^{(n)}}p(\vec{x})\\&\ge\sum_{\vec{x}\in A_\varepsilon^{(n)}}2^{-n(H(X)+\varepsilon)}\\&=2^{-n(H(X)+\varepsilon)}\cdot|A_\varepsilon^{(n)}|\end{aligned}$$

$$\therefore |A_\varepsilon^{(n)}|\le 2^{n(H(X)+\varepsilon)}$$

(4) $\because$ 当n充分大时，$Pr\{A_\varepsilon^{(n)}\}>1-\varepsilon$

$$\begin{aligned}1-\varepsilon&<Pr\{A_\varepsilon^{(n)}\}\\&\le\sum_{\vec{x}\in A_\varepsilon^{(n)}}2^{-n(H(X)-\varepsilon)}\\&=2^{-n(H(X)-\varepsilon)}|A_\varepsilon^{(n)}|\end{aligned}$$

$$\therefore |A_\varepsilon^{(n)}|\ge(1-\varepsilon)\cdot2^{n(H(X)-\varepsilon)}$$

要点总结：

1. AEP："几乎一切事件都令人同等的意外"。具体讲，若 $X_1, X_2, \cdots, X_n$ 为服从 $p(x)$ 的 i.i.d 序列，则

$$-\frac{1}{n}\log p(X_1, X_2, \cdots, X_n) \to H(X) \text{ 依概率}$$

定义：典型集 $A_\varepsilon^{(n)}$ 为满足如下条件的序列 $x_1, x_2, \cdots, x_n$ 所构成的集合

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

2. 典型集的性质：

1. 若 $(x_1, x_2, \cdots, x_n) \in A_\varepsilon^{(n)}$，则 $p(x_1, x_2, \cdots, x_n) = 2^{-n(H(X)\pm\varepsilon)}$
2. 当n充分大时，$Pr\{A_\varepsilon^{(n)}\} > 1-\varepsilon$
3. $|A_\varepsilon^{(n)}| \le 2^{n(H(X)+\varepsilon)}$，$|A|$表示集合A的元素个数.

7.6. 联合典型序列：

定义：服从分布 $p(x,y)$ 的联合典型序列 $\{(x^n, y^n)\}$ 所构成的集合 $A_\varepsilon^{(n)}$ 是指其经验熵与真实熵 ε 接近的n长序列构成的集合，即：

$$A_\varepsilon^{(n)} = \{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n :$$

$$\left|-\frac{1}{n}\log p(x^n) - H(X)\right| < \varepsilon \quad (\text{等同于 } 2^{-n(H(X)+\varepsilon)} \le p(x_1, x_2, \cdots, x_n) \le 2^{-n(H(X)-\varepsilon)})$$

$$\left|-\frac{1}{n}\log p(y^n) - H(Y)\right| < \varepsilon$$

$$\left|-\frac{1}{n}\log p(x^n, y^n) - H(X,Y)\right| < \varepsilon\}$$

其中 $p(x^n, y^n) = \prod_{i=1}^n p(x_i, y_i)$

定理 7.6.1（联合AEP） 设 (X^n, Y^n) 为服从 $p(x^n, y^n) = \prod_{i=1}^n p(x_i, y_i)$ 的 i.i.d 的n长序列，那么：

1. 当 $n \to \infty$ 时，$Pr\{(X^n, Y^n) \in A_\varepsilon^{(n)}\} \to 1$
2. $|A_\varepsilon^{(n)}| \le 2^{n(H(X,Y)+\varepsilon)}$
3. 如果 $(\tilde{X}^n, \tilde{Y}^n) \sim p(x^n)p(y^n)$，即 $\tilde{X}^n$ 与 $\tilde{Y}^n$ 是独立的且与 $p(x^n, y^n)$ 有相同的边际分布，那么：

$$Pr\{(\tilde{X}^n, \tilde{Y}^n) \in A_\varepsilon^{(n)}\} \le 2^{-n(I(X;Y)-3\varepsilon)}$$

而且，对于充分大的n，

$$Pr\{(\tilde{X}^n, \tilde{Y}^n) \in A_\varepsilon^{(n)}\} \ge (1-\varepsilon) \cdot 2^{-n(I(X;Y)+3\varepsilon)}$$

证明: 设 (X^n, Y^n) 为服从 $p(x^n, y^n) = \prod_{i=1}^{n} p(x_i, y_i)$ 的 i.i.d 的 n长序列

1. 当 $n \to \infty$ 时, $Pr\{(X_n, Y_n) \in A_\varepsilon^{(n)}\} \to 1$

首先证明,包含在典型集中的序列具有很高的概率.

由弱大数定律: $-\frac{1}{n}\log p(X^n) \to -E\log p(X) = H(X)$ 依概率

$\therefore$ 给定 $\varepsilon > 0$, $\exists n_1$, 使得对 $\forall n, n \geq n_1$,

$$Pr\{|-\frac{1}{n}\log p(X^n) - H(X)| \geq \varepsilon\} < \frac{\varepsilon}{3}$$

同理: $\varepsilon > 0$, $\exists n_2$, 使 $\forall n \geq n_2$

$$Pr\{|-\frac{1}{n}\log p(Y^n) - H(Y)| \geq \varepsilon\} < \frac{\varepsilon}{3}$$

$\varepsilon > 0$, $\exists n_3$, 使 $\forall n \geq n_3$

$$Pr\{|-\frac{1}{n}\log p(X^n, Y^n) - H(X,Y)| \geq \varepsilon\} < \frac{\varepsilon}{3}$$

选取 $n > \max(n_1, n_2, n_3)$ 则 三集合之并的概率必小于 ε

$\therefore$ 对于充分大的 n, 集合 $A_\varepsilon^{(n)}$ 的概率大于 $1-\varepsilon$

2. $|A_\varepsilon^{(n)}| \leq 2^{n(H(X,Y)+\varepsilon)}$

$$1 = \sum p(x^n, y^n)$$

$$\geq \sum_{A_\varepsilon^{(n)}} p(x^n, y^n)$$

$$\geq |A_\varepsilon^{(n)}| \cdot 2^{-n(H(X,Y)+\varepsilon)}$$

$$\therefore |A_\varepsilon^{(n)}| \leq 2^{n(H(X,Y)+\varepsilon)}$$

3. 如果 $(\tilde{X}^n, \tilde{Y}^n) \sim p(x^n)p(y^n)$, 即 $\tilde{X}^n$ 与 $\tilde{Y}^n$ 是独立的,且与 $p(x^n, y^n)$ 有相同的边际分布,那

$$Pr\{(\tilde{X}^n, \tilde{Y}^n) \in A_\varepsilon^{(n)}\} \leq 2^{-n(I(X;Y)-3\varepsilon)}$$

且对于充分大的 n, $Pr\{(\tilde{X}^n, \tilde{Y}^n) \in A_\varepsilon^{(n)}\} \geq (1-\varepsilon)2^{-n(I(X;Y)+3\varepsilon)}$

$$Pr\{(\tilde{X}^n, \tilde{Y}^n) \in A_\varepsilon^{(n)}\} = \sum_{(x^n, y^n) \in A_\varepsilon^{(n)}} p(x^n)p(y^n)$$

$$\leq |A_\varepsilon^{(n)}| \, 2^{-n(H(X)-\varepsilon)} \cdot 2^{-n(H(Y)-\varepsilon)}$$

$$= 2^{n(H(X,Y)+\varepsilon)} \cdot 2^{-n(H(X)-\varepsilon)} \cdot 2^{-n(H(Y)-\varepsilon)}$$

$$= 2^{-n(I(X;Y)-3\varepsilon)}$$

上方↑

对于充分大的 n, $Pr(A_\varepsilon^{(n)}) \geq 1-\varepsilon$,

$$1-\varepsilon \leq \sum_{(x^n, y^n) \in A_\varepsilon^{(n)}} p(x^n, y^n)$$

$$\leq |A_\varepsilon^{(n)}| \cdot 2^{-n(H(X,Y)-\varepsilon)}$$

$$\therefore |A_\varepsilon^{(n)}| \geq (1-\varepsilon) \cdot 2^{n(H(X,Y)-\varepsilon)}$$

$$Pr\{(\tilde{X}^n, \tilde{Y}^n) \in A_\varepsilon^{(n)}\}$$

$$= \sum_{A_\varepsilon^{(n)}} p(x^n)p(y^n)$$

$$\geq |A_\varepsilon^{(n)}| \, 2^{-n(H(X)+\varepsilon)} \cdot 2^{-n(H(Y)+\varepsilon)}$$

$$\geq (1-\varepsilon) \cdot 2^{n(H(X,Y)-\varepsilon)} \cdot 2^{-n(H(X)+\varepsilon)} \cdot 2^{-n(H(Y)+\varepsilon)}$$

$$= (1-\varepsilon) 2^{-n(I(X;Y)+3\varepsilon)}$$

3.1 马尔可夫不等式与切比雪夫不等式

(a) $\boxed{\text{马尔可夫不等式}}$ 对任意非负随机变量 X 以及任意的 $t>0$，证明

$$Pr\{X \geq t\} \leq \frac{EX}{t}$$

$$EX = \sum_{x \in \mathcal{X}} x \cdot P(X=x)$$

$$\geq \sum_{x \geq t} x \cdot P(X=x)$$

$$\geq t \sum_{x \geq t} P(X=x)$$

$$= t \cdot Pr\{X \geq t\}$$

$\therefore Pr\{X \geq t\} \leq \frac{EX}{t}$

(b) $\boxed{\text{切比雪夫不等式}}$ 设随机变量 Y 的均值与方差分别为 u 和 σ^2。设 $X=(Y-u)^2$

证明对 $\forall \varepsilon > 0$，

$$Pr\{|Y-u| > \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}$$

$$Pr\{|Y-u|^2 > \varepsilon^2\} \leq \frac{E|Y-u|^2}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2}$$

$$E|Y-u|^2 = E[Y^2] + u^2 - 2uE[Y] = \sigma^2 + u^2 - u^2 = \sigma^2$$

又 $\because Pr\{|Y-u| > \varepsilon\} = Pr\{|Y-u|^2 > \varepsilon^2\}$

$\therefore Pr\{|Y-u| > \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}$

(c) $\boxed{\text{弱大数定律}}$ 设 $Z_1, Z_2, \cdots, Z_n$ 为 i.i.d 随机变量序列，其均值和方差分别为 u 和 σ^2

令 $\overline{Z}_n = \frac{1}{n}\sum_{i=1}^{n} Z_i$ 为样本均值。证明：

$$Pr\{|\overline{Z}_n - u| > \varepsilon\} \leq \frac{\sigma^2}{n\varepsilon^2}$$

$$Pr\{|\overline{Z}_n - u| > \varepsilon\}$$

$$= Pr\{|\overline{Z}_n - u|^2 > \varepsilon^2\}$$

$$\leq \frac{1}{\varepsilon^2} \cdot E|\overline{Z}_n - u|^2$$

$$= \frac{1}{\varepsilon^2}\left(E\overline{Z}_n^2 + u^2 - 2uE[\overline{Z}_n]\right)$$

$$= \frac{1}{\varepsilon^2}\left(\frac{1}{n^2}E\left(\sum_{i=1}^{n} Z_i^2 + \sum_{i \neq j} Z_i Z_j\right) + u^2 - 2u \cdot \frac{1}{n} \cdot nu\right)$$

$$= \frac{1}{\varepsilon^2}\left(\frac{1}{n^2}\left(n \cdot (\sigma^2 + u^2) + n(n-1)u^2\right) - u^2\right)$$

$$= \frac{1}{\varepsilon^2}\frac{1}{n}\sigma^2$$

$\therefore Pr\{|\overline{Z}_n - u| > \varepsilon\} \leq \frac{\sigma^2}{n\varepsilon^2}$

因此，当 $n \to \infty$ 时，$Pr\{|\overline{Z}_n - u| > \varepsilon\} \to 0$.

即弱大数定律.

3.2 AEP与互信息.

设(X_i, Y_i)为i.i.d $\sim p(x,y)$. 假设X和Y独立, 假设X和Y相关的对数似然比.

求 $\frac{1}{n}\log\frac{p(X^n)p(Y^n)}{p(X^n,Y^n)}$ 的极限.

①: $\frac{1}{n}\log\frac{p(X^n)p(Y^n)}{p(X^n,Y^n)}$

$= \frac{1}{n}\log p(X^n) + \frac{1}{n}\log p(Y^n) - \frac{1}{n}\log p(X^n,Y^n)$

$= \frac{1}{n}\sum_{i=1}^{n}\log p(X_i) + \frac{1}{n}\sum_{i=1}^{n}\log p(Y_i) - \frac{1}{n}\sum_{i=1}^{n}\log p(X_i,Y_i)$

$\to E\log p(X) + E\log p(Y) - E\log p(X,Y)$

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

$= -I(X;Y)$

②. $\frac{1}{n}\log\frac{p(X^n)p(Y^n)}{p(X^n,Y^n)}$

$= \frac{1}{n}\sum_{i=1}^{n}\log\frac{p(X_i)p(Y_i)}{p(X_i,Y_i)}$

$\to E\log\frac{p(X)p(Y)}{p(X,Y)}$

$= -I(X;Y)$

3.4 AEP.

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

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

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

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

(a) 由AEP可知.

(b) ~~若$x^n\in A^n\cap B^n$, 则$x^n\in A^n$~~

大数定律: $Pr(X^n\in B^n)\to 1$

$\therefore \exists\varepsilon>0$, 当$n>N_1$时, $Pr(X^n\in A^n)>1-\frac{\varepsilon}{2}$

当$n>N_2$时, $Pr(X^n\in B^n)>1-\frac{\varepsilon}{2}$

$\therefore$ 当$n>\max(N_1,N_2)$时

$Pr\{X^n\in A^n\cap B^n\}=Pr\{X^n\in A^n\}+Pr\{X^n\in B^n\}-Pr\{X^n\in A^n\cup B^n\}$

$>(1-\frac{\varepsilon}{2})+(1-\frac{\varepsilon}{2})-1$

$=1-\varepsilon$

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

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

$1=\sum_{x}p(x^n)$

$\ge\sum_{A^n}p(x^n)$

$\ge|A^n|\cdot 2^{-n(H+\varepsilon)}$

$\therefore |A^n|\le 2^{n(H+\varepsilon)}$

(4). $\because Pr\{X^n\in A^n\cap B^n\}\to 1$

$\therefore\exists N$, 当$n>N$时 $Pr\{X^n\in A^n\cap B^n\}\ge\frac{1}{2}$

$\frac{1}{2}\le\sum_{A^n\cap B^n}p(x^n)\le|A^n\cap B^n|\cdot 2^{-n(H-\varepsilon)}$

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

3.5 由概率定义的集合. 设 $X_1, X_2, \cdots$ 为 i.i.d 离散随机变量序列, 熵为 $H(X)$.

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

表示概率 $\geq 2^{-nt}$ 的所有 n 长序列构成的子集。

(a) 证明: $|C_n(t)| \leq 2^{nt}$

(b) 当 t 为何值时, 有 $P(\{X^n \in C_n(t)\}) \to 1$? ?

(a)
$$1 = \sum_{x^n \in \mathcal{X}^n} p(x^n)$$
$$\geq \sum_{x^n \in C_n(t)} p(x^n)$$
$$\geq |C_n(t)| \cdot 2^{-nt}$$

$\therefore |C_n(t)| \leq 2^{nt}$

(b)

3.6. 类似于AEP的极限. 设 $X_1, X_2, \cdots, X_n$ 为 i.i.d 序列且服从概率密度函数 $p(x)$. 试求:

$$\lim_{n\to\infty} (p(X_1, X_2, \cdots, X_n))^{\frac{1}{n}}$$

$$\lim_{n\to\infty} \log (p(X_1, X_2, \cdots, X_n))^{\frac{1}{n}} = \lim_{n\to\infty} \frac{1}{n} \log p(X_1, X_2, \cdots, X_n)$$
$$= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \log p(X_i)$$
$$\to E \log p(X)$$
$$= -H(X)$$

$$\therefore \lim_{n\to\infty} (p(X_1, X_2, \cdots, X_n))^{\frac{1}{n}} = 2^{-H(X)}$$

3.9 AEP 设$X_1, X_2 \cdots$为独立同分布随机变量序列，服从概率密度函数$p(x)$，

$x \in \{1,2,\cdots,m\}$，于是$p(x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n} p(x_i)$。已知$-\frac{1}{n}\log p(X_1,X_2,\cdots,X_n) \to H(X)$依概率

设$q(x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n} q(x_i)$，其中$q$是$\{1,2,\cdots,m\}$上的另一个概率密度函数。

(a) 计算$\lim_{n\to\infty} -\frac{1}{n}\log q(X_1,X_2,\cdots,X_n)$，其中$X_1,X_2,\cdots$为服从$p(x)$的i.i.d序列

(b) 计算对数似然比$\frac{1}{n}\log\frac{q(X_1,X_2,\cdots,X_n)}{p(X_1,X_2,\cdots,X_n)}$的极限，其中$X_1,X_2,\cdots$为服从$p(x)$的i.i.d序列。由此说明当$p$为真实分布时，偏好分布$q$的优势将以指数衰减。

(a) $\lim_{n\to\infty} -\frac{1}{n}\log q(X_1,X_2,\cdots,X_n)$

$= \lim_{n\to\infty} -\frac{1}{n}\sum_{i=1}^{n}\log q(X_i)$

$\to -E\log q(X)$

$= \cancel{H_q(X)}\ -\sum p(x)\cdot\log q(x) = -\sum p(x)\cdot\log\frac{q(x)}{p(x)}\cdot p(x) = \sum p(x)\cdot\log\frac{p(x)}{q(x)} - \sum p(x)\cdot\log p(x)$

(b) $\lim_{n\to\infty}\frac{1}{n}\log\frac{q(X_1,X_2,\cdots,X_n)}{p(X_1,X_2,\cdots,X_n)}$

$= \cancel{H_p(X) - H_q(X)}$

$= \lim \frac{1}{n}\sum\log\frac{q(X_i)}{p(X_i)}$

$\to E\log\frac{q(X_i)}{p(X_i)}$

$= \sum \cancel{p(X_i)}\ p(x)\cdot\log\frac{q(x)}{p(x)}$

3.11.

(a) 给定任意两个集合A和B，使得$Pr(A)>1-\varepsilon_1$，$Pr(B)>1-\varepsilon_2$，证明：

$$Pr(A\cap B)>1-\varepsilon_1-\varepsilon_2$$

$Pr(A\cap B) = Pr(A)+Pr(B)-\underbrace{Pr(A\cup B)}_{<1}$

$> (1-\varepsilon_1)+(1-\varepsilon_2)-1$

$= 1-\varepsilon_1-\varepsilon_2$