考研数学概率论与数理统计基础讲义

主讲：张宇

张宇：新东方在线名师，博士，全国著名考研数学辅导专家，教育部"国家精品课程建设骨干教师"，全国畅销书《高等数学 18 讲》、《考研数学题源探析经典 1000 题》作者，高等教育出版社《全国硕士研究生入学统一考试数学考试参考书（大纲解析）》编者之一，2007 年斯洛文尼亚全球可持续发展大会受邀专家（发表 15 分钟主旨演讲）。首创"题源教学法"，对考研数学的知识结构和体系有全新的解读，对考研数学的命题与复习思路有极强的把握和预测能力，让学生轻松高效夺取高分。

目 录

第一讲　如何处理复杂事件?

1. 预备知识

（1）随机试验

称一个试验为随机试验，如果

1)试验可以在相同的条件下重复进行；

2)试验所有可能结果是明确可知道的，并且不止一个；

3)每一次试验会出现哪一个结果事先不能确定.

【评注】①我们是通过研究随机试验来研究随机现象的，为方便起见，将随机试验简称为试验，并用字母E或$E_1,E_2,\cdots,E_n,\cdots$表示.

②在一次试验中可能出现，也可能不出现的结果称为随机事件，简称为事件，并用大写字母A，B，C等表示，为讨论需要，将每次试验一定发生的事件称为必然事件，记为Ω. 每次试验一定不发生的事件称为不可能事件，记为$\varnothing$.

③随机试验每一最简单、最基本的结果称为基本事件或样本点，记为ω. 每次试验能且只能发生一个基本事件. 基本事件(或样本点)的全体称为基本事件空间(或样本空间)，记为Ω，即$\Omega=\{\omega\}$，随机事件A总是由若干个基本事件组成，即A是Ω的子集，$A\subset\Omega$. 事件A发生等价于构成A的基本事件有一个发生.

④在不少情况下，我们不能确切知道某一随机试验的全部可能结果，但可以知道它不超出某个范围. 这时，也可以用这个范围来作为该试验的全部可能结果. 例如我们需要记录某个城市一天的交通事故数量，则试验结果将是非负数x. 我们无法确定x的可能取值的确切范围，但可以把这范围取为$[0,+\infty)$，它总能包含一切可能的试验结果，尽管我们明知，某些结果，如 x＞10000，是不会出现的，我们甚至可以把这范围取为$(-\infty,+\infty)$也无妨. 这里就有了一定的数学抽象，它可以带来很大的方便.

（2）事件的运算关系

事件的关系、运算与集合的关系、运算相当，且具有相同的运算法则：

吸收律：若$A\subset B$，则$A\cup B=B$，$AB=A$；

交换律：$A\cup B=B\cup A$，$AB=BA$；

结合律：$(A\cup B)\cup C=A\cup(B\cup C)$，$(AB)C=A(BC)$；

分配律：$A(B\cup C)=AB\cup AC$，$A\cup BC=(A\cup B)(A\cup C)$；$A(B-C)=AB-AC$；

对偶律：$\overline{A\bigcup B}=\overline{A}\bigcap\overline{B}$，$\overline{A\bigcap B}=\overline{A}\bigcup\overline{B}$.

2. 古典概型

定义：如果其基本事件空间(样本空间)满足(1)只有有限个基本事件(样本点)；(2)每个基本事件(样本点)发生的可能性都一样. 称随机试验(随机现象)的概率模型为古典概型. 如果古典概型的基本事件总数为n，事件A包含k个基本事件，即有利于A的基本事件k

个．则A的概率定义为

$$P(A)=\frac{k}{n}=\frac{\text{事件}A\text{所含基本事件的个数}}{\text{基本事件总数}}$$

由上式计算的概率称为A的古典概率．

【例】将n个球随意放入N $(n\le N)$个盒子中，每个盒子可放任意多球，求P{恰有n个盒子中各有一球}．

【注】假设有 12 个人回母校参加校庆，则P{12 个人生日全不相同}=__________．

3. 几何概型

引例

定义：如果(1)样本空间(基本事件空间)Ω是一个可度量的几何区域；(2)每个样本点(基本事件)发生的可能性都一样，即样本点落入Ω的某一可度量的子区域A的可能性大小与A的几何度量成正比，而与A的位置及形状无关．称随机试验(随机现象)的概率模型为几何概型.

在几何概型随机试验中，如果S_A是样本空间Ω一个可度量的子区域，则事件$A=$“样本点落入区域S_A”的概率定义为

$$P(A)=\frac{S_A\text{的几何度量}}{\Omega\text{的几何度量}}$$

由上式计算的概率称为A的几何概率.

【例】甲、乙有约，上午 9:00——10:00 到校门口见面，等 20 分钟即离开. 求P{相遇}=__________.

4. 重要公式求概率

① $P(\overline{A})=1-P(A)$；

② $P(A-B)=P(A)-P(AB)$；

③ $P(A+B)=P(A)+P(B)-P(AB)$；

$P(A+B+C)=P(A)+P(B)+P(C)-P(AB)-P(BC)-P(AC)+P(ABC)$；

【注】当事件大于 3 个时，只考查

1） $A_1,A_2,\cdots A_n(n\geq 3)$ 两两互斥，则 $P(\bigcup\limits_{i=1}^{n}A_i)=\sum\limits_{i=1}^{n}P(A_i)$；

2） $A_1,A_2,\cdots A_n(n\geq 3)$ 相互独立，则

$$P(\bigcup_{i=1}^{n}A_i)=P(A_1\cup A_2\cup A_3\cup\cdots\cup A_n)=1-\prod_{i=1}^{n}[1-P(A_i)].$$

相互独立：设 $A_1,A_2,\cdots A_n$，若对其中任意有限个 $A_{i1},A_{i2},\cdots,A_{ik}(k\geq 2)$，都有

$P(A_{i1}A_{i2}\cdots A_{ik})=P(A_{i1})P(A_{i2})\cdots P(A_{ik})$，则称 $A_1,A_2,\cdots A_n$ 相互独立.

n 个事件相互独立 $\Leftrightarrow$ 它们中任意一部分事件换成各自的对立事件所得 n 个新事件相互独立

④ $P(B\mid A)=\dfrac{P(AB)}{P(A)}$ （$P(A)>0$）

⑤ $P(AB)=P(A)P(B|A)$； $P(ABC)=P(AB)P(C|AB)=P(A)P(B|A)P(C|AB)$；

⑥全概率公式（全集分解公式）

引例　假设一个村子里有三个小偷，求失窃的概率 P{失窃}.

定义：如果 $\bigcup\limits_{i-1}^{n}A_i=\Omega,A_iA_j=\phi(i\neq j),P(A_i)>0$，则对任一事件 B，有

$$B=\bigcup_{i-1}^{n}A_iB,P(B)=\sum_{i-1}^{n}P(A_i)P(B\mid A_i).$$

⑦贝叶斯公式

如果$\bigcup_{i=1}^{n}A_i=\Omega, A_iA_j=\phi(i\neq j), P(A_i)>0$，则对任一件事$B$，只要$P(B)>0$，有

$$P(A_i\mid B)=\frac{P(A_i)P(B\mid A_i)}{\sum_{i=1}^{n}P(A_i)(B\mid A_i)}(i=1,2,\cdots,n)$$

【例】有甲、乙两人射击，轮流独立对同一目标射击.P{甲命中}=α，P{乙命中}=β.甲先射击，谁先命中谁获胜.求甲、乙获胜的概率.

第二讲 如何求分布？

1. 基本概念

（1）随机变量

随机变量就是“其值随机会而定”的变量．设随机试验E的样本空间为$\Omega=\{\omega\}$，如果对每一个$\omega\in\Omega$，都有唯一的实数$X(\omega)$与之对应，并且对任意实数x，$\{\omega: X(\omega)\le x\}$是随机事件，则称定义在$\Omega$上的实单值函数$X(\omega)$为随机变量．简记为随机变量$X$．一般用大写字母$X$，$Y$，$Z\cdots$或希腊字母$\xi$，$\eta$，$\cdots$来表示随机变量．

（2）分布函数

设X是随机变量，x是任意实数，称函数$F(x)\triangleq P\{X\le x\}(x\in R)$为随机变量$X$的分布函数，或称$X$服从分布$F(x)$，记为$X\sim F(x)$．

（3）离散型随机变量

如果随机变量X只可能取有限个或可列个值$x_1,x_2,\cdots$，则称X为离散型随机变量，称$p_i=P\{X=x_i\}$，$i=1,2,\cdots$，为X的分布列、分布律或概率分布，记为$X\sim p_i$．概率分布常用表格形式或矩阵形式表示，即

X	x_1	x_2	$\cdots$
p	p_1	p_2	$\cdots$

或$X\sim\begin{pmatrix} x_1 & x_2 & \cdots \\ p_1 & p_2 & \cdots \end{pmatrix}$．

设离散型随机变量X的概率分布为$p_i=P\{X=x_i\}$，则X的分布函数

$$F(x)=P(X\le x)=\sum_{x_i\le x}P(X=x_i) \qquad (-\infty<x<+\infty)$$

（4）连续型随机变量

对于某一随便变量X，若存在非负可积函数$f(x)$，使得

$$F(x)=\int_{-\infty}^{x}f(t)dt，\ -\infty<x<+\infty$$

则称X为连续型随机变量.

2. 重要分布

（1）$0-1$分布（两点分布）

如果 X 的概率分布为 $X \sim \begin{pmatrix} 1 & 0 \\ P & 1-P \end{pmatrix}$ 即 $P\{X=1\}=p$，$P\{X=0\}=1-p$，则称 X 服从参数为 p 的 0-1 分布，记为 $X \sim B(1,p)(0<p<1)$.

（2）二项分布

【注】n 重伯努利试验：1）相互独立；2）每次试验出现 A 的概率相同；3）只有两个结果 A，$\overline{A}$. A 发生的次数记为 X，则 $P\{X=k\}=P\{X=k\}=C_n^k p^k (1-p)^{n-k}$，$k=0,1,\cdots,n$.

如果 X 的概率分布为 $p_k=P\{X=k\}=C_n^k p^k (1-p)^{n-k}$，$k=0,1,\cdots,n$，$0<p<1$，则称 X 服从参数为 (n,p) 的二项分布，记为 $X \sim B(n,p)$.

（3）几何分布（与几何无关!）

——“首中即停止”（等待型分布）

如果 X 的概率分布为 $p_k=P\{X=k\}=q^{k-1}p$，$k=1,2,\cdots,n$，$0<p<1$，$q=1-p$，则称 X 服从参数为 p 的几何分布.

（4）超几何分布

假如有 N 件产品，M 件正品，任取 n 件，取到 k 个正品的概率 p？

如果 X 的概率分布为 $p_k=P\{X=k\}=\dfrac{C_M^k C_{N-M}^{n-k}}{C_N^n}$，$k=0,1,\cdots$，$\min(M,n)$，$N$，$M$，$n$ 为正整数，则称 X 服从参数为 N，M，n 的超几何分布.

★（5）泊松分布

在某场合，某时间段内，源源不断地质点来流的个数 X 服从泊松分布.

如果 X 的概率分布为 $P\{X=k\}=\dfrac{\lambda^k}{k!}e^{-\lambda}$，$k=0,1,\cdots$，$\lambda>0$，则称 X 服从参数为 λ 的泊松分布.

（6）均匀分布（与几何概型联系）

若随机变量 X 在区间 I 上的任一子区间取值的概率与该子区间的长度成正比，则称 $X \sim U(I)$.也即 $X \sim f(x)=\begin{cases} \dfrac{1}{b-a}, & a<x<b, \\ 0, & \text{其他} \end{cases}$，记为 $X \sim U(a,b)$.

★（7）指数分布

若 $X \sim f(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0,\\ 0, & x\le 0\end{cases}$（$\lambda>0$），则称 $X \sim E(\lambda)$.

（8）正态分布

若 $X \sim f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$，$-\infty<x<+\infty$，则称 $X \sim N(\mu,\sigma^2)$.

3. 例题分析

【例】已知某设备由若干零部件组成，其中有 i 个 $(i=0,1,2)$ 非优质品零件的可能性相同. 当设备有 i 个非优质品零件时，其使用寿命服从参数 $\lambda=i+1$ 的指数分布，求设备使用寿命 X 的分布函数 $F(x)$ 与概率密度 $f(x)$.

4. 多维（二维）随机变量及其分布

——联合分布、边缘分布、条件分布、独立性

（1）二维随机变量

如果 X，Y 是定义在同一个样本空间 Ω 上的二个随机变量，则称二元总体 (X,Y) 为二维随机变量(或二维随机向量). 如果 $X_1,X_2,\cdots,X_n$ 是定义在同一样本空间 Ω 上的 n 个随机变量，则称 n 元总体 $(X_1,X_2,\cdots,X_n)$ 为 n 维随机变量或 n 维随机向量. X_i 称为第 i 个分量.

（2）联合分布函数

设 (X,Y) 为二维随机变量，对任意的实数 x,y，称二元函数

$$F(x,y)\triangleq P\{X\le x,Y\le y\}\ (x,y)\in R$$

为二维随机变量 (X,Y) 的联合分布函数，简称为分布函数，记为 $(X,Y)\sim F(x,y)$.

（3）离散型随机变量

如果二维随机变量(X,Y)只能取有限对值或可列对值(x_1,y_1)，(x_1,y_2)，…，(x_n,y_n)，…．则称(X,Y)为二维离散型随机变量．称

$$p_{ij} \triangleq P(X=x_i,Y=y_j)，\ i,j=1,2,\cdots$$

为(X,Y)的概率分布或联合分布，记为$(X,Y)\sim p_{ij}$．联合分布常用表格形式表示

$X \backslash Y$	y_1	$\cdots$	y_j	$\cdots$	$P(X=x_i)$
x_1	p_{11}	$\cdots$	p_{1j}	$\cdots$	$p_{1\cdot}$
$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$
x_i	p_{i1}	$\cdots$	p_{ij}	$\cdots$	$p_{i\cdot}$
$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$
$P(Y=y_i)$	$p_{\cdot 1}$	$\cdots$	$p_{\cdot j}$	$\cdots$	1

条件分布：$P(X=x_i|Y=y_j)=\dfrac{P(X=x_i,Y=y_j)}{P(Y=y_j)}=\dfrac{p_{ij}}{p_{.j}}$（$X$在$Y=y_j$条件下的条件分布）

$$P(Y=y_j|X=x_i)=\frac{P(X=x_i,Y=y_j)}{P(X=x_i)}=\frac{p_{ij}}{p_{i.}}$$（Y在$X=x_i$条件下的条件分布）

独立性：$p_{i.}\cdot p_{.j}=p_{ij}(i,j=1,2,\cdots)\Leftrightarrow X,Y$独立.

（4）连续型随机变量

对于二维随机变量(X,Y)，如果存在非负可积函数$f(x,y)$，使得

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(u,v)dudv\ \ (x,y)\in R$$

则称(X,Y)为二维连续型随机变量.

边缘概率密度：$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy$　　(X,Y)关于X的概率密度

$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)dx$　　(X,Y)关于Y的概率密度

条件概率密度：$f_{X|Y}(x|y)=\dfrac{f(x,y)}{f_Y(y)}(f_Y(y)>0)$　（X在$Y=y$条件下的条件密度）

$f_{Y|X}(y|x)=\dfrac{f(x,y)}{f_X(x)}(f_X(x)>0)$　（Y在$X=x$条件下的条件密度）

独立性：$f_X(x)f_Y(y)=f(x,y)\Leftrightarrow X,Y$独立.

【例】设$X\sim U(0,1)$，在$X=x\ (0<x<1)$的条件下，Y在$(0,x)$内服从均匀分布.

求：（Ⅰ） $f(x,y)$；（Ⅱ） $f_Y(y)$.

【总结】求谁不积谁，不积先定限，限内画直线，先交写下限，后交写上限.

第三讲 如何求数字特征？

1. 数学期望

（1）离散型随机变量

若 $X \sim \begin{pmatrix} x_1 & x_2 & \cdots & x_n & \cdots \\ p_1 & p_2 & \cdots & p_n & \cdots \end{pmatrix}$，则 $EX = \sum_{i=1}^{\infty} x_i p_i$.

（2）连续型随机变量

若 $X \sim f(x)$，则 $EX = \int_{-\infty}^{+\infty} xf(x)dx$.

【注】若 $Y = g(X)$，则 $EY = \begin{cases} \sum_{i=1}^{\infty} g(x_i)p_i & \text{离散型} \\ \int_{-\infty}^{+\infty} g(x)f(x)dx, & \text{连续型} \end{cases}$.

2.方差 DX

设 X 是随机变量，如果 $E(X-EX)^2$ 存在，则称 $E(X-EX)^2$ 为 X 的方差，记为 DX .

即 $DX = E(X-EX)^2$.

从定义角度，$DX = \begin{cases} \sum_{i=1}^{\infty} (x_i - EX)^2 p_i & \text{离散型} \\ \int_{-\infty}^{+\infty} (x-EX)^2 f(x)dx, & \text{连续型} \end{cases}$

【注】★ $DX = EX^2 - (EX)^2$， $EX^2 = DX + (EX)^2$

3.协方差

如果随机变量 X 与 Y 的方差 $DX > 0$， $DY > 0$ 存在，则称 $E(X-EX)(Y-EY)$ 为随机变量 X 与 Y 的协方差并记为 $\operatorname{cov}(X,Y)$.

$\operatorname{cov}(X,Y) = EXY - EX \cdot EY$.

若 $Y = X$，则 $\operatorname{cov}(X,Y) = \operatorname{cov}(X,X) = DX$.

4.相关系数

$\rho_{XY} = \dfrac{\operatorname{cov}(X,Y)}{\sqrt{DX}\cdot\sqrt{DY}}$，称为随机变量 X 与 Y 的相关系数.

$$\Rightarrow \rho_{XY} = \frac{EXY - EX \cdot EY}{\sqrt{DX} \cdot \sqrt{DY}}$$

【例】设某商品每周的需求量 $X \sim U[10,30]$.当商店进货数为$[10,30]$中的某一整数时，商店每售出一件商品可获利 500 元.1）若供大于求，则降价处理，处理的每件商品亏损 100 元；2）若供不应求，则可从外调货供应，此时每件商品仅获利 300 元.为使商店所获利润的期望值不少于 9280 元，试确定最少进货量.

第四讲　如何使用极限定理？

1.依概率收敛

设数列$\left\{X_n\right\}$，$n=1,2,\cdots$，X为一随机变量（或a为一常数）. 任给正数$\varepsilon>0$，恒有$\lim\limits_{n\to\infty}P\left\{\left|X_n-X\right|<\varepsilon\right\}=1$（或$\lim\limits_{n\to\infty}P\left\{\left|X_n-a\right|<\varepsilon\right\}=1$），则称随机变量序列$X_1,X_2,\cdots,X_n,\cdots$依概率收敛于$X$（或$a$）.

2. 大数定律

切比雪夫大数定律：假设$\left\{X_n,n\ge1\right\}$是相互独立的随机变量序列，如果方差$D(X_k)(k\ge1)$存在且一致有上界，即存在常数C，使$D(X_k)\le C$，（一切$k\ge1$），则$\left\{X_n,n\ge1\right\}$服从大数定律：$\frac{1}{n}\sum\limits_{i=1}^{n}X_i-\frac{1}{n}\sum\limits_{i=1}^{n}EX_i\xrightarrow{P}0.$

伯努利大数定律：假设μ_n是n重伯努利试验中事件A发生的次数，在每次试验中事件A发生的概率为$p(0<p<1)$，则$\frac{\mu_n}{n}\xrightarrow{P}p$，即对任意$\varepsilon>0$，有$\lim\limits_{n\to\infty}P\{|\frac{\mu_n}{n}-p|\ge\varepsilon\}=0.$

辛钦大数定律：假设$\left\{X_n,n\ge1\right\}$是独立同分布的随机变量序列，如果$EX_n=\mu(n\ge1)$存在，则$\frac{1}{n}\sum\limits_{i=1}^{n}X_i\xrightarrow{P}\mu$，即对任意$\varepsilon>0$，有$\lim\limits_{n\to\infty}P\{|\frac{1}{n}\sum\limits_{i=1}^{n}X_i-\mu|\ge\varepsilon\}=0.$

3. 中心极限定理

列维——林德伯格中心极限定理（独立同分布中心极限定理）：假设$\left\{X_n,n\ge1\right\}$是独立同分布的随机变量序列，如果

$EX_n=\mu$，$DX_n=\sigma^2(n\ge1)$存在，则$\left\{X_n,n\ge1\right\}$服从中心极限定理，即对任意的实数x，有

$$\lim_{n\to\infty}P\{\frac{\sum\limits_{i=1}^{n}X_i-n\mu}{\sqrt{n\sigma}}\le x\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-\frac{1}{2}x}\mathrm{d}t=\Phi(x).$$

棣莫弗——拉普拉斯中心极限定理（二项分布以正态分布为其极限分布定理）：假设随机变量$Y_n\sim B(n,p)$，$(0<p<1,n\ge1)$，则对任意实数x，有

$$\lim_{n\to\infty}P\{\frac{Y_n-np}{\sqrt{np(1-p)}}\le x\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-x/2}\mathrm{d}t=\Phi(x).$$

第五讲 如何作估计？

1.总体与样本

总体 X ——某全体研究对象的某一指标

样本——只研究“简单随机样本”： X_i 独立同分布于 X

2.统计量

设 $(X_1,\cdots,X_n)$ 为总体 X 的样本，$g(x_1,\cdots,x_n)$ 为 n 元函数，如果 g 中不含任何未知参数，则称 $g(X_1,\cdots,X_n)$ 为样本 $(X_1,\cdots,X_n)$ 的一个统计量．若 $(x_1,\cdots,x_n)$ 为样本值，则称 $g(x_1,\cdots,x_n)$ 为 $g(X_1,\cdots,X_n)$ 的观测值．

3.矩估计

$$EX=\begin{cases}\sum\limits_{i=1}^{\infty}x_i p_i, & \text{离散型}\\ \int_{-\infty}^{+\infty}xf(x)dx, & \text{连续型}\end{cases}\triangleq\frac{1}{n}\sum_{i=1}^{n}X_i$$

【例】设总体 $X\sim f(x,\theta)=\begin{cases}(\theta+1)x^{\theta}, & 0<x<1,\\ 0, & \text{其他}\end{cases}$，求 θ 的矩估计量．

4.最大似然估计

对未知参数 θ 进行估计时，在该参数可能的取值范围 Θ 内选取使“样本获此观测值 $(x_1,\cdots,x_n)$”的概率最大的参数值 $\hat\theta$ 作为 θ 的估计，这样选定的 $\hat\theta$ 最有利于 $(x_1,\cdots,x_n)$ 的出现．

设总体 X 是离散型，其概率分布为 $P\{X=x\}=p(x;\theta)$， θ 为未知参数，

$(X_1,\cdots,X_n)$ 为 X 的一个样本，则 $(X_1,\cdots,X_n)$ 取值为 $(x_1,\cdots,x_n)$ 的概率是

$$P\{X_1=x_1,\cdots,X_n=x_n\}=\prod_{i=1}^{n}P\{X_i=x_i\}=\prod_{i=1}^{n}P(x_i,\theta)$$

显然这个概率值是θ的函数，将其记为

$$L(\theta)=L(x_1,\cdots,x_n;\theta)=\prod_{i=1}^{n}P(x_1;\theta)$$

称$L(\theta)$为样本$(x_1,\cdots,x_n)$的似然函数．若$\hat{\theta}\in$Ⓗ使

$$L(x_1,\cdots,x_n;\hat{\theta})=\max_{\theta\in H}L(x_1,\cdots,x_n;\theta)$$

则称$\hat{\theta}=\hat{\theta}(x_1,\cdots,x_n)$为未知参数$\theta$的最大似然估计值，而相应的统计量$\hat{\theta}(X_1,\cdots,X_n)$称为参数$\theta$的极大(最大)似然估计量.

同理，如果总体X是连续型的随机变量，其概率密度函数为$f(x;\theta)$，$\theta\in$Ⓗ，则样本的似然函数为$L(\theta)=L(x_1,\cdots,x_n;\theta)=\prod\limits_{i=1}^{n}f(x_i,\theta)$.

若$\hat{\theta}=\hat{\theta}(x_1,\cdots,x_n)\in$Ⓗ，使$L(\hat{\theta})=\max\limits_{\theta\in H}\prod\limits_{i=1}f(x_i,\theta)$，则称$\hat{\theta}(x_1,\cdots,x_n)$为$\theta$的最大似然估计值，相应的统计量$\hat{\theta}(X_1,\cdots,X_n)$称为$\theta$的最大似然估计.

计算步骤:

①写出似然函数：$L(\theta)=\begin{cases}\prod\limits_{i=1}^{n}f(x_i,\theta) & \text{连续型}\\ \prod\limits_{i=1}^{n}p_i(x_i,\theta) & \text{离散型}\end{cases}$

②令$\dfrac{dL(\theta)}{d\theta}=0$，求出$\theta$，即为最大似然估计$\hat{\theta}$

【例】设$X\sim\begin{pmatrix}0 & 1 & 2 & 3\\ \theta^2 & 2\theta(1\text{-}\theta) & \theta^2 & 1-2\theta\end{pmatrix}$，其中$\theta(0<\theta<\dfrac{1}{2})$为未知参数. 利用以下样本值：3，1，3，0，3，1，2，3，求θ的矩估计值与最大似然估计值.