第一章 事件与概率

随机事件

1. 事件运算：

（1）事件 A 与 B 中至少有一个发生，记为 $A \cup B$；

（2）事件 A 与 B 同时发生，记为 $A \cap B$ 或 AB；

（3）事件 A 发生而 B 不发生，记为 $A-B$；

（4） A 的对立事件，即“ A 不发生”，记为 $\bar{A}$.

2. 概率为 0 不一定为不可能事件，概率为 1 不一定为必然事件.

3. 排列数 $A_n^k = \dfrac{n!}{(n-k)!} = n(n-1)\cdots(n-k+1)$.

4. 组合数 $C_n^k = \dbinom{n}{k} = \dfrac{n!}{k!(n-k)!} = \dfrac{n(n-1)\cdots(n-k+1)}{1\cdot 2\cdot\cdots\cdot k}$.

5. n 元集的循环 r 排列数为 $\dfrac{A_n^r}{r} = \dfrac{n!}{r(n-r)!}$

概率

1. 概率 P 的三个基本性质：

（1）非负性：$P(A) \geqslant 0$；

（2）规范性：$P(\Omega) = 1$；

（3）可列可加性：若 A_i 互斥，$i = 1, 2, \cdots$，则 $P\left(\bigcup_{i=1}^{+\infty} A_i\right) = \sum_{i=1}^{+\infty} P\left(A_i\right)$

2. 概率 P 的运算性质：

（1）减法公式(特定场合）：若 $A \supset B$, 则 $P(A-B) = P(A) - P(B)$；

（2）减法公式(一般场合）：$P(A-B) = P(A) - P(AB)$.

（3）加法公式 ：$P(A \cup B) = P(A) + P(B) - P(AB)$；

推广：对任意 n 个事件 $A_1, A_2, \cdots, A_n$ 有

$$P\left(\bigcup_{i=1}^{n} A_i\right)=\sum_{i=1}^{n} P\left(A_i\right)-\sum_{1 \leqslant i<j \leqslant n} P\left(A_i A_j\right)+\sum_{1 \leqslant i<j<k \leqslant n} P\left(A_i A_j A_k\right)+\cdots+(-1)^{n-1} P(A_1 A_2 \cdots A_n).$$

3. 条件概率：

（1）定义：若 $P(B)>0$，称 $P(A \mid B)=\dfrac{P(AB)}{P(B)}$ 为在 B 发生下的条件下 A 发生的条件概率.

（2）乘法公式：若 $P(B)>0$，则 $P(AB)=P(B)P(A \mid B)$.

（3）条件概率 $P(\cdot \mid B)$ 具有概率的三个基本性质：（因而具有与概率相同的运算性质）

（i） 非负性：$P(A \mid B) \geqslant 0$；

（ii） 规范性：$P(\Omega \mid B)=1$；

（iii）可列可加性：若 A_i 互斥，$i=1,2,\cdots$，则 $P\left(\bigcup_{i=1}^{+\infty} A_i \mid B\right)=\sum_{i=1}^{+\infty} P\left(A_i \mid B\right)$

4. 全概率公式：设 $B_1, B_2, \cdots, B_n$ 互不相容，且 $\bigcup_{i=1}^{n} B_i=\Omega$，如果 $P\left(B_i\right)>0$，$i=1,2,\cdots,n$，则对任一事件 A，有 $P(A)=\sum_{i=1}^{n} P\left(B_i\right) P\left(A \mid B_i\right)$.

5. 贝叶斯公式：设 $B_1, B_2, \cdots, B_n$ 互不相容，且 $\bigcup_{i=1}^{n} B_i=\Omega$，若 $P\left(B_i\right)>0$, $i=1,2,\cdots, n$，

则 $P\left(B_i \mid A\right)=\dfrac{P\left(B_i\right) P\left(A \mid B_i\right)}{\sum_{j=1}^{n} P\left(B_j\right) P\left(A \mid B_j\right)}$，$i=1,2,\cdots, n$. 其中 $P\left(B_i\right)$ 称为 B_i 的先验概率，$P\left(B_i \mid A\right)$ 称为 B_i 的后验概率(表示在事件 A 发生后对 B 的概率作出的修正).

独立性

1. 事件的独立性：

（1）定义：若 $P(AB)=P(A)P(B)$，则称事件 A，B 相互独立.

（2）若$\begin{cases} P(AB)=P(A)P(B) \\ P(BC)=P(B)P(C) \\ P(CA)=P(C)P(A) \\ P(ABC)=P(A)P(B)P(C) \end{cases}$，则称事件$A$，$B$，$C$相互独立.（同理可推广到$n$个事件）.

（3）A，B，C相互独立$\Rightarrow A$，B，C两两独立，反之不对.

（4）若A，B，C相互独立，则$A\cup B$与C，$A\cap B$与C，$A-B$与C，$\bar{A}$与C均独立.

推广：若$A_1, A_2, \cdots, A_n$相互独立，将它们划分为k组，每一组内进行和，差，积，对立等运算所得到的k个事件仍然相互独立.

2. 伯努利实验：若实验E只有两种可能的结果，将实验E独立地重复n次，称为一个n重伯努利实验（记作B^n）.

第二章 随机变量及其分布

随机变量及其分布

1. 随机变量$(r.v.)$：

（1）仅取有限或可列个值的随机变量称为离散型随机变量.

（2）取值充满某个区间(a,b)的随机变量称为连续型随机变量，这里a可为$-\infty$，b可为$+\infty$.

2. 分布函数：设X是一个随机变量，对任意实数x，称$F(x)=P(X\leqslant x)$为X的分布函数，分布函数具有如下三条基本性质：

（1）单调性：对任意的$x_1<x_2$，有$F\left(x_1\right)\leqslant F\left(x_2\right)$.

（2）有界性：对任意的x，有$0\leqslant F(x)\leqslant 1$，且$F(-\infty)=0$，$F(+\infty)=1$.

（3）右连续性：$F(x)$是x的右连续函数，即对任意的x，有$F\left(x+0\right)=F\left(x\right)$.

除此以外，还有：$P(X<x)=F(x-0)$；$P(X=x)=F(x)-F(x-0)$.

3. 分布列：若离散随机变量X的可能取值是$x_1, x_2, \cdots, x_n, \cdots$，则称

$$p_i=P\left(X=x_i\right), i=1,2,\cdots,n,\cdots \text{（一定要写明}i\text{的取值范围）}$$

或者

X	x_1	x_2	$\cdots$	x_n	$\cdots$
P	p_1	p_2	$\ldots$	p_n	$\cdots$

为随机变量 X 的分布列.

4. 概率密度函数：记连续随机变量 X 的分布函数为 $F(x)$，若存在一个非负可积函数 $f(x)$，使得对任意实数 x，有 $F(x)=\int_{-\infty}^{x} f(t)\mathrm{d}t$，则称 $f(x)$ 为 X 的概率密度函数.

5. 密度函数的基本性质：

（1）非负性：$f(x)\geqslant 0$.

（2）正则性：$\int_{-\infty}^{+\infty} f(x)\mathrm{d}x=1$.

6. 密度函数的其他重要性质：

（1）连续型 $r.v.$ 的分布函数 $F(x)$ 是 $(-\infty,+\infty)$ 上的连续函数，它可能在有限或可列个点上不可导，除此以外，有 $F'(x)=f(x)$.（故密度函数不唯一，但它们几乎处处相等）

（2）连续随机变量 X 仅取一点值的概率为零，即 $\forall x\in\mathbb{R}$，$P(X=x)=0$.（因此对离散型 $r.v.$ 来说，概率为 0 $\Leftrightarrow$ 不可能事件；对连续型 $r.v.$ 来说，概率为 0 不一定是不可能事件）

（3）$P(a\leqslant X\leqslant b)=\int_{a}^{b} f(x)\mathrm{d}x$.

常见的离散分布

1. $0-1$ 分布：分布列为 $P\left(X=k\right)=p^k(1-p)^{1-k}$，$k=0,1$

2. 二项分布：背景：n 重伯努利实验中成功的次数 X 服从二项分布. 记为 $X\sim b(n,p)$

分布列为 $P\left(X=k\right)=\binom{n}{k}p^k(1-p)^{n-k}$，$k=0,1,\cdots n$.（若 n 取 1 即为 $0-1$ 分布）

3. 泊松分布：分布列为 $P\left(X=k\right)=\dfrac{\lambda^k}{k!}e^{-\lambda}$，$k=0,1,2,\cdots$. 记为 $X\sim P(\lambda)$.

泊松定理：n 重伯努利实验中事件 A 发生的概率为 p_n（与 n 有关），若 $np_n=\lambda>0$，

则 $\lim\limits_{n\to\infty}\binom{n}{k}p_n^k(1-p_n)^{n-k}=\dfrac{\lambda^k}{k!}e^{-\lambda}$，$k=0,1,2,\cdots$（常用来估计二项分布）

4. 超几何分布：背景：N 个产品，其中有 M 个不合格品. 若从中不放回地随机抽取 n 个，

则其中含有的不合格品的个数 X 服从超几何分布.

5. 几何分布：背景：伯努利实验中首次成功时的试验次数 X 服从几何分布.

   分布列为 $P\left(X=k\right)=(1-p)^{k-1}p,\ k=1,2,\cdots$

   特点：无记忆性：对任意正整数 m,n，$P\left(X>m+n\,|\,X>m\right)=P(X>n)$

6. 负二项分布：背景：伯努利实验中第 r 次成功时的试验次数 X 服从负二项分布. 分布列为 $P\left(X=k\right)=\begin{pmatrix}k-1\\r-1\end{pmatrix}(1-p)^{k-r}p^r,\ k=r,r+1,\cdots$. r 取1即为几何分布.

常见的连续分布

1. 均匀分布：

   若随机变量 X 的密度函数和分布函数分别为

$$f(x)=\begin{cases}\dfrac{1}{b-a}, & a<x<b\\ 0, & \text{其他}\end{cases}\qquad F(x)=\begin{cases}0, & x<a,\\ \dfrac{x-a}{b-a}, & a\leqslant x<b,\\ 1, & x\geqslant b.\end{cases}$$

   则称 X 服从区间 (a,b) 上的均匀分布，记作 $X\sim U(a,b)$.

2. 指数分布：若随机变量 X 的密度函数和分布函数分别为

$$f(x)=\begin{cases}\lambda\mathrm{e}^{-\lambda x}, & x\geqslant 0,\\ 0, & x<0,\end{cases}\quad F(x)=\begin{cases}1-\mathrm{e}^{-\lambda x}, & x\geqslant 0,\\ 0, & x<0,\end{cases}$$

   则称 X 服从指数分布，记作 $X\sim E(\lambda)$ 或 $X\sim Exp(\lambda)$，其中参数 $\lambda>0$.

   特点：无记忆性：若 $X\sim E(\lambda)$，则 $\forall s,t>0$，有 $P(X>s+t\mid X>s)=P(X>t)$.

3. 正态分布：若随机变量 X 的密度函数和分布函数分别为

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

$$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}\mathrm{e}^{-\frac{(t-\mu)^2}{2\sigma^2}}\mathrm{d}t,\quad -\infty<x<+\infty$$

   则称 X 服从正态分布，记作 $X\sim N\left(\mu,\sigma^2\right)$. 称 $\mu=0,\sigma=1$ 时的正态分布 $N(0,1)$ 为标准正态分布. 标准正态分布的密度函数和分布函数分别为

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}},\quad -\infty<x<+\infty,\qquad \Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t,\quad -\infty<x<+\infty$$

性质：（1） $\Phi(0)=\frac{1}{2}$， $\Phi(-x)=1-\Phi(x)$.

（2）若 $X\sim N\left(\mu,\sigma^2\right)$，则 $F(x)=\Phi(\frac{x-\mu}{\sigma})$.

随机变量函数的分布

1. 离散型随机变量

若

X	x_1	x_2	$\cdots$	x_n	$\cdots$
P	p_1	p_2	...	p_n	$\cdots$

则

$Y=g(X)$	$g(x_1)$	$g(x_2)$	$\cdots$	$g(x_n)$	$\cdots$
P	p_1	p_2	...	p_n	$\cdots$

如果 $g(x_i)$ 中有重复的，需要将它们合并.

2. 连续型随机变量

方法一：分布函数法（万能方法）

设 X 的概率密度与分布函数分别为 $f_X(x),\ F_X(x)$. 下面求 $Y=g(X)$ 的分布.

（1）先求 $F_Y(y)$：

$F_Y(y)=P(Y\leqslant y)=P(g(X)\leqslant y)=P(X\in S)$，其中 $S=\{X\mid g(X)\leqslant y\}$.

（2）再对 $F_Y(y)$ 求导：

$$f_Y(y)=\begin{cases}\dfrac{\mathrm{d}F_Y(y)}{\mathrm{d}y}, & \text{当}F_Y(y)\text{在}y\text{处可导时}\\ 0, & \text{当}F_Y(y)\text{在}y\text{处不可导时}\end{cases}$$

方法二：公式法

设连续随机变量 X 的密度函数为 $f_X(x)$，$Y=g(X)$.

若 $y=g(x)$ 严格单调，其反函数 $h(y)$ 有连续导函数，则 $Y=g(X)$ 的密度函数为

$$f_Y(y)=\begin{cases} f_x[h(y)]\left|h'(y)\right|, & a<y<b, \\ 0, & \text{其余} \end{cases}$$

其中 $a=\min\{g(-\infty),g(+\infty)\}$，$b=\max\{g(-\infty),g(+\infty)\}$.

定理：正态分布具有线性不变性，即若 $X \sim N\left(\mu,\sigma^2\right)$，则当 $a\neq 0$ 时，有

$Y=aX+b\sim N\left(a\mu+b,a^2\sigma^2\right)$.

定理：若 $F_X(x)$ 为严格单增的连续函数， 则 $Y=F_X(X)\sim U(0,1)$.

注意：连续型随机变量的分布函数一定连续，但连续型随机变量的函数的分布函数不一定连续.（即若 X 是连续型随机变量，则 $F_X(x)$ 一定连续，但 $Y=g(X)$ 的分布函数 $F_Y(y)$ 不一定连续）.

第三章 多维随机变量及其分布

联合分布

1. 定义：$F(x,y)=P(X\leqslant x,Y\leqslant y)$ 称为二维随机变量 (X,Y) 的联合分布函数.
2. 性质：

（1）单调性 ：$F(x,y)$ 分别对 x 或 y 是单调不减的.

（2）有界性 ：$0\leqslant F(x,y)\leqslant 1$，且 $F(-\infty,y)=F(x,-\infty)=0$，　$F(+\infty,+\infty)=1$.

（3）右连续性：对每个变量右连续：$F(x+0,y)=F(x,y)$，　$F(x,y+0)=F(x,y)$.

（4）$P(x_1<X\leqslant x_2,y_1<Y\leqslant y_2)=F(x_2,y_2)-F(x_1,y_2)-F(x_2,y_1)+F(x_1,y_1)\geqslant 0$.

3. 联合分布列：若 (X,Y) 只取有限或可列个数对 (x_i,y_j)，则称

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

或

X	Y		
	y_1	y_2	$\cdots$
x_1	p_{11}	p_{12}	$\cdots$

x_2	p_{21}	p_{22}	$\cdots$
$\vdots$	$\vdots$	$\vdots$	

为 (X,Y) 的联合分布列.

4. 联合密度函数：若存在非负函数 $f(x,y)$，使得 $F(x,y)$ 可表示为

$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\mathrm{d}u\mathrm{d}v$，则称 $f(x,y)$ 为 $F(x,y)$ 的联合密度函数.

联合密度函数性质：

（1） 非负性： $f(x,y)\geqslant 0$

（2） 正则性： $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}x\mathrm{d}y=1$

（3） 在 $F(x,y)$ 偏导数存在的点上有 $\dfrac{\partial^2 F(x,y)}{\partial x\partial y}=f(x,y)$

（4） 若 G 为平面上的一个区域，则有 $P((X,Y)\in G)=\iint\limits_{G} f(x,y)\mathrm{d}x\mathrm{d}y.$

5. 常用连续型二维分布：

(1) 二维均匀分布：

概率密度 $f(x,y)=\begin{cases}\dfrac{1}{S_G}, & (x,y)\in G\\ 0, & 其余\end{cases}$，　记为 $(X,Y)\sim U(G)$.

二维均匀分布的边际分布不一定是一维均匀分布.

(2) 二维正态分布： $f(x,y)=$

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

$-\infty<x,y<+\infty$. 记为 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$.

二维正态分布的边际分布是一维正态分布. 即若 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，则

$X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$

边际分布

1. 边际分布函数：若二维随机变量 (X,Y) 的联合分布函数为 $F(x,y)$，则称

$F_X(x)=F(x,+\infty),\quad -\infty<x<+\infty$ 为 X 的边际分布. Y 的边际分布同理.

2. 边际分布列：若二维离散随机变量(X,Y)的联合分布列为$\{p_{ij}\}$，则称

$p_{i\cdot}=\sum_{j=1}^{\infty}p_{ij}$， $i=1,2,\cdots$为X的边际分布列. Y的边际分布列同理.

3. 边际概率密度：称$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y\ (-\infty<x<+\infty)$为$X$的边际概率密度. Y的边际概率密度同理.

4. 由联合分布能得到边际分布，但若只有边际分布则不能得到联合分布.

条件分布

1. 离散型随机变量:

若$P(Y=y_j)>0$, 称$P(X=x_i\,|\,Y=y_j)=\dfrac{P(X=x_i,Y=y_j)}{P(Y=y_j)}=\dfrac{p_{ij}}{p_{\cdot j}},i=1,2,\cdots$为在

$Y=y_j$条件下X的条件分布列.

2. 连续型随机变量:

对固定的y，若$f_Y(y)>0$，则称$f_{X|Y}(x\,|\,y)=\dfrac{f(x,y)}{f_Y(y)}$为在$Y=y$的条件下$X$的条

件概率密度, $F_{X|Y}(x\,|\,y)=\int_{-\infty}^{x}\dfrac{f(u,y)}{f_Y(y)}\mathrm{d}u$为条件分布函数.

3. 由条件分布＋边际分布能推出联合分布.

随机变量的独立性

1. 定义：对二维随机变量(X,Y),

（1） 若$F(x,y)=F_X(x)F_Y(y)$，则称X,Y相互独立.

（2） 对离散型随机变量，若$p_{ij}=p_{i\cdot}p_{\cdot j}$，则称$X,Y$相互独立.

（3） 对连续型随机变量，若$f(x,y)=f_X(x)f_Y(y)$，则称X,Y相互独立.

2. 若X,Y相互独立，则可由边际分布确定联合分布.

3. 定理：若X,Y相互独立， $f(x),g(y)$连续, 则$f(X),g(Y)$也相互独立.

二维随机变量函数的分布

1. 离散型随机变量：$Z=g(X,Y)$，则 $P(Z=z_k)=P(g(X,Y)=z_k)=\sum\limits_{g(x_i,y_j)=z_k} p_{ij}$.

2. 连续型随机变量：$Z=g(X,Y)$，则 $F_Z(z)=P(Z\leqslant z)=P(g(X,Y)\leqslant z)$

$=\iint\limits_{D_z} f(x,y)\mathrm{d}x\mathrm{d}y$，其中 $D_z=\{(x,y)\mid g(x,y)\leqslant z\}$

3. 变量替换法：若 $\begin{cases}u=g_1(x,y)\\ v=g_2(x,y)\end{cases}$ 存在反函数 $\begin{cases}x=x(u,v)\\ y=y(u,v)\end{cases}$，且雅可比行列式 $J=\begin{vmatrix}\frac{\partial x}{\partial u} & \frac{\partial x}{\partial v}\\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v}\end{vmatrix}$

$\neq 0$，则 $\begin{cases}U=g_1(X,Y)\\ V=g_2(X,Y)\end{cases}$ 的联合密度函数为 $f_{UV}(u,v)=f_{XY}(x(u,v),y(u,v))\,|J|$.

4. 增补变量法：求 $Z=g(X,Y)$ 的概率密度，可先增补变量 $W=h(X,Y)$，用变量变换法求 (Z,W) 的联合概率密度 $f_{ZW}(z,w)$，再求其边际概率密度 $f_Z(z)$. 用此方法可导出 $X+Y$，XY，$\dfrac{Y}{X}$ 的分布（为简单起见，令 $h(X,Y)=Y$ 即可）.

5. $Z=X+Y$ 的分布：$f_{X+Y}(z)=\int_{-\infty}^{+\infty} f(x,z-x)\mathrm{d}x=\int_{-\infty}^{+\infty} f(z-y,y)\mathrm{d}y$. 若 X,Y 相互独立，则 $f_Z(z)=\int_{-\infty}^{+\infty} f_X(x)f_Y(z-x)\mathrm{d}x=\int_{-\infty}^{+\infty} f_X(z-y)f_Y(y)\mathrm{d}y$.（卷积公式）

6. $Z=XY$ 的分布：$f_{XY}(z)=\int_{-\infty}^{+\infty}\dfrac{1}{|x|}f(x,\dfrac{z}{x})\mathrm{d}x$.

7. $Z=\dfrac{Y}{X}$ 的分布：$f_{Y/X}(z)=\int_{-\infty}^{+\infty}|x|\,f(x,xz)\mathrm{d}x$.

8. $\max\{X,Y\}$ 的分布：X,Y 相互独立，则 $F_{\max}(z)=P(\max\{X,Y\}\leqslant z)=P(X\leqslant z,Y\leqslant z)$

$=P(X\leqslant z)P(Y\leqslant z)=F_X(z)F_Y(z)$.

推广：$X_1,X_2,\cdots,X_n$ 独立同分布，密度函数与分布函数分别为 $f(x)$ 与 $F(x)$，

$Z=\max\{X_1,X_2,\cdots,X_n\}$，则 $F_{\max}(z)=[F(z)]^n$，$f_{\max}(z)=n[F(z)]^{n-1}f(z)$.

9. $\min\{X,Y\}$ 的分布：X,Y 相互独立，则 $F_{\min}(z)=P(\min\{X,Y\}\leqslant z)$

$=1-P(\min\{X,Y\}>z)=1-P(X>z)P(Y>z)=1-[1-P(X\leqslant z)][1-P(Y\leqslant z)]$

$=1-[1-F_X(z)][1-F_Y(z)]$.

推广：$X_1,X_2,\cdots,X_n$ 独立同分布，密度函数与分布函数分别为 $f(x)$ 与 $F(x)$，

$Z=\min\{X_1,X_2,\cdots,X_n\}$，则 $F_{\min}(z)=1-[1-F(z)]^n$，$f_{\min}(z)=n[1-F(z)]^{n-1}f(z)$.

分布的可加性

1. 定义：若同一类分布的相互独立的随机变量的和仍是此类分布，则称此类分布具有可加性.

2. 具有可加性的常用分布：（此处均假设 X,Y 相互独立）

（1） 二项分布：$X\sim B(n_1,p)$, $Y\sim B(n_2,p)$，则 $X+Y\sim B(n_1+n_2,p)$

（2） 泊松分布：$X\sim P(\lambda_1)$, $Y\sim P(\lambda_2)$，则 $X+Y\sim P(\lambda_1+\lambda_2)$

（3） 正态分布：$X\sim N(\mu_1,\sigma_1^2)$, $Y\sim N(\mu_2,\sigma_2^2)$, 则 $X+Y\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$.

推广：$X_i\sim N(\mu_i,\sigma_i^2)$，$i=1,2,\cdots n$. X_i 相互独立，a_i 不全为零，则

$\sum_{i=1}^{n}a_iX_i\sim N(\sum_{i=1}^{n}a_i\mu_i,\sum_{i=1}^{n}a_i^2\sigma_i^2)$（独立正态随机变量的线性组合仍服从正态分布）

第四章 随机变量的数字特征

数学期望

1. 定义：若 $\begin{cases}\sum_{i=1}^{\infty}|x_i|p_i<+\infty\text{，} & X\text{为离散型}r.v.\\ \int_{-\infty}^{+\infty}|x|f(x)\mathrm{d}x<+\infty\text{，} & X\text{为连续型}r.v.\end{cases}$，则称

$E(X)=\begin{cases}\sum_{i=1}^{\infty}x_ip_i\text{，} & X\text{为离散型}r.v.\\ \int_{-\infty}^{+\infty}xf(x)\mathrm{d}x\text{，} & X\text{为连续型}r.v.\end{cases}$ 为 X 的数学期望, 否则称 X 的数学期望不存在. 其中 $p_i=p(X=x_i)$，$f(x)$ 为 X 的概率密度.（以下所出现数学期望均假设其存在）

2. 定理：

$$E(g(X))=\begin{cases}\sum_{i=1}^{\infty}g(x_i)p_i\text{，} & X\text{为离散型}r.v.\\ \int_{-\infty}^{+\infty}g(x)f(x)\mathrm{d}x\text{，} & X\text{为连续型}r.v.\end{cases}$$

$$E(g(X,Y))=\begin{cases}\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}g(x_i,y_j)p_{ij}\ , & X\text{为离散型}r.v.\\ \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x,y)f(x,y)\mathrm{d}x\mathrm{d}y\ , & X\text{为连续型}r.v.\end{cases}$$

3. 性质：

（1） $E(C)=C$；

（2）线性性：$E(aX+bY)=aE(X)+bE(Y)$；

推广：$E(\sum_{i=1}^{n}a_iX_i+C)=\sum_{i=1}^{n}a_iE(X_i)+C$；

（3）若 X 与 Y 相互独立，则 $E(XY)=E(X)E(Y)$；但反之，由 $E(XY)=E(X)E(Y)$ 不能得出 X 与 Y 相互独立；

（4）施瓦茨不等式：$|E(XY)|^2\leqslant E(X^2)E(Y^2)$；

4. 条件数学期望：

（1）定义：$E(X\,|\,Y=y)=\begin{cases}\sum_{i=1}^{\infty}x_iP\left(X=x_i\,|\,Y=y\right)\ , & (X,Y)\text{为离散型}r.v.\\ \int_{-\infty}^{+\infty}xf_{X|Y}(x\,|\,y)\mathrm{d}x\ , & (X,Y)\text{为连续型}r.v.\end{cases}$

称为 X 的条件期望．Y 的条件期望 $E(Y\,|\,X=x)$ 同理.

（2）性质：条件期望 $E(X\,|\,Y=y)$ 具有数学期望的一切性质，它是关于 y 的函数.

（3）重期望公式：$E(X)=E(E(X\,|\,Y))$

方差

1. 定义：$D(X)=E[X-E(X)]^2$ 称为 X 的方差．$\sigma(X)=\sqrt{D(X)}$ 称为 X 的标准差.

2. 方差常用计算公式：$D(X)=E(X^2)-[E(X)]^2$

3. 性质：

（1） $D(C)=0$

（2） $D(CX)=C^2D(X)$

（3） $D(X+Y)=D(X)+D(Y)+2E[(X-E(X))(Y-E(Y))]$.

若 X 与 Y 相互独立, 则 $D(X+Y)=D(X)+D(Y)$. 但 $D(X+Y)=D(X)+D(Y)$

不能推出 X 与 Y 相互独立.

推广：若 X_i 相互独立，则 $D(\sum_{i=1}^{n} a_i X_i + C) = \sum_{i=1}^{n} a_i^2 D(X_i)$；

（4） $D(X)=0$ 当且仅当 X 取常值 C 的概率为 1，即 $P(X=C)=1$；

（5） $D(X) \leqslant E(X-C)^2$，当且仅当 $C=E(X)$ 时取等号；

4. 切比雪夫不等式：设 X 的数学期望为 μ，方差为 σ^2，则 $P(|X-\mu| \geqslant \varepsilon) \leqslant \dfrac{\sigma^2}{\varepsilon^2}$ 或 $P(|X-\mu|<\varepsilon) \geqslant 1-\dfrac{\sigma^2}{\varepsilon^2}$.

5. 随机变量的标准化：称 $X^* = \dfrac{X-E(X)}{\sqrt{D(X)}}$ 为 X 的标准化随机变量. 此时 $E(X^*)=0$，$D(X^*)=1$

常见分布的数学期望与方差

常见分布	分布列/密度函数	期望	方差
0-1 分布	$P(X=k)=p^k(1-p)^{1-k},\ k=0,1$	p	$p(1-p)$
二项分布	$P(X=k)=\binom{n}{k} p^k(1-p)^{n-k},\ k=0,1,\cdots,\ n$	np	$np(1-p)$
泊松分布	$P(X=k)=\dfrac{\lambda^k}{k!}e^{-\lambda},\ k=0,1,2,\cdots$	λ	λ
均匀分布	$f(x)=\dfrac{1}{b-a}, a<x<b$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$
指数分布	$f(x)=\lambda \mathrm{e}^{-\lambda x}, x \geqslant 0$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$
正态分布	$f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}},\quad -\infty<x<+\infty,$	μ	σ^2

协方差

1. 定义：$\operatorname{Cov}(X,Y)=E[(X-E(X))(Y-E(Y))]$ 称为 X 与 Y 的协方差.

2. 性质：

（1）$\operatorname{Cov}(X,Y)=E(XY)-E(X)E(Y)$，从而若 X 与 Y 相互独立，则 $\operatorname{Cov}(X,Y)=0$.

（2）$\operatorname{Cov}(aX,bY)=ab\operatorname{Cov}(X,Y)$

（3）$\operatorname{Cov}(X+Y,Z)=\operatorname{Cov}(X,Z)+\operatorname{Cov}(Y,Z)$

3. 协方差与方差的关系：

（1）$\operatorname{Cov}(X,X)=D(X)$

（2）$D(X+Y)=D(X)+D(Y)+2\operatorname{Cov}(X,Y)$

推广：$D(\sum_{i=1}^{n}X_i)=\sum_{i=1}^{n}D(X_i)+2\sum_{i<j}\operatorname{Cov}(X_i,X_j)$

（3）施瓦茨不等式：$|\operatorname{Cov}(X,Y)|^2\leqslant D(X)D(Y)$

相关系数

1. 定义：$\rho_{XY}=\dfrac{\operatorname{Cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$ 称为随机变量 X 与 Y 的相关系数，用于表征 X 与 Y 线性关系的强弱.

2. 若 $\rho_{XY}=1$，称 X 与 Y 正线性相关；$\rho_{XY}=-1$，称 X 与 Y 负线性相关；若 $\rho_{XY}=0$，称 X 与 Y 不相关.

3. 性质：

（1）$|\rho_{XY}|\leqslant 1$.

（2）$\rho(kX,kY)=\rho(X,Y)$.

（3）$\rho_{XY}=\operatorname{Cov}(X^*,Y^*)$，其中 X^*,Y^* 分别为 X 与 Y 标准化随机变量.

（4）$|\rho_{XY}|=1\Leftrightarrow$ X 与 Y 几乎处处有线性关系，即 $P(Y=aX+b)=1$，其中 $a\neq 0$.

$\rho_{XY}=1$ 时 $a>0$，$\rho_{XY}=-1$ 时 $a<0$.

4. X 与 Y 独立 $\Rightarrow$ X 与 Y 不相关，反之不对.

5. 对二维正态分布 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，$X$ 与 Y 的相关系数 $\rho_{XY}=\rho$，且 X 与 Y 独立 $\Leftrightarrow$ X 与 Y 不相关

矩与协方差矩阵

1. 定义：

   （1） $E(X^k)$ 称为 X 的 k 阶(原点)矩.

   （2） $E[X-E(X)]^k$ 称为 X 的 k 阶中心矩.

   （3） $E(X^kY^l)$ 称为 X 与 Y 的 $k+l$ 阶混合(原点)矩.

   （4） $E[(X-E(X))^k(Y-E(Y))^l]$ 称为 X 与 Y 的 $k+l$ 阶混合中心矩.

2. （1）数学期望 $E(X)$ 是一阶原点矩.

   （2）方差 $D(X)=E[X-E(X)]^2$ 是二阶中心矩.

   （3）协方差 $\mathrm{Cov}(X,Y)=E[(X-E(X))(Y-E(Y))]$ 是二阶混合中心矩.

3. 定义：n 维随机变量 $(X_1,X_2,\cdots X_n)$ 的协方差矩阵为

$$\boldsymbol{C}=\begin{pmatrix} D(X_1) & \mathrm{Cov}(X_1,X_2) & \dots & \mathrm{Cov}(X_1,X_n) \\ \mathrm{Cov}(X_2,X_1) & D(X_2) & \cdots & \mathrm{Cov}(X_2,X_n) \\ \vdots & \vdots & & \vdots \\ \mathrm{Cov}(X_n,X_1) & \mathrm{Cov}(X_n,X_2) & \cdots & D(X_n) \end{pmatrix}$$，是一个对称非负定矩阵.

n维正态分布的性质

1. $(X_1,X_2,\cdots X_n)$ 服从 n 维正态分布，则每个 X_i 服从一维正态分布;若每个 X_i 服从一维正态分布，且 X_i 相互独立，则 $(X_1,X_2,\cdots X_n)$ 服从 n 维正态分布.

2. $(X_1,X_2,\cdots X_n)$ 服从 n 维正态分布 $\Leftrightarrow$ 任一线性组合 $l_1X_1+l_2X_2+\cdots+l_nX_n$ 服从一维正态分布.

3. $(X_1,X_2,\cdots X_n)$ 服从 n 维正态分布, $Y_1,Y_2,\cdots,Y_k$ 均为 $X_1,X_2,\cdots,X_n$ 的线性组合，则 $(Y_1,Y_2,\cdots,Y_k)$ 服从 k 维正态分布（正态分布的线性不变性）.

4. $(X_1,X_2,\cdots X_n)$ 服从 n 维正态分布, 则 $X_i,i=1,2,\cdots n$ 独立 $\Leftrightarrow$ $X_i,i=1,2,\cdots n$ 不相关.

5. n 维正态分布 $(X_1,X_2,\cdots X_n)$ 的密度函数为

$$f(x_1,x_2,\cdots,x_n)=\frac{1}{(2\pi)^{n/2}|\boldsymbol{C}|^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}C^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}$$，其中 $\boldsymbol{\mu}$ 为

$(X_1,X_2,\cdots X_n)$ 的数学期望向量，$\boldsymbol{C}$ 为 $(X_1,X_2,\cdots X_n)$ 的协方差矩阵.

第五章 大数定律与中心极限定理

依概率收敛与按分布收敛

1. 定义：

（1）若 $\forall\varepsilon>0$, $\lim\limits_{n\to\infty}P(|X_n-X|<\varepsilon)=1$, 则称 $\{X_n\}$ 依概率收敛于 X , 记作

$X_n\xrightarrow{P}X$.

（2）若对 $F_n(x)$ 的任意连续点 x ，都有 $\lim\limits_{n\to\infty}F_n(x)=F(x)$, 则称 $\{F_n(x)\}$ 弱收敛于

$F(x)$, $\{X_n\}$ 按分布收敛于 X . 分别记作 $F_n(x)\xrightarrow{W}F(x)$, $X_n\xrightarrow{L}X$.

2. 依概率收敛的运算性质：设 $X_n\xrightarrow{P}a$, $Y_n\xrightarrow{P}b$, 则

（1） X_n 与 Y_n 之间的四则运算分别依概率收敛到对应的 a 与 b 之间的四则运算.

（2）若 $g(x,y)$ 在点 (a,b) 连续，则 $g(X_n,Y_n)\xrightarrow{P}g(a,b)$.

3. 依概率收敛与按分布收敛的关系：

（1） $X_n\xrightarrow{P}X\Rightarrow X_n\xrightarrow{L}X$.

（2） $X_n\xrightarrow{P}c\Leftrightarrow X_n\xrightarrow{L}c$, 其中 c 为常数.

大数定律

1. 定义：若 $\forall\varepsilon>0$, $\lim\limits_{n\to\infty}P\left(\left|\frac{1}{n}\sum\limits_{i=1}^{n}X_i-\frac{1}{n}\sum\limits_{i=1}^{n}E(X_i)\right|<\varepsilon\right)=1$, 则称 $\{X_n\}$ 服从大数定律.

2. 依概率收敛与大数定律的关系：记 $Y_n=\frac{1}{n}\sum\limits_{i=1}^{n}X_i$, 则 $\{X_n\}$ 服从大数定律 $\Leftrightarrow$

$Y_n-E(Y_n)\xrightarrow{P}0$

3. 伯努利大数定律：记 f_A 为 n 重伯努利试验中 A 发生的次数，则 $\forall\varepsilon>0$ ，

$$\lim_{n\to\infty} P\left(\left|\frac{f_A}{n}-p\right|<\varepsilon\right)=1 \ .$$

4. 切比雪夫大数定律：若 X_i 互不相关， $D(X_i)\leqslant c$, $i=1,2,\cdots$, 则 $\{X_n\}$ 服从大数定律.

5. 马尔可夫大数定律：若 $\frac{1}{n^2}D(\sum_{i=1}^{n}X_i)\to 0(n\to\infty)$，则 $\{X_n\}$ 服从大数定律.

6. 辛钦大数定律(弱大数定律)：若 $\{X_n\}$ 独立同分布，数学期望存在，则 $\{X_n\}$ 服从大数定律.

7. 切比雪夫不等式： $P(|X-E(X)|\geqslant\varepsilon)\leqslant\frac{D(X)}{\varepsilon^2}$.

8. 马尔可夫不等式： $P(|X|\geqslant\varepsilon)\leqslant\frac{E(|X|^k)}{\varepsilon^k}$.

9. 重要不等式：若随机变量 X 只取非负值，则 $\forall\varepsilon>0$, $P(X\geqslant\varepsilon)\leqslant\frac{E(X)}{\varepsilon}$.

中心极限定理

1. 林德伯格—莱维中心极限定理：设 $\{X_n\}$ 独立同分布， $E(X_k)=\mu$, $D(X_k)=\sigma^2>0$,

记 $Y_n=\frac{\sum_{k=1}^{n}X_k-n\mu}{\sqrt{n}\sigma}$, 则对任意 y , $\lim_{n\to\infty}P(Y_n\leqslant y)=\Phi(y)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t$.

2. 棣莫弗—拉普拉斯中心极限定理：设 n 重伯努利实验中事件 A 在每次实验中发生的概率为 $p(0<p<1)$, S_n 为 n 次实验中 A 出现的次数. 记 $Y_n=\frac{S_n-np}{\sqrt{np(1-p)}}$ ，则对任意 y ,

$$\lim_{n\to\infty}P(Y_n\leqslant y)=\Phi(y)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t \ .$$

第六章 统计量及其分布

总体与样本

1. 定义：一个统计问题中, 研究对象的全体称为总体, 每个成员称为个体. 从总体中随机抽取的部分个体所成的集合称为样本.

2. 定义：若样本 $X_1,X_2,\cdots,X_n$ 相互独立, 且与总体同分布, 则称该样本为简单随机样本,

仍简称样本.（以后所出现样本均指简单随机样本）

3. 若总体的分布函数为 $F(x)$，则样本的（联合）分布函数为 $F(x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}F(x_i)$，称为样本分布.

4. 若总体 X 为连续型随机变量，概率密度为 $f(x)$，则样本的（联合）密度函数为

$$f(x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}f(x_i).$$

5. 若总体 X 为离散型随机变量，分布列为 $p(x_i)=P(X=x_i)$，则样本的（联合）分布列为

$$p(x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}p(x_i).$$

6. 定义：将样本观测值 $x_1,x_2,\cdots,x_n$ 从小到大排序，得到 $x_{(1)}\leqslant x_{(2)}\leqslant\cdots\leqslant x_{(n)}$，定义

$F_n(x)=\begin{cases}0, & x<x_{(1)}\\ \dfrac{k}{n}, & x_{(k)}\leqslant x<x_{(k+1)},\ k=1,2,\cdots,n-1\\ 1, & x\geqslant x_{(n)}\end{cases}$，称 $F_n(x)$ 为该样本的经验分布函数，它满足分布函数的性质.

7. 格利文科定理：设 $F_n(x)$ 为样本 $x_1,x_2,\cdots,x_n$ 的经验分布函数，则对任意的 x，

$$P(\lim_{n\to\infty}\sup_{-\infty<x<+\infty}|F_n(x)-F(x)|=0)=1.$$

统计量及其分布

1. 定义： $X_1,X_2,\cdots,X_n$ 为 X 的样本，$g(X_1,X_2,\cdots,X_n)$ 为 $X_1,X_2,\cdots,X_n$ 的函数，若 $g(X_1,X_2,\cdots,X_n)$ 不含未知参数，则称其为一个统计量. 统计量的分布称为抽样分布.

2. 样本均值：$\overline{X}=\dfrac{1}{n}\sum_{i=1}^{n}X_i$ 称为 $X_1,X_2,\cdots,X_n$ 的样本均值，其观测值 $\overline{x}=\dfrac{1}{n}\sum_{i=1}^{n}x_i$.

性质：

（1） $\sum_{i=1}^{n}(x_i-\overline{x})=0$

（2）对任意的 c，$\sum_{i=1}^{n}(x_i-\overline{x})^2\leqslant\sum_{i=1}^{n}(x_i-c)^2$

（3）若总体分布为$N(\mu,\sigma^2)$，则$\bar{x}$精确分布于$N(\mu,\dfrac{\sigma^2}{n})$

（4）若总体分布未知或不为正态分布，但$E(X)=\mu$，$D(X)=\sigma^2$，则当n较大时，$\bar{x}$的渐进分布为$N(\mu,\dfrac{\sigma^2}{n})$

3. 样本方差：$S^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$，其观察值$s^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$.

4. 样本标准差：$S=\sqrt{S^2}=\sqrt{\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2}$，其观察值$s=\sqrt{\dfrac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$.

5. 定理：若总体X有期望$E(x)=\mu$，方差$D(x)=\sigma^2$，样本$x_1,x_2,\cdots,x_n$的样本均值为$\bar{x}$，样本方差为s^2，则$E(\bar{x})=\mu$，$D(\bar{x})=\dfrac{\sigma^2}{n}$，$E(s^2)=\sigma^2$（不依赖于总体的分布形式）.

6. 样本k阶（原点）矩：$A_k=\dfrac{1}{n}\sum_{i=1}^{n}X_i^k$，其观测值$a_k=\dfrac{1}{n}\sum_{i=1}^{n}x_i^k$. 样本均值$\bar{x}$为一阶原点矩.

7. 样本k阶中心矩：$B_k=\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k$，其观测值$b_k=\dfrac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^k$.

8. 注意：样本方差S^2与样本二阶中心矩S_n^2不同：

（1）$S^2=\dfrac{n}{n-1}S_n^2$

（2）$E(S_n^2)=\dfrac{n-1}{n}\sigma^2$，$E(S^2)=\sigma^2$

9. 定理：若总体X的k阶矩$E(X^k)$存在（记为μ_k），则当$n\to\infty$时，$A_k\xrightarrow{P}\mu_k$. 即是说样本k阶矩依概率收敛到总体k阶矩.（该结论是后面矩估计法的理论根据）

10. 次序统计量：

（1）定义：设样本为$X_1,X_2,\cdots,X_n$，观测值为$x_1,x_2,\cdots,x_n$. 若$X_{(i)}$的取值是将观测值从小到大排序后得到的第i个数，则称$X_{(i)}$为该样本的第i个次序统计量.

（2）$X_{(1)}=\min\{X_1,X_2,\cdots,X_n\}$称为最小次序统计量；

$X_{(n)}=\max\{X_1,X_2,\cdots,X_n\}$称为最大次序统计量.

（3）$D_n = X_{(n)} - X_{(1)}$ 称为样本极差.

（4）次序统计量 $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ 既不独立, 也不同分布.

11. p 分位数：

（1）定义：将样本观测值 $x_1, x_2, \cdots, x_n$ 从小到大排序, 得到 $x_{(1)} \leqslant x_{(2)} \leqslant \cdots \leqslant x_{(n)}$, 对

$$0 < p < 1, \text{称 } x_p = \begin{cases} x_{([np]+1)}, & np\text{不为整数} \\ \dfrac{1}{2}[x_{(np)} + x_{(np+1)}], & np\text{为整数} \end{cases} \text{为样本的 } p \text{ 分位数.}$$

（2）样本的 0.5 分位数 $x_{0.5}$ 又称为中位数, 记为 Q_2 或 M ；

0.25 分位数 $x_{0.25}$ 又称为第一四分位数, 记为 Q_1；

0.75 分位数 $x_{0.75}$ 又称为第三四分位数, 记为 Q_3.

三大抽样分布

1. 定义：设 $X \sim N(0,1)$,

（1）若 $P(X > z_\alpha) = \alpha$, 则称 z_α 为标准正态分布的上 α 分位数.（且由正态分布对称性知 $z_{1-\alpha} = -z_\alpha$）;

（2）若 $P(|X| > z_{\frac{\alpha}{2}}) = \alpha$, 则称 $z_{\frac{\alpha}{2}}$ 为标准正态分布的双侧 α 分位数.

2. χ^2 分布：

（1） 定义：设 $X_1, X_2, \cdots, X_n$ 为来自总体 $N(0,1)$ 的样本, 则称统计量

$\chi^2 = X_1^2 + X_2^2 + \cdots + X_n^2$ 服从自由度为 n 的 χ^2 分布, 记为 $\chi^2 \sim \chi^2(n)$.

（2） 性质：

①可加性：若 $\chi_1^2 \sim \chi^2(n_1)$, $\chi_2^2 \sim \chi^2(n_2)$, 且 χ_1^2, χ_2^2 相互独立, 则

$\chi_1^2 + \chi_2^2 \sim \chi^2(n_1 + n_2)$.（可推广到多个变量）

②期望与方差：若 $\chi^2 \sim \chi^2(n)$, 则 $E(\chi^2) = n$, $D(\chi^2) = 2n$.

③极限分布：若 $\chi^2 \sim \chi^2(n)$, 则当 $n \to \infty$ 时, $\chi^2(n) \to$ 正态分布.

（3） 上 α 分位点：对给定的数 $\alpha \in (0,1)$, 称满足 $P(\chi^2 > \chi_\alpha^2(n)) = \alpha$ 的点 $\chi_\alpha^2(n)$ 为

$\chi^2(n)$ 的上 α 分位点.

3. t 分布:

（1） 定义：若 $X \sim N(0,1)$, $Y = Y_1^2 + Y_2^2 + \cdots + Y_n^2 \sim \chi^2(n)$, 且 X,Y 相互独立, 则称 $t = \dfrac{X}{\sqrt{\dfrac{Y_1^2 + Y_2^2 + \cdots + Y_n^2}{n}}}$ 服从自由度为 n 的 t 分布, 记为 $t \sim t(n)$, 是对称分布.

（2） 期望与方差： $n = 1$ 时, 期望不存在； $n > 1$ 时, 期望为 0 ；$n > 2$ 时, 方差为 $\dfrac{n}{n-2}$.

（3） 上 α 分位点：对给定的数 $\alpha \in (0,1)$, 称满足 $P(t > t_\alpha(n)) = \alpha$ 的点 $t_\alpha(n)$ 为 $t(n)$ 的上 α 分位点. 且 $t_\alpha(n) + t_{1-\alpha}(n) = 0$ (互为相反数).

4. F 分布:

（1） 定义：若 $X_1^2 + X_2^2 + \cdots + X_m^2 \sim \chi^2(m)$ ， $Y_1^2 + Y_2^2 + \cdots + Y_n^2 \sim \chi^2(n)$ ，则称

$$F = \frac{\dfrac{X_1^2 + X_2^2 + \cdots + X_m^2}{m}}{\dfrac{Y_1^2 + Y_2^2 + \cdots + Y_n^2}{n}}$$ 服从自由度为 (m,n) 的 F 分布, 记为 $F \sim F(m,n)$.

（2） 由定义易知若 $F \sim F(m,n)$, 则 $\dfrac{1}{F} \sim F(n,m)$.

（3） 上 α 分位点：对给定的数 $\alpha \in (0,1)$, 称满足 $P(F > F_\alpha(m,n)) = \alpha$ 的点 $F_\alpha(m,n)$ 为 $F(m,n)$ 的上 α 分位点. 且 $F_\alpha(m,n)F_{1-\alpha}(n,m) = 1$.

（4） 若 $X \sim t(n)$, 则 $X^2 \sim F(1,n)$, 且 $[t_{1-\frac{\alpha}{2}}(n)]^2 = F_\alpha(1,n)$

5. 定理(一个正态总体): 总体 $N(\mu,\sigma^2)$, 样本 $X_1, X_2, \cdots, X_n$ 的样本均值为 $\bar{X}$, 样本方差为 S^2 , 则

（1） $\bar{X} \sim N(\mu, \dfrac{\sigma^2}{n})$

（2） $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

（3） $\bar{X}$ 与 S^2 独立

（4） $\dfrac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t(n-1)$（由前三条结论推得）

6. 定理(两个正态总体):设 $X_1, X_2, \cdots, X_{n_1}$ 与 $Y_1, Y_2, \cdots, Y_{n_2}$ 分别为 $N(\mu_1, \sigma_1^2)$，$N(\mu_2, \sigma_2^2)$ 的样本, 且相互独立, 样本均值分别为 $\overline{X}$ 与 $\overline{Y}$, 样本方差分别为 S_1^2 与 S_2^2, 则

（1） $\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}} \sim N(0,1)$

（2） $\dfrac{S_1^2 / S_2^2}{\sigma_1^2 / \sigma_2^2} \sim F(n_1-1, n_2-1)$

（3） $\sigma_1^2 = \sigma_2^2$ 时， $\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} \sim t(n_1+n_2-2)$, 其中

$$S_w^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2},\ S_w = \sqrt{S_w^2}.$$

第七章 参数估计

基本概念

1. 定义：从总体中抽出一个样本，用某种方法对未知参数进行估计，称为参数估计. 常用 θ 表示参数. 参数 θ 的所有可能取值组成的集合称为参数空间，常用 Θ 表示.
2. 分类：

（1）点估计：设分布函数形式已知，但它的一个或多个参数未知，借助样本来估计总体未知参数的值，称为点估计.

（2）区间估计：估计未知参数的取值范围, 并使此范围包含未知参数真值的概率为给定的值，称为区间估计.

3. 点估计的一般提法：

设总体 X 的分布函数 $F(x;\theta)$ 形式已知，参数 θ 未知. $X_1, X_2, \cdots, X_n$ 为 X 的样本.

$x_1, x_2, \cdots, x_n$ 为相应样本值. 点估计就是要构造一个适当的统计量 $\hat{\theta}(X_1, X_2, \cdots, X_n)$，用它的观测值 $\hat{\theta}(x_1, x_2, \cdots, x_n)$ 来估计未知参数 θ.

4. 常用构造估计量的方法 $\begin{cases} \text{矩估计法} \\ \text{最大似然估计法} \end{cases}$

矩估计法

1. 定义：用样本矩及其函数去替换相应的总体矩及其函数的方法称为矩估计法. (称为替换原理)
2. 思想：用样本 k 阶矩作为总体 k 阶矩的估计量.
3. 例子：
   - 用样本均值估计总体均值，即 $\hat{E}(X) = \overline{x}$；
   - 用样本二阶中心矩估计总体方差，即 $\hat{D}(X) = s_n^2$；
4. 实质：用经验分布函数替换总体分布函数.
5. 优点：简单易行，不需要事先知道总体的分布.
6. 缺点：当总体类型已知时，没有充分利用信息；此外，矩估计可能不唯一，此时通常采用低阶矩给出未知参数的估计.

最大似然估计法

1. 总体 X 为离散型:

   设总体 X 的分布律为 $P(X = x) = p(x;\theta), \theta \in \Theta$. $X_1, X_2, \cdots, X_n$ 是来自总体 X 的一个样本.

   （1） 写出样本的似然函数 $L(\theta) = L(\theta; x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} p(x_i;\theta)$；

   （2） 选取使似然函数 $L(\theta)$ 取最大值的 $\hat{\theta}$ 作为未知参数 θ 的估计值，即 $\hat{\theta}$ 满足 $L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta)$.

2. 总体 X 为连续型:

   设总体 X 的概率密度为 $f(x;\theta), \theta \in \Theta$. $X_1, X_2, \cdots, X_n$ 是来自总体 X 的一个样本.

（1） 写出样本的似然函数 $L(\theta)=L(\theta;x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}f(x_i;\theta)$

（2） 选取使似然函数 $L(\theta)$ 取最大值的 θ 作为未知参数 θ 的估计值，即 θ 满足 $L(\hat{\theta})=\max_{\theta\in\Theta}L(\theta)$

3. 定义：若 $L(\hat{\theta};x_1,\cdots,x_n)=\max_{\theta\in\Theta}L(\theta;x_1,\cdots,x_n)$，则称 $\hat{\theta}(x_1,\cdots,x_n)$ 为参数 θ 的最大似然估计值，相应的 $\hat{\theta}(X_1,\cdots,X_n)$ 称为参数 θ 的最大似然估计量.

4. 对数似然估计法：使对数似然函数 $\ln L(\theta)$ 最大的 $\hat{\theta}$ 也使似然函数 $L(\theta)$ 最大，因此寻找最大值可以从定义出发，也可以对 $\ln L(\theta)$ 微分(且更为常用). 步骤如下：

（1） 写出似然函数 $L(\theta)=L(\theta;x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}p(x_i;\theta)$ 或 $\prod_{i=1}^{n}f(x_i;\theta)$

（2） 取对数：$\ln L(\theta)=\sum_{i=1}^{n}\ln p(x_i;\theta)$ 或 $\ln L(\theta)=\sum_{i=1}^{n}\ln f(x_i;\theta)$

（3） 求导：$\dfrac{\mathrm{d}\ln L(\theta)}{\mathrm{d}\theta}$，并令 $\dfrac{\mathrm{d}\ln L(\theta)}{\mathrm{d}\theta}=0$，解方程即得最大似然估计值 $\hat{\theta}$.（注意要检验 $\dfrac{\mathrm{d}^2\ln L(\theta)}{\mathrm{d}\theta^2}$ 的符号）

5. 最大似然估计的不变性：若 $\hat{\theta}$ 为 θ 的最大似然估计，则对任一函数 $g(\theta)$，$g(\hat{\theta})$ 为其最大似然估计.

6. 注意：待估参数的最大似然估计不一定存在，即使存在也不一定惟一.

7. 多个未知参数的情形：

（1）未知参数可以不止一个. 若 $\theta_1,\theta_2,\cdots,\theta_k$ 均为未知参数，则定义似然函数为

$$L(\theta_1,\cdots,\theta_k)=L(\theta_1,\cdots,\theta_k;x_1,\cdots,x_n)=\prod_{i=1}^{n}f(x_i;\theta_1,\cdots,\theta_k)$$

（2）对于某组给定的样本值 $x_1,x_2,\cdots,x_n$，参数 $\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_k$ 使似然函数 $L(\theta_1,\cdots,\theta_k)$ 取得最大值，即 $L(\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_k)=\max_{(\theta_1,\theta_2,\cdots,\theta_k)\in\Theta}L(\theta_1,\theta_2,\cdots,\theta_k)$，则称 $\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_k$ 为 $\theta_1,\theta_2,\cdots,\theta_k$ 的极大似然估计值.

（3）若 $L(\theta_1,\cdots,\theta_k)$ 关于 $\theta_1,\theta_2,\cdots,\theta_k$ 可微，则称 $\dfrac{\partial L}{\partial \theta_i}=0,\ \ i=1,2,\cdots,k$ 为似然方程组.

解之可得未知参数的极大似然估计值 $\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_k$，然后再求得极大似然估计量.

8. 注意：若 $L(\theta_1,\cdots,\theta_k)$ 不是 $\theta_1,\cdots,\theta_k$ 的可微函数，则需要用其他方法求极大似然估计值.

9. 在统计问题中，往往先使用最大似然估计法，在最大似然估计法使用不便时，再用矩估计法.

估计量的评选标准

1. 常用标准：
   （1） 无偏性
   （2） 有效性
   （3） 一致性（相合性）
2. 无偏性：

   （1） 定义：若估计量 $\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$ 的数学期望 $E(\hat{\theta})$ 存在，且对任意的 $\theta\in\Theta$，有 $E(\hat{\theta})=\theta$，则称 $\hat{\theta}$ 是 θ 的无偏估计量.

   （2） 结论：不论总体服从什么分布，样本 k 阶矩 $A_k=\dfrac{1}{n}\sum_{i=1}^{n}X_i^k$ 都是总体 k 阶矩 μ_k 的无偏估计.
3. 有效性：

   (1) 定义：设 $\hat{\theta}_1=\hat{\theta}_1(X_1,X_2,\cdots,X_n)$，$\hat{\theta}_2=\hat{\theta}_2(X_1,X_2,\cdots,X_n)$ 均为 θ 的无偏估计量，若对任意的 $\theta\in\Theta$，均有 $D(\hat{\theta}_1)\leqslant D(\hat{\theta}_2)$，且至少有一个 $\theta\in\Theta$ 使得不等式严格成立，则称 $\hat{\theta}_1$ 较 $\hat{\theta}_2$ 有效.

   (2) 结论：算数均值比加权均值更有效. 例如取总体 X 的一个样本 X_1,X_2,X_3，则 $\hat{\mu}_1=\dfrac{1}{3}X_1+\dfrac{1}{3}X_2+\dfrac{1}{3}X_3$ 与 $\hat{\mu}_2=\dfrac{1}{6}X_1+\dfrac{1}{3}X_2+\dfrac{1}{2}X_3$ 均为总体期望的无偏估计，但 $\hat{\mu}_1$ 比 $\hat{\mu}_2$ 更有效.
4. 一致性(相合性)：

   (1) 定义：设 $\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$ 是参数 θ 的估计量. 若对任意的 $\theta\in\Theta$，$\hat{\theta}$ 依概率收

敛于θ，即$\forall\varepsilon>0$，$\lim\limits_{n\to\infty}P(|\hat{\theta}-\theta|\geqslant\varepsilon)=0$，则称$\hat{\theta}$为$\theta$的一致（或相合）估计量. 相合性被认为是估计量的一个最基本要求.

(2) 判定定理:

定理一：设$\hat{\theta}=\hat{\theta}(x_1,x_2,\cdots,x_n)$是参数$\theta$的一个估计量，若$\lim\limits_{n\to\infty}E(\hat{\theta})=\theta$，$\lim\limits_{n\to\infty}D(\hat{\theta})=0$，则$\hat{\theta}$为$\theta$的一致估计量.

定理二：若$\hat{\theta}_{n1},\cdots,\hat{\theta}_{nk}$分别为$\theta_1,\cdots,\theta_k$的一致估计，$\eta=g(\theta_1,\cdots,\theta_k)$是$\theta_1,\cdots,\theta_k$的连续函数，则$\hat{\eta}_n=g(\hat{\theta}_{n1},\cdots,\hat{\theta}_{nk})$为$\eta$的一致估计.

(3) 常用结论:

(i) 样本k阶矩是总体k阶矩的一致估计量；

(ii) 设$\hat{\theta}$是θ的无偏估计量，且$\lim\limits_{n\to\infty}D(\hat{\theta})=0$（其中$n$为样本容量），则$\hat{\theta}$是$\theta$的一致估计量.

5. 均方误差:

(1) 定义：设$\hat{\theta}$为参数θ的一个估计，则称$\mathrm{MSE}(\hat{\theta})=E(\hat{\theta}-\theta)^2$为$\hat{\theta}$的均方误差.

(2) 由于$\mathrm{MSE}(\hat{\theta})=E(\hat{\theta}-\theta)^2=D(\hat{\theta})+[E(\hat{\theta})-\theta]^2$，所以均方误差较小不仅意味着$\hat{\theta}$的方差较小，也意味着$\hat{\theta}$与$\theta$的偏差较小. 因此均方误差是评价点估计的最一般标准.

充分统计量与充分性原则

1. 充分统计量:

(1) 定义：设$x_1,x_2,\cdots,x_n$为来自总体X的样本. 总体X的分布函数为$F(x;\theta)$. 统计量$T=T(x_1,x_2,\cdots,x_n)$称为θ的充分统计量（同时也称为该分布的充分统计量），如果在给定T的取值后，$x_1,x_2,\cdots,x_n$的条件分布与待估参数θ无关，即

$P(X_1=x_1,X_2=x_2,\cdots,X_n=x_n|T=t)$或$f(x_1,x_2,\cdots,x_n|T=t)$与$\theta$无关.

(2) 由上述定义知，只有统计量T变了，才能对θ产生影响，即是说θ的所有信息都被T控制，这就是充分性的含义.

(3) 因子分解定理：设总体概率函数为$f(x;\theta)$，$X_1,X_2,\cdots,X_n$为来自总体的一个样本，

则 $T = T(X_1, X_2, \cdots, X_n)$ 为充分统计量的充要条件为：存在两个函数 $g(t;\theta)$ 与 $h(x_1, x_2, \cdots, x_n)$，使得 $f(x_1, x_2, \cdots, x_n;\theta) = g\left(T(x_1, x_2, \cdots, x_n);\theta\right)h(x_1, x_2, \cdots, x_n)$.

其中 $g(t;\theta)$ 是通过统计量 T 的取值 t 而依赖于样本的函数.

2. 充分性原则：

(1) 定理(Rao-Blackwell)：设总体概率函数为 $f(x;\theta)$，$x_1, x_2, \cdots, x_n$ 为其样本，$T = T(x_1, x_2, \cdots, x_n)$ 为 θ 的充分统计量, 则对 θ 的任一无偏估计 $\hat{\theta} = \hat{\theta}(x_1, x_2, \cdots, x_n)$，令 $\tilde{\theta} = E(\hat{\theta} \mid T)$，那么 $\tilde{\theta}$ 也是 θ 的无偏估计, 且 $D(\tilde{\theta}) \leqslant D(\hat{\theta})$.

(2) 说明：

(i) 好的无偏估计都是充分统计量的函数；

(ii) 如果无偏估计不是充分统计量的函数, 则将其对充分统计量求条件期望可以得到一个新的无偏估计, 且新估计的方差不超过原估计的方差；

(iii) 考虑 θ 的估计时, 只需要在基于充分统计量的函数中进行即可, 该说法对所有统计推断都是正确的, 这便是充分性原则.

第八章 假设检验

基本思想与概念

1. $\begin{cases} \text{对参数一无所知} \Longrightarrow \text{用参数估计的方法处理；} \\ \text{对参数有所了解, 但有猜测怀疑需要证实之时} \Longrightarrow \text{用假设检验的方法处理.} \end{cases}$

2. 含义：假设检验就是根据样本对所提出的假设做出判断, 是接受还是拒绝.

3. 依据：小概率事件在一次试验中很难发生.

4. 常用的假设检验问题(其中 θ_0 为已知常数)：

(1) 原假设 $H_0: \theta \leqslant \theta_0$，备择假设 $H_1: \theta > \theta_0$；(单侧检验)

(2) 原假设 $H_0: \theta \geqslant \theta_0$，备择假设 $H_1: \theta < \theta_0$；(单侧检验)

(3) 原假设 $H_0: \theta = \theta_0$，备择假设 $H_1: \theta \neq \theta_0$.(双侧检验)

5. 检验统计量：用于对原假设 H_0 做出判断的统计量称为检验统计量.

6. 拒绝域：使原假设被拒绝的样本观测值所在区域称为拒绝域，拒绝域的边界点称为临界点.
7. 注意：假设检验的着力点不在于说明原假设正确，而在于说明原假设不正确. 因此建立检验的拒绝域是重要的.
8. 两类错误：因为假设检验的依据是小概率事件在一次试验中很难发生,但很难发生不等于不发生,因此假设检验所作出的结论有可能是错误的.这种错误有两类:

   (1) 原假设 H_0 正确,但被拒绝,称为第一类错误.犯第一类错误的概率是显著性水平 α.

   (2) 原假设 H_0 错误,但被接受,称为第二类错误.犯第二类错误的概率常记为 β.

   注意：在样本容量 n 一定的情况下,犯两类错误的概率不能同时减小!(就像跷跷板一样)
9. 两类错误的控制准则：首先控制第一类错误,使其犯错概率不超过 α.然后若有必要,通过增大样本容量的方法来减少 β.
10. 显著性检验：只对犯第一类错误的概率加以控制,不考虑犯第二类错误的概率,这种检验称为**显著性检验**.
11. 单侧检验:

    (1) 左侧检验：$H_0:\mu\geqslant\mu_0$, $H_1:\mu<\mu_0$ (临界值与拒绝域均在左侧),也称下限检验.

    (2) 右侧检验：$H_0:\mu\leqslant\mu_0$, $H_1:\mu>\mu_0$ (临界值与拒绝域均在右侧),也称上限检验.
12. 假设检验的**一般步骤**:

    (1) 提出原假设 H_0 与备择假设 H_1;

    (2) 给定显著性水平 α 以及样本容量 n;

    (3) 确定检验统计量以及拒绝域形式;

    (4) 按 $P\{H_0\text{为真拒绝}H_0\}\leqslant\alpha$ 求出拒绝域

    (5) 取样，根据样本观测值确定接受还是拒绝 H_0;

    (6) 回到原问题作答.
13. 通常把有把握,有经验的结论作为原假设,或者尽可能使后果严重的错误成为第一类错误.即**原假设不会轻易被拒绝,希望得到的结论不会轻易被接受**.

正态总体的假设检验

1. 结论：对单个正态总体 $N(\mu,\sigma^2)$ 而言, 当显著性水平为 α 时，检验问题

(1) $H_0: \mu = \mu_0$, $H_1: \mu > \mu_0$

(2) $H_0: \mu \leqslant \mu_0$, $H_1: \mu > \mu_0$

有相同的拒绝域(尽管 H_0 的形式不同, 实际意义也不同). 因此第二类形式的问题可归结为第一类形式的讨论. 该结论对另一边的单侧检验也成立.

2. 单个正态总体 $N(\mu,\sigma^2)$ 的均值 μ 的检验

检验方法	原假设 H_0	备择假设 H_1	检验统计量	拒绝域
U 检验法 (用于 σ^2 已知时)	$\mu = \mu_0$	$\mu \neq \mu_0$	$U = \dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$ （H_0 为真时服从 $N(0,1)$）	$\lvert U \rvert \geqslant z_{\frac{\alpha}{2}}$
	$\mu \leqslant \mu_0$	$\mu > \mu_0$		$U \geqslant z_\alpha$
	$\mu \geqslant \mu_0$	$\mu < \mu_0$		$U \leqslant -z_\alpha$
t 检验法 (用于 σ^2 未知时)	$\mu = \mu_0$	$\mu \neq \mu_0$	$T = \dfrac{\bar{X} - \mu_0}{S / \sqrt{n}}$ （H_0 为真时服从 $t(n-1)$）	$\lvert T \rvert \geqslant t_{\frac{\alpha}{2}}$
	$\mu \leqslant \mu_0$	$\mu > \mu_0$		$T \geqslant t_\alpha$
	$\mu \geqslant \mu_0$	$\mu < \mu_0$		$T \leqslant -t_\alpha$

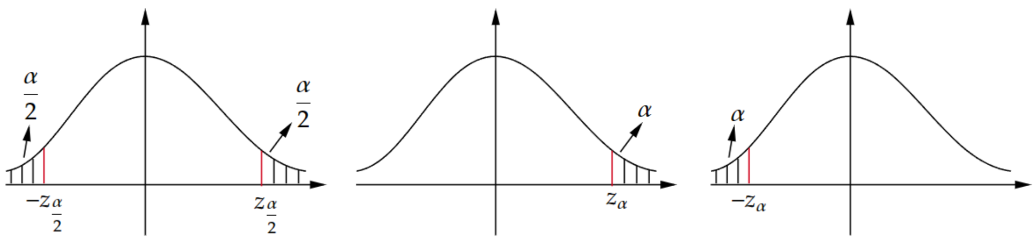

图 1 U 检验法的三种拒绝域(t 检验法的拒绝域同理)

3. 单个正态总体 $N(\mu,\sigma^2)$ 的方差 σ^2 的检验

检验方法	原假设 H_0	备择假设 H_1	检验统计量	拒绝域
χ^2 检验法 (用于 μ 未知时)	$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\chi^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$ （H_0 为真时服从 $\chi^2(n-1)$）	$\chi^2 \leqslant \chi^2_{1-\frac{\alpha}{2}}(n-1)$ 或 $\chi^2 \geqslant \chi^2_{\frac{\alpha}{2}}(n-1)$
	$\sigma^2 \geqslant \sigma_0^2$	$\sigma^2 < \sigma_0^2$		$\chi^2 \leqslant \chi^2_{1-\alpha}(n-1)$
	$\sigma^2 \leqslant \sigma_0^2$	$\sigma^2 > \sigma_0^2$		$\chi^2 \geqslant \chi^2_\alpha(n-1)$

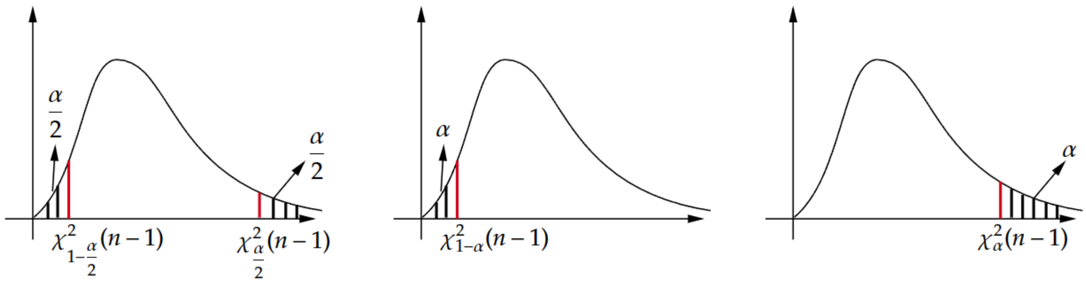

图 2 χ^2 检验法的三种拒绝域