“概率论与数理统计”复习提纲

一、古典概率

1. 样本空间的定义，样本点的定义，随机事件的定义，特殊的随机事件：不可能事件$\emptyset$和必然事件Ω

   注:样本空间不是唯一的，随机事件的表示依赖于样本空间的选择；选择样本空间的标准是该样本空间足以表示我们想要考察的随机事件。

2. 随机事件的关系和运算

   (1) 并、交、补：“或”、“且”、”否”， $A\bigcup B$、$A\bigcap B=AB$ 、$\bar{A}$; 差事件 $A-B=A\ B$;

   (2) 蕴含、相等、互斥、对立：$A\subseteq B$、$A=B$、$AB=\emptyset$、$A=\bar{B}$;

   (3) 不交并、包含差：$A+B$、$A-B$.

3. 集合运算的一些法则：集合运算的优先级为”逆¿交¿并”

   交换律，结合律，分配律，对偶律

4. 随机事件的集合表示

5. 古典概型和几何概型

   (1) 古典概型的定义：只有有限个基本随机事件，且每个基本随机事件发生的可能性相同

   (2) 计算公式：$P(A)=\dfrac{|A|}{|\Omega|}$

   注 古典概型的概率计算归结为计数，要用到排列和组合的知识。

   (3) 抽样和分组

   1) 设有N件产品，其中有M件次品，抽样n次，每次放回，则抽样总数为$|\Omega|=N^n$;抽到r件次品的情形共有$C_N^r M^r(N-M)^{n-r}$;

   2) 假设同上，若抽取的产品不放回，则称为无放回抽样，不考虑抽样的次序，则抽样总数为$|\Omega|=C_N^n$;抽到r件次品的情形共有$C_M^r C_{N-M}^{m-r}$;

   注: 将n件物品分成r组，每组人数分别为$n_1,n_2,\cdots,n_r$,其中$n_1+n_2+\cdots+n_r=n$,则全部分组的方式共有$\left(\begin{array}{c} n \\ n_1,n_2,\cdots,n_r \end{array}\right)=\dfrac{n!}{n_1!n_2!\cdots n_r!}$.

6. 几何概型：样本空间Ω为平面内可求面积的区域，假设概率和面积成正比，则$P(A)=\dfrac{S(A)}{S(\omega)}$.

7. 概率计算的一般原理

   (1) 概率的性质：非负性、规范性、可加性、可数可加性

   (2) 概率的加法原理

   1) 两个随机事件的并 $P(A\bigcup B)=P(A)+P(B)-P(AB)$

   2) 三个随机事件的并 $P(A\bigcup B\bigcup C)=P(A)+P(B)+P(C)-P(AB)-P(AC)-P(BC)+P(ABC)$

   3) 有限个随机事件的并(容斥原理)

   4) 特殊情况：$P(\bar{A})=1-P(A)$,

   若$B\subseteq A$, 则$P(A-B)=P(A)-P(B)$

8. 乘法原理和条件概率

(1) 条件概率的定义：$P(B|A)=\dfrac{P(AB)}{P(A)}$表示在事件A发生的条件下事件B发生的概率；

若A,B独立，则$P(AB)=P(A)P(B)$; 若A,B依赖，则$P(AB)=P(A)P(B|A)=P(B)P(A|B)$

(2) 两个事件独立性的定义：$P(B|A)=P(B)$、$P(A|B)=P(A)$、$P(AB)=P(A)P(B)$

注:以$P(B|A)=P(B)$作为独立性的定义，需要A的概率大于0.

注:条件概率仍然是一个概率.

(a) 非负性：$P(B|A)\geq 0$;

(b) 规范性：$P(\Omega|A)=1$;

(c) 可加性：$P(B_1+B_2|A)=P(B_1|A)+P(B_2|A)$;可数可加性类似可得.

(3)三个事件的独立性： 若A,B,C两两独立,且$P(ABC)=P(A)P(B)P(C)$,则称A,B,C相互独立.

(4) 多个事件的独立性：若对任意$1\leq i_1<i_2<\cdots<i_k\leq n, k=2,3,\cdots,n$,

$$P(A_{i_1}A_{i_2}\cdots A_{i_k})=P(A_{i_1})P(A_{i_2})\cdots P(A_{i_k})$$

(5) 多个随机事件的乘法原理:

$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1})$$

当$A_1,A_2,\cdots A_n$相互独立时，其积事件的概率是概率的乘积.

9. 全概率公式和贝叶斯公式

(1)样本空间的划分：$A_1,A_2,\cdots,A_n$ 称为是一个划分，若

(a) $A_1,A_2,\cdots,A_n$ 两两不交；

(b) $A_1\cup A_2\cup\cdots\cup A_n=\Omega$。

注：所谓划分，又称“互不相交的完备组”，即指一组随机事件，其中有且仅有一个随机事件发生。

(2)全概率公式和贝叶斯公式：

1） 全概率公式：$P(B)=\sum_{i=1}^{n}P(BA_i)=\sum_{i=1}^{n}P(A_i)P(B|A_i)$

2） 贝叶斯公式：$P(A_j|B)=\frac{P(A_j)P(B|A_j)}{\sum_{i=1}^{n}P(A_i)P(B|A_i)}$。

二、一维随机变量的分布、及数字特征

1. 随机变量的定义与分类

2. 离散型随机变量的分布、数字特征和随机变量函数的分布

(1)离散型随机变量的分布律

1)离散型随机变量X的取值范围为$\{x_1,x_2,\cdots\}$, 则其分布律为$P(X=x_i)=p_i$

2)p_i的性质：(a)$p_i\geq 0$; (b) $\sum_i p_i=1$

(2)离散型随机变量的分布函数

1)离散型随机变量的分布函数的定义: $F(x)=P(X\leq x), P(a<X\leq b)=F(b)-F(a)$

2)分布函数的性质:

(a)$F(x)$是单调不减的函数;

(b)$F(-\infty)=0, F(\infty)=1$;

(c) $F(x)$ 是右连续的函数，即

$$F(x+)=\lim_{y>x, y\to x} F(y)=F(x)$$

注: 任意随机变量在一点的概率 $P(X=x)=P(X\le x)-P(X<x)=F(x)-F(x-)$；从而随机变量在 x 点的 概率为 0 当且仅当其分布函数在 x 是连续的；随机变量在所有点的概率为 0 当且仅当其分布函数在所有点都是连续的。 由此我们得到随机变量在任意区间的概率

$$P(a<X\le b)=F(b)-F(a),\quad P(a<X<b)=F(b-)-F(a)$$
$$P(a\le X\le b)=F(b)-F(a-),\quad P(a\le X<b)=F(b-)-F(a-)$$

(3)离散型随机变量的期望和方差

1)离散型随机变量的期望和方差的定义

(a)离散型随机变量的期望的定义为

$$EX=\sum_i x_i P(X=x_i)$$

(b)离散型随机变量的方差的定义为

$$Var(X)=E(X-EX)^2=EX^2-(EX)^2$$

(c)离散型随机变量的标准差的定义为

$$\sigma(X)=\sqrt{Var(X)}$$

(4)常见的离散型随机变量

1) $0-1$ 分布，只能取值为 0 或 1 的随机变量，其分布为

$$P(X=i)=p^i(1-p)^{1-i},\ \ i=0,1$$

其中 $0\le p\le 1$ 是未知参数，则称 X 服从参数为 p 的 $0-1$ 分布，记为 $X\sim B(1,p)$. $E(X)=p, Var(X)=p(1-p)$

2)二项分布：对一堆产品进行 n 次重复抽样，所得次品个数是随机变量 X，设总次品率为 $p=\frac{M}{N}$， 则

$$P(X=k)=C_n^k p^k(1-p)^{n-k},\ \ k=0,1,\cdots,n$$

称 X 服从参数为 n,p 的二项分布，n 是自然数，$0\le p\le 1$，记为 $X\sim B(n,p)$,$EX=np, Var(X)=np(1-p)$.

$0-1$ 分布 即为参数为 $1,p$ 的二项分布，因而使用了二项分布的记号.

注：二项分布的另一个模型是 n 次独立重复随机试验中成功的次数，每次的成功率为 p.

3)超几何分布：对一堆产品进行 n 次无放回抽样，所得次品个数记为 X，设总产品数为 N，总次品数为 M， 则

$$P(X=m)=\frac{C_M^m C_{N-M}^{n-m}}{C_N^n},\ \ n=0,1,2,\cdots$$

称 X 服从参数为 $N,M;n$ 的超几何分布，记为 $H(N,M;n)$, $EX=\dfrac{nM}{N}$.

注：有可能出现 $m>M$ 的情形，此时约定当 $m>M$ 时二项式系数 $C_M^m=0$ 即可。

注：由概率函数的性质，可得恒等式

$$C_N^n=\sum_{m=0}^{n} C_M^m C_{N-M}^{n-m}$$

注： 当 $N>>n$ 时，每一次抽样对概率的影响极小，因此可以视为重复抽样，由此我们有如下的近似， $H(N,M;n)\approx B(n,\frac{M}{N})$。

4) 泊松分布：泊松分布一般表示一段时间内事故发生的次数，一段时间内到某商店里购物的人数等，其分布律如下，

$$P(X=k)=e^{-\lambda}\frac{\lambda^k}{k!},\ \ k=0,1,2,\cdots$$

其中 $\lambda>0$，称为泊松分布的强度，记为 $X\sim P(\lambda)$，或者 $\pi(\lambda)$, $EX=\lambda, Var(X)=\lambda$

注： 当 n 比较大，而 np 比较适中时，$B(n,p)\approx P(np)$.

5) 几何分布：设有一随机试验，每次的成功率为 p，第一次成功时的试验次数为 X，则

$$P(X=k)=(1-p)^{k-1}p,\ \ k=1,2,3,\cdots$$

称 X 服从参数为 p 的几何分布，因为其概率函数构成一几何级数，记为 $X\sim G(p)$. $EX=\dfrac{1}{p}, Var(X)=\dfrac{1-p}{p^2}$

注： 服从几何分布的随机变量有如下性质：

$$P(X>k)=(1-p)^k p+(1-p)^{k+1}p+\cdots=(1-p)^k$$

注： 几何分布的无记忆性：几何分布具有如下性质，设 $X\sim G(p)$， 则

$$P(X>k+l|X>l)=P(X>k)$$

(5)离散型随机变量的函数

1) 离散型随机变量的函数的定义

若 X 是离散型随机变量，其概率函数为 $x_i\to p_X(x_i)$， $Y=g(X)$ 必然也是离散型随机变量，设 Y 的全部取值为 $y_1,y_2,\cdots$， 则 $P(Y=y_j)=P(g(X)=y_j=\sum_{g(x_i)=y_j}p_X(x_i)$。

2) 离散型随机变量的函数的期望

设 X 是离散型随机变量，$Y=g(X)$，则 $P(Y=y_j)=\sum_{g(x_i)=y_j}P(X=x_i)$，从而可计算 Y 的期望，

$$\begin{aligned}EY&=\sum_j y_j P(Y=y_j)=\sum_j\sum_{g(x_i)=y_j}y_j P(X=x_i)\\&=\sum_i g(x_i)P(X=x_i)\end{aligned}$$

3. 连续型随机变量的分布、数字特征和随机变量函数的分布

(1)连续性随机变量的概率密度函数

1)定义

随机变量 X 称为是连续型的，若存在一个函数 $f(x)$，使得

$$P(X \in A) = \int_A f(x)dx$$

特别地，$P(a < X < b) = \int_a^b f(x)dx$，其中 $f(x)$ 称为是 X 的密度函数。

2)性质

(a)$f(x) \geq 0$，(b)$\int_{-\infty}^{\infty} f(x)dx = 1$

(2)连续性随机变量的分布函数

1)定义

连续随机变量的分布函数 $F(x) = \int_{-\infty}^{x} f(s)ds$

注连续随机变量的分布函数是连续函数，因此连续随机变量在任何一点的概率为零；计算连续随机变量在一个区间的概率不需要考虑端点的情况。

2)性质

(a)$F(x)$是单调不减的函数; (b)$F(-\infty) = 0, F(\infty) = 1$; (c) $F(x)$是连续函数.

(3)连续性随机变量的数字特征

1) 连续性随机变量的期望

连续型随机变量取值在 x 附近的概率为 $f(x)dx$，因此连续型随机变量的期望定义为

$$EX = \int_{-\infty}^{\infty} xf(x)dx$$

2) 连续性随机变量的方差

$$Var(X) = \int_{-\infty}^{\infty}(x - EX)^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - (\int_{-\infty}^{\infty} xf(x)dx)^2$$

(4)常见的连续性随机变量

1) 均匀分布：通常为截断误差的分布或者确定范围情形下的先验分布，记为 $U[a,b]$，其密度函数为

$$f(x) = \begin{cases} C, & a < x < b \\ 0, & \text{其它} \end{cases}$$

注： 均匀分布即一维区间上的几何概型，服从均匀分布的随机变量取值在 A 上的概率就是 A 的长度除以区间的总长度。

注： 均匀分布的分布函数为

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases}$$

注： 均匀分布的期望$EX=\dfrac{a+b}{2}$, 方差$Var(X)=\dfrac{1}{12}(b-a)^2$.

2) 指数分布：指数分布通常描述的是电子元件的使用寿命的分布或者服务时间的分布，记为 $E(\lambda)$，其 密度函数为

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x\geq 0\\ 0, & x<0\end{cases}$$

其中 $\lambda>0$ 称为“速率”。

注： $\int_0^\infty \lambda e^{-\lambda x}dx=[-e^{-\lambda x}]|_0^\infty=1$，可得规范性；

注： 类似地，若 $x>0$，可计算概率 $G(x)=P(X>x)=e^{-\lambda x}$，该函数常称为尾概率函数， 指数分布因其尾概率函数是指数函数而得名。

注： 指数分布的分布函数为

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

注： 指数分布的无记忆性：$P(X>t+s|X>s)=P(X>t)$，和几何分布的无记忆性类似，指数分布的无记忆性对所有 的非负实数成立，具有这种无记忆性的分布只有指数分布，而几何分布仅对非负整数成立。

注： 指数分布的期望$EX=\dfrac{1}{\lambda}$, 方差$Var(X)=\dfrac{1}{\lambda^2}$.

3) 正态分布：正态分布是概率统计中最常用最重要的分布，又称高斯分布和“钟形”分布，其密度函数为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

记为 $N(\mu,\sigma^2)$。

注： $N(0,1)$ 通常称为标准正态分布，标准正态分布 的密度函数一般记成 $\phi(x)$，其分布函数常记为 $\Phi(x)$。

注：正态分布在一个区间内的概率可以通过标准正态分布的分布函数进行计算，

$$P(a<X<b)=\Phi(\frac{b-\mu}{\sigma})-\Phi(\frac{a-\mu}{\sigma})$$

注： 对于标准正态分布，利用其对称性，得$\Phi(-x)=1-\Phi(x)$

注： 标准正态分布的分布函数 $\Phi(x)$ 对于正的 x，可以通过标准正态分布表查到，对于负的 x，要利用上面的对称性来求其数值。

注： 求分布函数数值的反问题是求分布函数的分位点，设 X 是一个连续型随机变量，其分布函数为 $F(x)$， 对任意 $0<\alpha<1$，满足 $P(X>u)=1-F(u)=\alpha$ 的数称为随机变量 X 的或者分布 F 的上 α-分位点， 记为 X_α 或 F_α，类似可定义下 α-分位点，即满足 $P(X<u)=F(u)=\alpha$，显然下 α- 分位点 即为上 $1-\alpha$-分位点。分位点可理解为高考中的"分数线"，若有 15% 人录取一本专业，则一本分数线为上 0.15 分为点。

注：标准正态分布的上 α-分位点记为 u_α，其数值可从标准正态分布表差得， 比如欲求 $u_{0.05}$，则从标准正态分布表中查到数值 0.95，可得分位点的数值为 $u_{0.05}=1.65$。

注： 正态分布的期望是$EX=\mu$, 方差是$Var(X)=\sigma^2$.

(5)连续性随机变量的函数

1) 连续性随机变量的函数

由于密度函数没有和概率直接联系，因此求连续型随机变量函数的分布，一般先求分布函数。设 X 是连续型随机变量，密度为 $f_X(x)$，$Y=g(X)$ 是 X 的函数，求 Y 的分布。

$$\begin{aligned}F_Y(y)=&P(Y\le y)=P(g(X)\le y)\\=&P(X\in A_y)=\int_{A_y}f_X(x)dx\\f_Y(y)=&\frac{d}{dy}F_Y(y)\end{aligned}$$

注： 对于比较好的函数，我们可以直接给出随机变量函数的密度公式，

设 X 是连续型随机变量，值域为 (a,b)，a,b 可以为 $\pm\infty$，$Y=g(X)$，g 在 (a,b) 上是严格单调函数（可以单调增加、也可以单调递减），其反函数为 h，并且假设 g 是可导的，Y 的取值范围为 (c,d)，则

$$f_Y(y)=\begin{cases}f_X(h(y))|h'(y)|, & c<y<d\\0, & 其它\end{cases}$$

2) 连续性随机变量的函数的期望

$$EY=\int_{-\infty}^{\infty}g(x)f_X(x)dx$$

4. 期望的性质

1）$Ec=c$；

2）$E(aX+b)=aEX+b$；从而 $E(X-c)=EX-c$，特别地，$E(X-EX)=0$；

3）$E[f(X)+g(X)]=Ef(X)+Eg(X)$；比如，$E(X-c)^2=EX^2-2cEX+c^2$。

4）$X\ge Y$，则 $EX\ge EY$；若 $X\ge Y$，$EX=EY$，则 $P(X=Y)=1$。

5. 方差的性质

1）$Var(X)\ge 0$；若 $Var(X)=0$，则 $P(X=0)=1$；

2）$Var(aX+b)=a^2Var(X)$

注： 若随机变量 X 的期望、方差都存在，$X^\star=\frac{X-EX}{\sigma(X)}$ 称为是 随机变量 X 的标准化，容易得到 $EX^\star=0$，$Var(X^\star)=1$。

3）$E(X-c)^2=E[(X-EX)+(EX-c)]^2=Var(X)+(EX-c)^2\ge Var(X)$；

三、二维随机变量的分布、数字特征和随机变量的函数的分布

1. 二维随机变量的分布

(1)二维离散型随机变量的联合概率函数

$$P(X=x_i,Y=y_j)=p_{i,j}$$

注：给定联合概率函数，则

$$P(X \in A) = \sum_{(x_i,y_j)\in A} p_{i,j}$$

(2)二维离散型随机变量的边缘概率函数

X, Y 的边缘概率函数分别记为 $P(X = x_i) = p_{i\cdot}$，$P(Y = y_j) = p_{\cdot j}$， 有时候也直接记为 p_X, p_Y，则

$$p_{i\cdot} = P(X = x_i) = \sum_j P(X = x_i, Y = y_j) = \sum_j p_{i,j}$$
$$p_{\cdot j} = P(Y = y_j) = \sum_i P(X = x_i, Y = y_j) = \sum_i p_{i,j}$$

注：若概率函数以分布表形式给出，则边缘概率函数的数值分别的各行各列的数值之和，则边缘概率函数 可以在原分布表上增加一行一列给出。

(3) 二维离散型随机变量的条件分布

在 $X = x$ 条件 Y 的分布称为 Y 的条件分布，$X = x$ 条件下的条件概率函数通常记为

$$p_{Y|X}(y|x) = P(Y = y|X = x) = \frac{P(Y = y, X = x)}{P(X = x)}$$

在 $Y = y$ 条件下 X 的分布；若 X 的所有可能取值为 x_i，Y 的所有可能取值为 y_j，则

$$p_{Y|X}(y_j|x_i) = \frac{p_{i,j}}{p_{i\cdot}}$$
$$p_{X|Y}(x_i|y_j) = \frac{p_{i,j}}{p_{\cdot j}}$$

(4) 二维离散型随机变量的独立性

若随机变量 X 的取值对 Y 的取值没有影响，则称 X, Y 是独立的；此时，

$$\begin{aligned} &\forall\ i,j \quad p_{\cdot j} = p_{Y|X}(y_j|x_i) = \frac{p_{i,j}}{p_{i\cdot}} \\ \Leftrightarrow &\forall\ i,j \quad p_{i,j} = p_{i\cdot}\ p_{\cdot j} \\ \Leftrightarrow &\forall\ i,j \quad p_{i\cdot} = p_{X|Y}(x_i|y_j) = \frac{p_{i,j}}{p_{\cdot j}} \end{aligned}$$

注：若 X, Y 独立，则 $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$，即对任意 A, B， 随机事件 $\{X \in A\}$ 和随机事件 $\{Y \in B\}$ 是独立的。

注：若 X, Y 独立，则 X, Y 的联合概率函数可以由边缘概率函数的乘积得到。

注：若 X, Y 独立，其联合概率函数由分布表给出，则任意两行成比例，任意两列也成比例。

注：若 X, Y 独立，其联合概率函数由分布表给出，则任意矩形四个顶点构成的两条对角线数值的乘积相同。

(5) 二维离散型随机变量的联合分布函数

二维随机变量 (X, Y) 的联合分布函数为

$$F(x, y) = P(X \le x, Y \le y)$$

注： 若 $y_1 < y_2$，$F(x,y_2) - F(x,y_1) = P(X \le x, y_1 < Y \le y_2)$；同理，若 $x_1 < x_2$， 我们有 $F(x_2,y) - F(x_1,y) = P(x_1 < X \le x_2, Y \le y)$；由概率的非负性，联合分布函数关于 两个自变量都是单调不减的；

注： 若 $x_1 < x_2$，$y_1 < y_2$，利用上面的式子可得

$$P(x_1 < X \le x_2, y_1 < Y \le y_2) = F(x_2,y_2) - F(x_1,y_2) - F(x_2,y_1) + F(x_1,x_2) \ge 0$$

从而二维随机变量落在一个左开右闭矩形区域内的概率可以由其联合分布函数在该矩形四个顶点的数值 求得；

注： 和一维类似，联合分布函数关于两个变量都是右连续的；而联合分布函数的规范性则和 一维情形略有不同，

$$F(-\infty,y) = F(x,\infty) = 0, \quad \forall\, x,y \qquad\qquad F(\infty,\infty) = 1$$

两个变量中只要有一个为负无穷，则其值为 0；两个值同时为正无穷时，其值为 1。

(6) 二维离散型随机变量的边缘分布的分布函数

$$F(x,+\infty) = P(X \le x, Y < \infty) = F(X \le x) = F_X(x)$$
$$F(+\infty,y) = P(X < \infty, Y \le y) = F(Y \le y) = F_Y(y)$$

(7) 二维连续型随机变量的联合密度函数

一个二维随机变量 (X,Y) 称为是连续型的，若存在函数 $f(x,y)$，使得

$$P((X,Y) \in A) = \iint_A f(x,y)dxdy$$

$f(x,y)$ 称为 X,Y 的联合密度函数

注： 联合密度函数 $f(x,y)$ 非负，并且满足规范性：

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1$$

(8) 二维连续型随机变量的联合分布函数

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(s,t)dtds$$

注： 二维连续型随机变量 (X,Y) 的联合分布函数 $F(x,y)$，则其密度函数为

$$f(x,y) = \frac{\partial^2}{\partial x \partial y}F(x,y)$$

(8) 二维连续型随机变量的边缘分布函数与边缘密度函数

$$F_X(x) = F(x,\infty) = \int_{-\infty}^{x}\int_{-\infty}^{\infty} f(s,y)dyds$$
$$f_X(x) = \frac{d}{dx}F_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$$

$$F_Y(y) = F(\infty,y) = \int_{-\infty}^{y}\int_{-\infty}^{\infty} f(x,s)dxds$$
$$f_Y(y) = \frac{d}{dy}F_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

(9) 二维连续型随机变量的条件分布密度函数

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}$$
$$f_{Y|X}(y|x)=\frac{f(x,y)}{f_X(y)}$$

(10) 二维连续型随机变量的独立性

$$X,Y\ \text{独立}\Leftrightarrow f(x,y)=f_X(x)f_Y(y)$$

注：二维连续型随机变量是独立的，若其联合密度函数是变量可分离的；类似于分布函数， 该结论的逆也是成立的。因此连续型随机变量的独立性可直接根据联合密度函数的形式来判断。

注：对于具有两个变量的分区域函数 $f(x,y)$ 的变量分离性，必须同时考虑其区域和函数形式.

2. 二维随机向量的数字特征

(1) 二维随机向量函数的期望

设 (X,Y) 是二维随机向量，则

$$Eg(X,Y)=\begin{cases}\sum_i\sum_j g(x_i,y_j)P(X=x_i,Y=y_j), & \text{离散型}\\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}g(x,y)f(x,y)dxdy\end{cases}$$

注： 边缘 X,Y 的期望和方差

$$EX=\begin{cases}\sum_i\sum_j X_iP(X=x_i,Y=y_j), & \text{离散型}\\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xf(x,y)dxdy\end{cases}$$

$$EY=\begin{cases}\sum_i\sum_j y_jP(X=x_i,Y=y_j), & \text{离散型}\\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}yf(x,y)dxdy\end{cases}$$

期望和方差的性质：

(a) $E(aX+bY)=aEX+bEY$；

(b) 若 X,Y 独立，则 $E(XY)=EX\ EY$；

(c) $Var(X+Y)=Var(X)+Var(Y)+2E(X-EX)(Y-EY)$；特别地， 若 X,Y 独立，$Var(X+Y)=Var(X)+Var(Y)$.

(2)协方差和相关系数

1) 协方差：$Cov(X,Y)=E(X-EX)(Y-EY)$ 称为随机变量 X,Y 的协方差。

(a) 计算公式：$Cov(X,Y)=E(XY)-EXEY$；

(b) 和自身的协方差：$Cov(X,X)=Var(X)$；

(c) 对称双线性：$Cov(X,Y)=Cov(Y,X)$；$Cov(aX+bY,Z)=aCov(X,Z)+bCov(Y,Z)$；

(d) 利用对称双线性性：

$Cov(X+Y,X-Y)=Var(X)-Var(Y)$；

$Cov(X+Y,X+Y)=Var(X)+Var(Y)+2Cov(X,Y)$；

$Cov(X-Y,X-Y)=Var(X)+Var(Y)-2Cov(X,Y)$；

(e) $[Cov(X,Y)]^2 \le Var(X)Var(Y)$；特别地，$Cov(X,c)=0$；

2) 相关系数：$\rho(X,Y)=\rho_{X,Y}=\frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$

(a) 相关系数是无量纲量，$|\rho(X,Y)| \le 1$ 反映了 X,Y 之间线性关系的程度；

(b) 若 $\rho(X,Y)=\pm 1$，则 $f(t)=Var(X+tY)=0$ 有唯一解，因此 $X+tY=c$，Y,X 之间是线性关系；

(c) 若 $\rho(X,Y)=0$，则 X,Y 之间无线性关系，此时称 X,Y 不相关。

注： X,Y 不相关 $\Leftrightarrow Cov(X,Y)=0 \Leftrightarrow EXY=EXEY \Leftrightarrow Var(X+Y)=Var(X)+Var(Y) \Leftrightarrow Var(X+Y)=Var(X-Y)$

注： 若 X,Y 独立，则 X,Y 不相关；反过来，则不一定成立。

注： 若 X,Y不相关且 X,Y服从二点分布，则X,Y 独立.

注： 若 (X,Y)服从二维正态分布且不相关，则X,Y 独立.

3. 常见的二维随机变量

(1) 二维均匀分布：$U(\Omega)$，Ω 是平面上可求面积的区域，二维均匀分布的密度函数为

$$f(x,y)=\begin{cases} C, & (x,y)\in\Omega \\ 0, & (x,y)\notin\Omega \end{cases}$$

注： 记 $S(A)$ 为区域 A 的面积，则由规范性

$$1=\iint_{\Omega} C dxdy = CS(\Omega) \Rightarrow C=\frac{1}{S(\Omega)}$$

注： 若 $A\subseteq\Omega$，则

$$P((X,Y)\in A)=\iint_A C dxdy=\frac{S(A)}{S(\Omega)}$$

二维均匀分布实际上就是几何概型。

(2) 二维正态分布：二维正态分布记为 $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2;\rho)$，其联合密度函数

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

1) 边缘分布：$X\sim N(\mu_1,\sigma_1^2)$，类似可得 $Y\sim N(\mu_2,\sigma_2^2)$

注： 二维正态分布的两个边缘分布都是正态分布，其参数由 二维正态分布的前四个参数决定。另一个参数 ρ 和边缘分布无关，这也说明了 联合分布决定边缘分布，但反过来并不正确.

2) 条件分布：二维正态分布的条件分布也是正态分布。

$$\begin{aligned} f_{Y|X}(y|x) &= \frac{f(x,y)}{f_X(x)} \\ &= \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2}\frac{1}{1-\rho^2}[\frac{y-\mu_2}{\sigma_2}-\rho\frac{x-\mu_1}{\sigma_1}]^2} \end{aligned}$$

注： 由独立性的定义，二维正态随机变量 (X,Y) 独立当且仅当 $\rho=0$。

4. 二维随机变量的函数的分布

(1) 离散型随机变量函数的分布

设二维随机变量 (X,Y) 的联合概率函数为 $p_{i,j}=P(X=x_i,Y=y_j)$， $Z=g(X,Y)$，要求 Z 的概率函数，先找出 Z 能够取到的所有数值，记为 z_k，则

$$P(Z=z_k)=P(g(X,Y)=z_k)=\sum_{g(x_i,y_j)=z_k} g_{i,j}$$

注： 若二维随机变量 (X,Y) 的概率函数以分布表形式给出，则可以直接在表格上计算。

1)二维离散型随机变量函数和的分布

$$P(Z=z_k)=P(X+Y=z_k)=\sum_{i)} P(x=x_i,Y=z_k-x_i)=\sum_{j} P(x=z_k-y_j,Y=y_j)$$

注： $X\sim B(n,p),Y\sim B(m,p)$，$X,Y$ 独立，则$X+Y\sim B(m+n,p)$

特别地， $X_i\sim B(1,p),i=1,\cdots,n$且$X_i(i=1,\cdots,n)$相互独立，则$X_1+X_2+\cdots+X_n\sim B(n,p)$

注： $X\sim P(\lambda),Y\sim P(\mu)$，$X,Y$ 独立，$Z=X+Y$，则$X+Y\sim P(\lambda+\mu)$

2)二维离散型随机变量极大极小值的分布

已知二维离散型随机变量 (X,Y)，其概率分布函数是$F(x,y)$

$$F_{max}(z)=P(Z_{max}\leq z)=P(X\leq z,Y\leq z)=F(z,z)$$
$$F_{min}(z)=1-P(X>z,Y>z)=F_X(z)+F_Y(z)-F(z,z)$$

特别地，当X,Y相互独立时，

$$F_{max}(z)=P(Z_{max}\leq z)=P(X\leq z,Y\leq z)=F(z,z)=F_X(z)F_Y(z)$$
$$F_{min}(z)=1-P(X>z,Y>z)=1-[1-F_X(z)][1-F_Y(z)]$$

(2) 连续型随机变量函数的分布：二维连续型随机变量的函数比一维要复杂得多，因此一般的函数很难求其分布，我们考虑比较简单的函数.

1) 二维连续性随机变量和函数的分布

已知二维连续型随机变量 (X,Y)，其联合密度函数为 $f(x,y)$，$Z=X+Y$， Z 的分布函数，

$$\begin{aligned}F_Z(z)&=P(Z\leq z)=P(X+Y\leq z)\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f(x,y)dydz=\int_{-\infty}^{\infty}\int_{-\infty}^{z-y} f(x,y)dxdz\\f_Z(z)&=\frac{d}{dz}F_Z(z)=\int_{-\infty}^{\infty} f(x,z-x)dx=\int_{-\infty}^{\infty} f(z-y,y)dy\end{aligned}$$

注： $X+Y=z$ 是平面上的一条直线，$f_Z(z)dz$ 是 (X,Y) 在这条直线附近的概率，因此 $f_Z(z)$ 之和联合密度函数在这条直线上的数值有关；

注： 若 X,Y 独立，则 $f(x,y)=f_X(x)f_Y(y)$，此时有

$$f_Z(z)=\int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx$$

此时 f_Z 称为是 f_X 和 f_Y 的卷积。

注： $X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$，$X,Y$ 独立，则 $X+Y\sim N(\mu_1+\mu_2,\sigma_1^2,\sigma_2^2)$；

注： (X,Y) 服从二维正态分布，则 $aX+bY$ 也是服从正态分布；

注： (X,Y) 服从二维正态分布，则 $(aX+bY,cX+dY)$ 也服从二维正态分布。

2)二维连续型随机变量极大极小值的分布

随机变量 X,Y 相互独立，分布函数分别为 F_X,F_Y，密度函数分别为 f_X,f_Y，则

$$\begin{aligned}F_{max}(z)&=P(Z_{max}\le z)=P(X\le z,Y\le z)=F_X(z)F_Y(z)\\ f_{max}(z)&=f_X(z)F_Y(z)+F_X(z)f_Y(z)\\ 1-F_{min}(z)&=P(X>z,Y>z)=(1-F_X(z))(1-F_Y(z))\\ f_{min}(z)&=f_X(z)[1-F_Y(z)]+f_Y(z)[1-F_X(z)]\end{aligned}$$

设 $X_1,X_2,\cdots,X_n$ 是独立同分布的随机变量，共同的分布函数和密度函数为 F 和 f，则

$$\begin{aligned}F_{max}(z)&=P(X_1\le z,\cdots,X_n\le z)\\ &=P(X_1\le z)\cdots P(X_n\le z)=[F(z)]^n\\ f_{max}(z)&=n[F(z)]^{n-1}f(z)\\ 1-F_{min}(z)&=P(X_1>z,\cdots,X_n>z)\\ &=P(X_1>z)\cdots P(X_n>z)=[1-F(z)]^n\\ f_{min}(z)&=n[1-F(z)]^{n-1}f(z)\end{aligned}$$

四、大数定律

1. 大数定律

(1) 依概率收敛：若对任意 $\epsilon>0$，

$$\lim_{n\to\infty}P(|X_n-c|<\epsilon)=1$$

或者等价地，$\lim_{n\to\infty}P(|X_n-c|>\epsilon)=0$，称随机变量序列 $\{X_n\}$ 依概率收敛于常数 c, 记为 $X_n\xrightarrow{P}c$。

注:依概率收敛要比通常的收敛性弱.

注： 若 X_n 依概率收敛于 c，g 是连续函数，则 $g(X_n)$ 依概率收敛于 $g(c)$；若 $X_n\xrightarrow{P}a$，$Y_n\xrightarrow{P}b$，g 是二元连续函数，则 $g(X_n,Y_n)\xrightarrow{P}g(a,b)$。

(2)伯努利大数定律：

用 n_A 表示 n 次独立重复试验中随机事件 A 发生的次数，则

$$f_n(A)=\frac{n_A}{n}\xrightarrow{P}P(A)$$

注:伯努利大数定理可以叙述为：频率依概率收敛于概率。从而概率可以视为频率的极限，而频率可以视为概率的近似。

(3)辛钦大数定律： 设 $X_1, X_2, \cdots X_n$ 是独立同分布的随机变量序列，$EX_1 = \mu, Var(X_1) = \sigma^2$ 存在，则

$$\frac{1}{n}(X_1 + X_2 + \cdots + X_n) \xrightarrow{P} \mu$$

注: 在伯努利试验中，令 $X_i = \begin{cases} 1, & \text{第 } i \text{ 次试验中 } A \text{ 发生} \\ 0, & \text{第 } i \text{ 次试验中 } A \text{ 不发生} \end{cases}$ ，则 $f_n(A) = \frac{1}{n}(X_1 + \cdots + X_n)$，从而辛钦大数定律是伯努利大数定律的推广。

注: $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ 是 n 次试验结果的平均值，这个随机变量通常称为 n 次试验的样本均值，从而辛钦大数定律可表述为：样本均值依概率收敛于均值。

注: 辛钦大数定律是最常见形式的大数定律，因此也常常直接称为大数定律或弱大数定律。

(4)切比雪夫不等式：伯努利大数定律的严格证明要归功于切比雪夫，主要工具是以其命名的不等式，

设随机变量 X 的期望 EX、方差 $Var(X)$ 存在，则对任意 $\epsilon > 0$，

$$P(|X - EX| \geq \epsilon) \leq \frac{Var(X)}{\epsilon^2}$$

或等价的，$P(|X - EX| < \epsilon) \geq 1 - \frac{Var(X)}{\epsilon^2}$。

(5)切比雪夫大数定律

设 $X_1, X_2, \cdots X_n$ 是两两不相关的随机变量序列，$EX_i = \mu_i, Var(X_1) = \sigma_i^2$ 都存在，且存在常数M, 使得$|\sigma_i^2| \leq M$ 恒成立，则对$\forall \varepsilon > 0$,

$$\frac{1}{n}(X_1 + X_2 + \cdots + X_n) \xrightarrow{P} \frac{1}{n}(\mu_1 + \mu_2 + \cdots + \mu_n)$$

注: 切比雪夫大数定律蕴含辛钦大数定律.

2. 中心极限定理

(1)依分布收敛：若 X_n 的分布函数 $F_n(x)$ 收敛于 X 的分布函数 $F(x)$，称随机变量序列 X_n 依分布收敛于随机变量 X，记为 $X_n \xrightarrow{d} X$。 **注:** 依分布收敛只和随机变量的分布有关，因此有时候我们也称 X_n 依分布收敛于 F；

(2)棣莫佛-拉普拉斯中心极限定理：

设 $X_1, \cdots, X_n, \cdots$ 是独立同分布随机变量序列，$X_1 \sim b(1, p)$，则

$$P(a < \frac{X_1 + \cdots + X_n - np}{\sqrt{np(1-p)}} < b) \to \Phi(b) - \Phi(a)$$

即 $\frac{X_1 + \cdots + X_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0, 1)$。

注: 作一个线性变换，$X_1 + \cdots + X_n$ 近似服从正态分布 $N(np, np(1-p))$，从而这是二项分布 $b(n, p)$ 的近似分布。二项分布还有一个近似分布泊松分布，它们的近似条件是不一样的，一般当 np 比较大时，我们将其近似为正态分布，而当 np 不太大，n 比较大时将其近似为泊松分布。

(3) 勒维中心极限定理：

设 $X_1, X_2, X_3, \cdots$ 是独立同分布的随机变量序列，$EX_1 = \mu, Var(X_1) = \sigma^2$ 存在，则

$$\frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{d} N(0, 1)$$

注： 注意到，$0-1$ 分布的期望和方差分别为 $p,p(1-p)$，从而勒维中心极限定理是棣莫佛-拉普拉斯 中心极限定理的推广；

注： 作一个变换，$X_1+\cdots+X_n$ 近似服从正态分布 $N(n\mu,n\sigma^2)$。 中心极限定理的实质是：任何形式随机性的大量累加都近似于正态分布。这也说明了中心极限定理和 正态分布的重要性。

五、统计基础

1. 统计的基本概念

(1)总体与样本、样本观测值

(2) 样本函数和统计量

(3) 常见统计量

1） 样本均值：$\overline{X}=\frac{1}{n}(X_1+\cdots+X_n)$，其观测值记为 $\overline{x}$；

2） 样本方差：$S^2=\frac{1}{n-1}[(X_1-\overline{X})^2+\cdots+(X_n-\overline{X})^2]$，其观测值记为 s^2； $S=\sqrt{S^2}$ 和 $s=\sqrt{s^2}$ 分别为样本标准差及其观测值；

3） 样本原点矩：$A_k=X_1^k+\cdots+X_n^k$，观测值记为 a_k；

4） 样本中心矩：$M_k=(X_1-\overline{X})^2+\cdots+(X_n-\overline{X})^2$，观测值为 m_k； **注：** 样本方差 S^2 不同于二阶样本中心矩 M_2，$S^2=\frac{n}{n-1}M_2$； **注：** 中心矩和原点矩之间的关系等同于样本中心矩和样本原点矩之间的关系。

2. 统计学分布

(1)标准正态分布

1） 记 Z_α 或 u_α 为标准正态分布的上 α-分位点，即若 $X\sim N(0,1)$，

$$P(X>Z_\alpha)=\alpha$$

2） 设 X 的密度函数是偶函数，则 $P(X>a)=P(X<-a)$。因此我们有 $Z_{1-\alpha}=-Z_\alpha$。

3） 设 $X\sim N(\mu,\sigma^2)$，$X_1,\cdots,X_n$ 是其样本，则

$$u=\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$$

(2) $\mathcal{X}^2$-分布

1） 设 $X_1,\cdots,X_n$ 独立同分布，均服从 $N(0,1)$，则

$$X_1^2+X_2^2+\cdots+X_n^2$$

的分布是自由度为 n 的 $\mathcal{X}^2$-分布，记为 $\mathcal{X}_n^2$。

2） 对 $n\le 2$，其密度函数是单调减少的；对 $n>2$，其密度函数是先增后减的单峰函数；$\mathcal{X}^2$- 分布 的上 α-分位点记为 $\mathcal{X}_n^2(\alpha)$。

3） 若 $X\sim\mathcal{X}_n^2$，$Y\sim\mathcal{X}_m^2$，X,Y 相互独立，则 $X+Y\sim\mathcal{X}_{n+m}^2$。

4） 若 $X\sim N(\mu,\sigma^2)$，$X_1,\cdots,X_n$ 是其样本，则

$$\mathcal{X}^2=\frac{\sum_{k=1}^n(X_k-\mu)^2}{\sigma^2}\sim\mathcal{X}_n^2$$

(3) $t-$分布

1）设 $X \sim N(0,1)$，$Y \sim \mathcal{X}_n^2$，则

$$t = \frac{X}{\sqrt{Y/n}}$$

服从自由度为 n 的 t-分布，记为 t_n。

2） t-分布的密度函数是关于 y-轴对称的，记 t-分布的分位点为 $t_n(\alpha)$，则 $t_n(1-\alpha) = -t_n(\alpha)$。

3）设 $X \sim t_1$，则 $\frac{1}{X} \sim t_1$。

4）设 $X \sim N(\mu, \sigma^2)$，$X_1, \cdots, X_n$ 是其样本，则其样本均值、样本方差

$$\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$$
$$S^2 = \frac{1}{n-1}[\sum_{k=1}^{n}(X_k - \overline{X})^2]$$

是相互独立的。

5）假设同上，则

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

(4) $F-$分布

1）设 $X \sim \mathcal{X}_{n_1}^2$，$Y \sim \mathcal{X}_{n_2}^2$，$X, Y$ 独立，则

$$\frac{X/n_1}{Y/n_2}$$

服从自由度为 (n_1, n_2) 的 F-分布，记为 F_{n_1,n_2}。

2） F-分布的分位点记为 $F_{n_1,n_2}(\alpha)$。

3） $X \sim t_n$，则 $X^2 \sim F(1,n)$；

4） $X \sim F_{n_1,n_2}$，则 $\frac{1}{X} \sim F_{n_2,n_1}$；

5） $F_{n_1,n_2}(1-\alpha) = \frac{1}{F_{n_2,n_1}}(\alpha)$.

六、参数估计

1. 矩估计：矩估计是参数点估计的一种方法，由 Pearson 提出，其利用样本矩估计总体的矩，其基本思想如下：

(1) 由大数定律，$\frac{1}{n}(X_1+X_2+\cdots+X_n) \xrightarrow{P} EX$，故样本均值 $\frac{1}{n}(X_1+\cdots+X_n)$ 可作为 EX 的估计；

(2) 同样由大数定律，可得 $\frac{1}{n}(X_1^k + X_2^k + \cdots + X_n^k) \xrightarrow{P} EX^k$，即 $A_k \xrightarrow{P} v_k$；

(3) 由依概率收敛的性质，设 g 是连续函数，则 $g(A_1, \cdots, A_k) \xrightarrow{P} g(v_1, \cdots, v_k)$。特别地，总体 X 的 k 阶中心矩可以用原点矩的连续函数表示，$u_k = g(v_1, \cdots, v_n)$，此时必有 $M_k = g(A_1, \cdots, A_k)$，故 $M_k \xrightarrow{P} u_k$。

(4)矩估计的计算方法

2. 极大似然估计

(1) 极大似然估计的思想：A 已发生，$P(A|\theta_1) > P(A|\theta_2)$，则认为 θ 更可能为 θ_1。

(2) 离散情形的似然函数：

$X \sim F(\theta)$ 为离散随机变量，$A = \{X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n\}$ 发生，则 $L(\theta) = P(A|\theta)$ 为离散情形的似然函数，即

$$L(\theta) = \Pi_{i=1}^{n} P(X_i = x_i|\theta)$$

则极大似然估计 $\tilde{\theta}$ 是 $L(\theta)$ 的最大值点。

(3) 连续情形的似然函数：

连续随机变量的情形，则用密度函数代替概率：

$$L(\theta) = \Pi_{i=1}^{n} f(x_i|\theta)$$

1)最大值点的求法（一）：求导数

(a) $X \sim P(\lambda)$；

(b) $X \sim e(\lambda)$；

2) 最大值点的求法（二）：单调性

(a) $X \sim U(0, \theta)$；

(b) $X \sim U(\theta, 2\theta)$；

3. 极大似然估计的问题：

(1) 不变性原理：若 $\tilde{\theta}$ 是 θ 的极大似然估计，则 $g(\tilde{\theta})$ 是 $g(\theta)$ 的极大似然估计。

(2) 唯一性：$X \sim U(\theta, \theta+1)$，一组观测值为 $(0.3, 0.7, 0.4, 0.8, 1.1, 0.5, 0.9)$。

4. 点估计的评价标准

(1) 均方误差：$E(\theta - \tilde{\theta})^2$，均方误差的分解：

(2) 无偏性：

1) 定义：

2) 期望和方差的无偏估计

3) 均匀分布的无偏估计

4) 期望的线性无偏估计

(3) 有效性：

1) 定义：

2) 期望的最优线性无偏估计

(4) 一致性：

1) 一致性的定义：

2) 期望和方差的一致估计；

3) 一致估计的判别法：

5. 渐近无偏性：$E(\tilde{\theta}) \to \theta$；

6. 渐近稳定性：$D(\tilde{\theta}) \to 0$。

七、区间估计与假设检验

1. 区间估计的定义：

   形如 $\hat{\theta}_1 < \theta < \hat{\theta}_2$ 的估计称为是区间估计，形如 $\theta < \tilde{\theta}_2$ 或 $\theta > \tilde{\theta}_1$ 的估计称为是单侧区间估计。$1-\alpha = P(\hat{\theta}_1 < \theta < \hat{\theta}_2)$ 称为是区间估计的置信度（系数、水平），置信水平 一般是人为确定的，置信区间和置信水平的选择有关，称为置信水平为 $1-\alpha$ 的（单侧）置信区间。$\hat{\theta}_1, \hat{\theta}_2$ 称为双侧置信下（上）限，$\tilde{\theta}_1, \tilde{\theta}_2$ 称为单侧置信下（上）限。

   注：置信水平的意义在于在参数的多次估计中，所得置信区间能够覆盖真值的比例约为 $1-\alpha$；其表示的是估计 方法的可靠性，不代表单次估计结果的可靠性。

2. 区间估计的枢变量法：

   (1) U 是除待估计参数 θ 外，没有其它未知参数的样本函数；

   (2) U 的分布已知，设其上下 $\frac{\alpha}{2}$ 分位点为 u_2, u_1；

   (3) U 关于 θ 是单调增（减）的。

   满足上述条件的 U 称为是枢变量，则

$$
\begin{aligned}
1-\alpha &= P(u_1 < U(\theta) < u_2) \\
&= P(\hat{\theta}_1 < \theta < \hat{\theta}_2)
\end{aligned}
$$

   则 $(\hat{\theta}_1, \hat{\theta}_2)$ 是求得的置信水平为 $1-\alpha$ 的置信区间。

3. 区间估计的解题规范：

   (1) 写出枢变量及其分布；

   (2) 写出求得的置信区间；

   (3) 代入数据计算。

4. 假设检验：所谓假设检验，即根据试验数据对总体分布的某个结论作出判断，该结论通常记为 H_0，称为原假设，H_0 的 对立面记为 H_1，称为备择假设。假设检验的常见形式为，H_0 为 $\theta = \theta_0$，H_1 为 $\theta \neq \theta_0$； 或者 H_0 为 $\theta \leq (\geq)\theta_0$，$H_1$ 则为 $\theta \geq (\leq)\theta_0$，此时称为单边假设检验。在假设检验问题中， H_0 成立，而检验结果为 H_1 的情形称为第一类错误，通常记第一类错误发生的概率为 α；在 H_0 不成立，而检验结论为 H_0 的情形称为第二类错误， 记第二类错误的概率为 β。两类错误的危害通常是不同的，一般我们控制第一类发生错误的概率 $\alpha \leq \alpha_0$，α_0 是 人为选择的，一般称为显著水平。

5. 假设检验的枢变量法：

   (1) $U(\theta)$ 是 θ 的枢变量，其观测值为 u；

   (2) 令 u_1, u_2 为 $U(\theta)$ 分布的下、上 α 分位点，则$P(u_1 < U(\theta_0) < u_2 | H_0) = 1-\alpha$；

   (3) 若 $u_1 < u < u_2$，则接受原假设；反之，接受备择假设，则 $P(\text{拒绝 } H_0 | H_0) = \alpha$。

   注：(u_1, u_2) 称为接受域，$(-\infty, u_1) \cup (u_2, \infty)$ 称为拒绝域。

6. 假设检验的解题规范：

(1) 写出原假设和备择假设；

(2) 写出枢变量并计算其观测值；

(3) 写出拒绝域；

(4) 写出结论。

7. 几个问题：

(1) 假设检验和区间估计的一致性；$\hat{\theta}_1 < \theta < \hat{\theta}_2$ 等价于接受原假设；

(2) 假设检验的第一类错误发生的概率和第二类错误发生的概率，一般一个减少时，另一个增加；

(3) 假设检验的两个假设之间不是平等的，仅当有充分证据时才推翻原假设。

8. 总结：

枢变量	置信区间	拒绝域
$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$	$(\bar{X} - \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}}, \bar{X} + \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}})$	$(-\infty, -Z_{\frac{\alpha}{2}}) \cup (Z_{\frac{\alpha}{2}}, \infty)$
$\frac{\bar{X}-\mu}{S/\sqrt{n}}$	$(\bar{X} - \frac{S}{\sqrt{n}} t_{n-1}(\frac{\alpha}{2}), \bar{X} + \frac{S}{\sqrt{n}} t_{n-1}(\frac{\alpha}{2}))$	$(-\infty, -t_{n-1}(\frac{\alpha}{2})) \cup (t_{n-1}(\frac{\alpha}{2}), \infty)$
$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}$	$(\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\mathcal{X}_n^2(\frac{\alpha}{2})}, \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\mathcal{X}_n^2(1-\frac{\alpha}{2})})$	$(0, \mathcal{X}_n^2(1 - \frac{\alpha}{2})) \cup (\mathcal{X}^2)n(\frac{\alpha}{2}), \infty)$
$\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2}$	$(\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\mathcal{X}_{n-1}^2(\frac{\alpha}{2})}, \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\mathcal{X}_{n-1}^2(1-\frac{\alpha}{2})})$	$(0, \mathcal{X}_{n-1}^2(1 - \frac{\alpha}{2})) \cup (\mathcal{X}_{n-1}^2(\frac{\alpha}{2}), \infty)$

9. 单侧区间估计和单侧假设检验：

(1) 单侧置信上（下）限即将双侧置信上（下）限中的 $\frac{\alpha}{2}$ 换为 α；

(2) 单侧假设检验的原假设和备择假设的判断：包含有等号的部分为原假设；显著的部分为备择假设的部分；

(3) 单侧假设检验的拒绝域的方向和备择假设的方向一致。