引言 (概率论部分)

排列组合：

- 排列：n个里面挑m个组成一组，并排序号

$$P_n^m = n \times (n-1) \times (n-2) \times \cdots \times (n-m+1) = \frac{n!}{(n-m)!}$$

- 组合：n个里面挑m个组成一组

$$C_n^m = \frac{P_n^m}{m!} = \frac{n!}{(n-m)!m!}$$
$$C_n^m = C_n^{n-m}$$

排列组合计算方法

- 分类计数法（加法原理）
  - 一件事有n中不同套路，其中套路1有m_1种不同方法，套路2有m_2种不同的方法，$\cdots$，套路n有m_n种不同的方法，那么解决这件事有$\sum_{i=1}^{n} m_i$种不同的方法。
- 分布计数法
  - 完成一件事有n个步骤，步骤1有m_1种方法，步骤2有m_2种方法，$\cdots$，步骤n有m_n种方法，那么完成这件事一共有$\prod_{i=1}^{n} m_i$种不同的方法
- 捆绑法、插空法、插板法
- 不同的物品分配涉及排列，相同的物品分配涉及组合

基本概念和随机事件关系

试验

满足以下三点的试验，被称为"随机试验"，简称试验

- 可以在**相同条件**下重复进行
- 每次试验的可能结果不止一个，且试验前可以明确结果
- 进行试验之前不明确哪个结果会发生

样本空间

所有可能的结果的一个集合称为样本空间。如扔骰子的结果的样本空间为$S = \{1,2,3,4,5,6\}$，电灯泡寿命$S = \{t|t \geq 0\}$。

随机事件

指的样本是空间S的一个子集，通常用A, B, C表示。如"扔骰子扔出的点数是偶数"事件$A = \{2,4,6\}$。

事件发生

当随机事件结果中的集合元素出现，则称为事件发生。

基本事件

有单个样本点组成的集合。如扔骰子中的{1}、{2}

必然事件

S本身也是S的子集，包含了所有的样本点。如扔骰子扔出1-6中的一个点数。

不可能事件

用ϕ表示，是S的一个子集，不包含任何样本点。如扔骰子扔出10000点。

完备事件组

n个事件，事件的集合两两没有交际，并且它们的并集恰好就是样本空间S，则称这n个事件为一个完备事件组。如一个样本空间S，有子集A, B, C，则$A \cap B = \phi, A \cap C = \phi, B \cap C = \phi, A \cup B \cup C = S$

事件的运算律

基础

$$A + B = A \cup B$$
$$A \times B = A \cap B$$

交换律

$$A \cup B = B \cup A \quad A \cap B = B \cap A$$

结合律

$$A \cup (B \cup C) = (A \cup B) \cup C$$
$$A \cap (B \cap C) = (A \cap B) \cap C$$

分配律

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

德摩根律

$$\overline{A \cup B} = \bar{A} \cap \bar{B}$$
$$\overline{A \cap B} = \bar{A} \cup \bar{B}$$
$$\text{长杠变短杠，开口变方向}$$

频率与概率

概率的古典定义

事件A在n次测试中发生了m次，则事件A的概率为$P(A) = \frac{m}{n}$

频率

试验得到的结果

概率的频率

大量试验中，随着试验次数的增加，频率总是围绕着某个固定的数值波动，这个"固定的数值"被称为概率。

概率的公理化定义

设E是随机试验，S是E的样本空间，对于E的一个事件A赋予一个实数，记为P(A)，称为事件A的概率，如果集合函数$P(\cdot)$满足下列条件：

1. 对于每一个事件A，有$P(A) \geq 0$
2. 对于必然事件S，有$P(S) = 1$
3. 设$A_1, A_2, \cdots$是两两不相容的事件，即$A_iA_j = \phi \quad i \pm j, i, j = 1, 2, 3, \ldots$，有$P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$

概率的性质

1. 对于每一个事件A，有$0 \leq P(A) \leq 1$
2. 对于必然事件S，$P(S) = 1$,对于不可能事件ϕ，$P(\phi) = 0$
3. 有限可加性：对于有限事件$A_1, A_2, \ldots, A_n$，如果两两不相容，有$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$
4. 设两事件A,B，若$A \subset B$，则有$P(A) \leq P(B) \quad P(B - A) = P(B) - P(A)$
5. 对于任意事件A，对立事件$\bar{A}$的概率$P(\bar{A}) = 1 - P(A)$
$$P(\overline{A}) = P(S - A) = P(S) - P(A) = 1 - P(A)$$
6. 减法公式：对于任意两个事件A,B，有$P(B - A) = P(B) - P(AB)$
$$P(B - A) = P(B - AB) = P(B) - P(AB), AB \subset B$$
7. 加法公式：
   对于任意两个事件A,B，$P(A + B) = P(A) + P(B) - P(AB)$
   对于任意三个事件A,B,C，$P(A + B + C) = P(A) + P(B) + P(C) - P(AB) - P(BC) - P(AC) + P(ABC)$

古典概型

一个随机试验E，它的基本事件有有限个，并且每个基本事件出现的概率相等。把这类试验称为拉普拉斯试验，这类概率模型被称为古典概型，也叫等可能概型。

特点

- 有限样本点
- 每个样本点等可能发生

几何概型

特点

- 无限样本点
- 每个样本点等可能发生

条件概率与乘法公式

条件概率

事件A发生的条件下事件B发生的概率。

$$P(B|A)=\frac{P(AB)}{P(A)},P(A)>0$$

条件概率的性质

1. 对于每一个事件B都有$P(B|A)\geq 0$
2. 对于必然事件S，有$P(S|A)=1$
3. 可列可加性：

   设B_1,B_2,...两两互不相容，则有:

$$P(B_1|A\cup B_2|A\cup\ldots)=P(\cup_{i=1}^{\infty}B_i|A)=P(B_1|A)+P(B_2|A)+\cdots=\sum_{i=1}^{\infty}P(B_i|A)$$

4. $P(B|A)=1-P(\overline{B}|A)$

乘法公式

$$\text{由：}P(B|A)=\frac{P(AB)}{P(A)}$$
$$\text{可得：}P(AB)=P(A)P(B|A)$$

全概率公式

$$P(A)=\sum_{i=1}^{n}P(B_i)P(A|B_i)$$

贝叶斯公式

$$P(B_i|A)=\frac{P(AB_i)}{P(A)}=\frac{P(B_i)P(A|B_i)}{\sum\limits_{i=1}^{n}P(A|B_i)}$$

事件独立性

设A,B为两事件，若$P(AB)=P(A)P(B)$，则称事件A,B互相独立【可推广】

独立性的结论

1. $P(AB)=P(A)P(B)$
2. $P(B)=P(B|A),(P(A)\geq 0)$
3. $P(B|A)=P(B|\overline{A}),(0<P(A)<1)$
4. $P(A|B)=P(A|\overline{B}),(0<P(B)<1)$
5. $A\text{与}\overline{B},B\text{与}\overline{A},\overline{A}\text{与}\overline{B}\text{都互相独立}$
6. $\text{若}P(A)>0.P(B)>0,\text{则}A\text{与}B\text{互相独立和互不相容不同时独立}$
7. 必然事件与任何事件都互相独立
8. 不可能事件与任何事件都相互独立

随机变量

随机变量（函数）：

- 随机变量X是定义在随机试验样本空间S(e)上的单实值函数X=X(e),可以把样本空间上的值对应到实数数轴上。

1. 随机变量X=X(e)是一个单实值函数，随机试验的每一个结果都对应一个单实值
2. X(e)体现的是随机事件的描述
3. X(e)的每种取值都有一定的概率

$$\text{掷硬币为例，}H\text{表示正面}T\text{表示反面。}$$
$$X=\begin{cases}1, e=H \\ 0, e=T\end{cases}$$
$$P(X=1)=P(H)=\frac{1}{2}$$
$$P(X=0)=P(T)=\frac{1}{2}$$

4. 在随机试验之前可以确定X(e)所有可能出现的取值，但是不确定出现哪个值

离散型随机变量及其分布律

离散型随机变量指的是随机变量的全部可能取值，是有限个或可列无限个。

- 可列无限个指的是可以与自然数一一对应
- 分布律：随机变量X各个可能取值$x_k(k=1,2,...)$对应的概率$P(X=x_k)=P_k,(k=1,2,...)$称为X的分布律，一般用表格表示

X	x_1	x_2	...	x_n	...
P_k	P_1	P_2	...	P_n	...

以扔硬币为例：

X	0	1
P	$\frac{1}{2}$	$\frac{1}{2}$

- 分布律注意：
  - $P_i \geq 0, i=1,2,...$
  - $\sum_{i=1}^{\infty} P_i=1$

几种常见的离散型随机变量

0-1分布

试验只有两个结果，X的取值只有0和1

分布律：

X	0	1
P	p(0<p<1)	1-p(0<p<1)

二项分布 $B(n,p)$

n重伯努利试验（伯努利试验指的是要么A发生，要么$\overline{A}$发生），A发生的次数服从二项分布，A发生的概率记为p(0<p<1)，则$\overline{A}$发生的概率为1-p，记作B(n,p)

分布律：

$$B(n,p)=P(X=k)=C_n^k \cdot p^k \cdot (1-p)^{n-k},(k=0,1,2,\cdots,n)$$

例：某一学生咨询中心服务电话接通率为$\frac{3}{4}$,某班3位同学商定明天就同一问题询问该服务中心，且每人仅拨打一次电话，求他们中成功咨询的人数X的分布律。

解：$B(3,\frac{3}{4}),P(X=k)=C_3^k \times (\frac{3}{4})^k \times (\frac{1}{4})^{3-k}$

X	0	1	2	3
P	$P(X=0)=C_3^0 \times (\frac{3}{4})^0 \times (\frac{1}{4})^3=\frac{1}{64}$	$P(X=1)=C_3^1 \times (\frac{3}{4})^1 \times (\frac{1}{4})^2=\frac{9}{64}$	$P(X=2)=C_3^2 \times (\frac{3}{4})^2 \times (\frac{1}{4})^1=\frac{27}{64}$	$P(X=3)=C_3^3 \times (\frac{3}{4})^3 \times (\frac{1}{4})^0=\frac{27}{64}$

泊松分布 $\pi(\lambda)$或$P(\lambda)$

适用于描述单位时间或空间上随机事件发生的次数，如一小时内加油站到达的车辆数、单位面积上细菌的分布等，用$\pi(\lambda)$或$P(\lambda)$表示，其中λ是参数，指的是单位时间或空间事件发生的平均次数。

分布律：

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

口诀：拉k e负拉，比上k全家

例：设随机变量$X\sim\pi(\lambda)$(表示变量X服从泊松分布)，且$P(X=0)=e^{-2}$，则常数λ=_____,概率$P(X\le 2)=$_____。

解：$P(X=k)=\frac{\lambda^k e^{-\lambda}}{k!}$

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

$P(X\le 2)=P(X=0)+P(X=1)+P(X=2)=\frac{2^0e^{-2}}{0!}+\frac{2^1e^{-2}}{1!}+\frac{2^2e^{-2}}{2!}=5e^{-2}$

泊松定理

设$X\sim B(n,p)$，当**n较大且p较小**时，近似地，$X\sim\pi(np)$。

例：设一个工厂生产的零件次品率为0.1%，并且各个零件是否成为次品是互相独立的，求1000个零件中至少有2个次品的概率。

解：令X=次品的数量 $X\sim B(1000,0.001)$

$P(X\ge 2)=1-P(X<2)=1-P(X=0)-P(X=1)=$

$1-C_1^0 000\times(0.001)^0\times(0.999)^1 000-C_1^1 000\times(0.001)^1\times(0.999)^9 99=0.26424$

$P(X=k)=\frac{\lambda^k e^{-\lambda}}{k!},\lambda=np=1000\times 0.1\%=1$

$P(X\ge 2)=1-P(X=0)-P(X=1)=1-\frac{1^0\times e^{-1}}{0!}-\frac{1^1\times e^{-1}}{1!}=0.26424$

离散型随机变量的分布函数

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

将X所有小于等于x的取值对应的概率相加。例如：

X	1	2	3	4
P	0.1	0.2	0.3	0.4

$$F(x)=\begin{cases}0,x<1\\0.1,1\le x<2\\0.3,2\le x<3\\0.6,3\le x<4\\1,x\ge 4\end{cases}$$

注意：

1. $F(x)$是一个不减函数
2. $P\{a<x\le b\}=F(b)-F(a)$
3. $0\le F(x)\le 1,F(-\infty)=0,F(+\infty)=1$
4. $F(x)$右连续

例：设随机变量X的分布函数为：

$$F(x)=\begin{cases}0,x<-1\\0.1,-1\le x<0\\0.6,0\le x<1\\1,x\ge 1\end{cases}$$

求随机变量X的分布律。

解：

X	-1	0	1
P	0.1	0.5	0.4

一维连续型随机变量

一维离散型随机变量的分布律如下：

X	X_1	X_2	...	X_n	...
P	P_1	P_2	...	P_n	...

$$\sum_{k=1}^{\infty} P_k = 1(k=1,2,3,...)$$

对于一维连续型随机变量，其单点的概率为0，所以我们不研究其单点的概率，一般研究取值在一个区间上的概率。

连续型随机变量的概率密度函数$f(x)$

- 对于概率密度函数$f(x)$：
    1. $f(x) \geq 0$
    2. $\int_{-\infty}^{+\infty} f(x)dx = 1$
- 离散型随机变量的分布用分布律表示，连续型随机变量的分布用概率密度函数$f(x)$表示

连续型随机变量的分布函数$F(x)$

$$\text{对于离散型随机变量，} F(x) = P(X \leq x_n) = \sum_{i=1}^{n} P(x_i)$$

$$\text{对于连续型随机变量，} F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

$$\text{对于连续型随机变量，亦有：} F(X) = P(x_1 < X \leq x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(t)ft = \int_{-\infty}^{x_2} f(t)ft - \int_{-\infty}^{x_1} f(t)ft$$

例1：设随机变量X的概率密度函数为$f(x) = \begin{cases} A\cos x, 0 < x < \frac{\pi}{2} \\ 0, else \end{cases}$，则$A =$ ______。

解：$1 = \int_{-\infty}^{+\infty} f(x)dx = \int_{0}^{\frac{\pi}{2}} A\cos x = A \cdot \sin x|_{0}^{\frac{\pi}{2}} = A \to A = 1$

例2：设随机变量X的概率密度为$f(x) = \begin{cases} A\cos x, |x| < \frac{\pi}{2} \\ 0, else \end{cases}$,试求：

(1)常数A；　(2)分布函数$F(x)$；　(3)概率$P\{0 < X < \frac{\pi}{4}\}$。

解：(1)$1 = \int_{-\infty}^{+\infty} f(x)dx = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} A\cos x = A \cdot \sin x|_{-\frac{\pi}{2}}^{\frac{\pi}{2}} = 2A \to A = \frac{1}{2}$

可得$f(x) = \begin{cases} \frac{1}{2}\cos x, -\frac{1}{2} < x < \frac{\pi}{2} \\ 0, else \end{cases}$

(2)$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} \int_{-\infty}^{x} f(t)dt, x < -\frac{\pi}{2} \\ \int_{-\frac{\pi}{2}}^{x} \frac{\cos t}{2}dt, -\frac{\pi}{2} \leq x < \frac{\pi}{2} \\ 1, x \geq \frac{\pi}{2} \end{cases} = \begin{cases} 0, x < -\frac{\pi}{2} \\ \frac{\sin x + 1}{2}, -\frac{\pi}{2} \leq x < \frac{\pi}{2} \\ 1, x \geq \frac{\pi}{2} \end{cases}$

(3)$P\{0 < X < \frac{\pi}{4}\} = \int_{0}^{\frac{\pi}{4}} \frac{\cos x}{2} = \frac{\sqrt{2}}{4}$

一维连续型随机变量

一维离散型随机变量的分布律如下：

X	X_1	X_2	...	X_n	...
P	P_1	P_2	...	P_n	...

$$\sum_{k=1}^{\infty} P_k = 1(k=1,2,3,...)$$

对于一维连续型随机变量，其单点的概率为0，所以我们不研究其单点的概率，一般研究取值在一个区间上的概率。

连续型随机变量的概率密度函数$f(x)$

- 对于概率密度函数$f(x)$：
    1. $f(x) \geq 0$
    2. $\int_{-\infty}^{+\infty} f(x)dx = 1$
- 离散型随机变量的分布用分布律表示，连续型随机变量的分布用概率密度函数$f(x)$表示

连续型随机变量的分布函数$F(x)$

$$\text{对于离散型随机变量，}F(x)=P(X\le x_n)=\sum_{i=1}^{n}P(x_i)$$

$$\text{对于连续型随机变量，}F(x)=P(X\le x)=\int_{-\infty}^{x}f(t)dt$$

$$\text{对于连续型随机变量，亦有：}F(X)=P(x_1<X\le x_2)=F(x_2)-F(x_1)=\int_{x_1}^{x_2}f(t)ft=\int_{-\infty}^{x_2}f(t)ft-\int_{-\infty}^{x_1}f(t)ft$$

例1：设随机变量X的概率密度函数为$f(x)=\begin{cases}A\cos x,0<x<\dfrac{\pi}{2}\\0,else\end{cases}$，则$A=$______。

解：$1=\int_{-\infty}^{+\infty}f(x)dx=\int_{0}^{\frac{\pi}{2}}A\cos x=A\cdot\sin x|_{0}^{\frac{\pi}{2}}=A\to A=1$

例2：设随机变量X的概率密度为$f(x)=\begin{cases}A\cos x,|x|<\dfrac{\pi}{2}\\0,else\end{cases}$,试求：

(1)常数A；　(2)分布函数$F(x)$；　(3)概率$P\{0<X<\dfrac{\pi}{4}\}$。

解：(1)$1=\int_{-\infty}^{+\infty}f(x)dx=\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}A\cos x=A\cdot\sin x|_{-\frac{\pi}{2}}^{\frac{\pi}{2}}=2A\to A=\dfrac{1}{2}$

可得$f(x)=\begin{cases}\dfrac{1}{2}\cos x,-\dfrac{1}{2}<x<\dfrac{\pi}{2}\\0,else\end{cases}$

(2)$F(x)=P(X\le x)=\displaystyle\int_{-\infty}^{x}f(t)dt=\begin{cases}\displaystyle\int_{-\infty}^{x}f(t)dt,x<-\dfrac{\pi}{2}\\\displaystyle\int_{-\frac{\pi}{2}}^{x}\dfrac{\cos t}{2}dt,-\dfrac{\pi}{2}\le x<\dfrac{\pi}{2}\\1,x\ge\dfrac{\pi}{2}\end{cases}=\begin{cases}0,x<-\dfrac{\pi}{2}\\\dfrac{\sin x+1}{2},-\dfrac{\pi}{2}\le x<\dfrac{\pi}{2}\\1,x\ge\dfrac{\pi}{2}\end{cases}$

(3)$P\{0<X<\dfrac{\pi}{4}\}=\displaystyle\int_{0}^{\frac{\pi}{4}}\dfrac{\cos x}{2}=\dfrac{\sqrt{2}}{4}$

例3：设随机变量X的概率密度函数为$f(x)=\begin{cases}x,0\le x<1\\2-x,1\le x\le 2\\0,else\end{cases}$，则$P(X\le 1.5)=$(　)

$(A)0.875$　$(B)\int_{0}^{1.5}(2-x)dx$　$(C)\int_{1}^{1.5}(2-x)dx$　$(D)\int_{-\infty}^{1.5}(2-x)dx$

解：$P(X\le 1.5)=F(1.5)=\int_{-\infty}^{1.5}f(x)dx=\int_{0}^{1}xdx+\int_{1}^{1.5}(2-x)dx=0.875\to A$

另解：$P(X\le 1.5)=1-P(X>1.5)=1-\int_{1.5}^{2}(2-x)dx=0.875\to A$

常用的连续型随机变量

均匀分布$U(a,b)$

$$f(x)=\begin{cases}\dfrac{1}{b-a},a<x<b\\0,else\end{cases}$$

例：车辆一小时内任意时刻到达目的地的概率相等，服从均匀分布，问这辆车在$20\sim 30$分钟到达的概率相等？

解：$\int_{20}^{30}\dfrac{1}{60}=\dfrac{1}{6}$

指数分布 $e(\lambda)$或$Exp(\lambda)$

$$f(x)=\begin{cases}\lambda e^{-\lambda x},x>0\\0,x\le 0\end{cases}$$

$$F(x)=\begin{cases}\displaystyle\int_{0}^{x}\lambda e^{-\lambda t}dt=-e^{\lambda t}|_{0}^{x}=1-e^{-\lambda x},x>0\\0,x\le 0\end{cases}$$

$$P(X_1\le X\le X_2)=F(X_2)-F(X_1)$$

泊松分布($P(X=k)=\dfrac{\lambda^k e^{-\lambda}}{k!}$)：单位时间内事件发生一次的概率($\lambda$代表单位时间内事件发生的平均次数)

指数分布：事件下一次发生间隔时间对应的概率

例(区分)：一个医院平均每小时出生三个婴儿

对于泊松分布：一个小时内出生婴儿的概率$\to\lambda=3$

对于指数分布：下一个婴儿在多长时间后出生→ $\lambda = 3$，如求下一个婴儿在1-3小时内出生的概率
$P(1 < X < 3) = F(3) - F(1) = 1 - e^{-9} - (1 - e^{-3}) = e^{-3} - e^{-9}$

例：假设人的寿命服从指数分布，则一个60岁的老人和一个刚出生的婴儿再活十年的概率是否相等？

解：设人的寿命$X \sim e(\lambda)$，
$P(X \geq 70|X \geq 60) = \dfrac{P(X \geq 70 \cap X \geq 60)}{P(X \geq 60)} = \dfrac{P(X \geq 70)}{P \geq 60} = \dfrac{1 - P(X < 70)}{1 - P(X < 60)} = \dfrac{1 - F(70)}{1 - F(60)} = \dfrac{1 - (1 - e^{-70\lambda})}{1 - (1 - e^{-60\lambda})} = e^{-10\lambda}$

$P(X \geq 10|X \geq 0) = \dfrac{P(X \geq 0 \cap X \geq 0)}{P(X \geq 0)} = \dfrac{P(X \geq 10)}{P \geq 0} = \dfrac{1 - P(X < 10)}{1 - P(X < 0)} = \dfrac{1 - F(10)}{1 - F(0)} = \dfrac{1 - (1 - e^{-10\lambda})}{1 - (1 - e^{-0\lambda})} = e^{-10\lambda}$

指数分布的无记忆性

$$P(X > s + t|X > s) = P(X > t)$$

正态分布/高斯分布 $N(\mu, \sigma^2)$

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

性质

1. 关于$x = \mu$对称
2. 在$x = \mu$处取得最大值
3. 当$x < \mu$时，单调递增，$x > \mu$时单调递减
4. 在$x = \mu \pm \sigma$处有拐点
5. $y = 0$是水平渐近线

标准正态分布 $N(0, 1)$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\dfrac{x^2}{2}}, (-\infty < x < +\infty)$$

性质

1. $f(-x) = f(x)$，与y轴对称
2. 分布函数$\Phi(-x) = 1 - \Phi(x)$
3. 若$X \sim N(\mu, \sigma^2)$，则$Z = \dfrac{X - \mu}{\sigma} \sim N(0, 1)$
   1. 若$X \sim N(\mu, \sigma^2)$，则$F(x) = P\{X \leq x\} = P\{\dfrac{X - \mu}{\sigma} \leq \dfrac{x - \mu}{\sigma}\} = P\{Z \leq \dfrac{x - \mu}{\sigma}\} = \Phi(\dfrac{x - \mu}{\sigma})$
   2. $\begin{aligned} P(x_1 < X < x_2) = F(X_2) - F(X_1) = P(\dfrac{x_1 - \mu}{\sigma} < \dfrac{x - \mu}{\sigma} < \dfrac{x_2 - \mu}{\sigma}) \\ = P(\dfrac{x_1 - \mu}{\sigma} < Z < \dfrac{x_2 - \mu}{\sigma}) = \Phi(\dfrac{x_2 - \mu}{\sigma}) - \Phi(\dfrac{x_1 - \mu}{\sigma}) \end{aligned}$
   3. $P(X \geq x) = P\{\dfrac{X - \mu}{\sigma} \geq \dfrac{x - \mu}{\sigma}\} = 1 - P\{\dfrac{X - \mu}{\sigma} \leq \dfrac{x - \mu}{\sigma}\} = 1 - \Phi(\dfrac{X - \mu}{\sigma})$

例：假设某市高考考生成绩X服从正态分布$N(500, 100^2)$，现有考生25000名，计划招生10000名，试估计录取分数线。

解：设分数线为a，$P\{X \geq a\} = \dfrac{10000}{25000} = \dfrac{2}{5}$，由$X \sim N(500, 100^2)$得$Z = \dfrac{X - 500}{100} \sim N(0, 1)$

$P\{X \geq a\} = P\{\dfrac{X - \mu}{\sigma} \geq \dfrac{a - \mu}{\sigma}\} = P\{\dfrac{X - 500}{100} \geq \dfrac{a - 500}{100}\} = P\{Z \geq \dfrac{a - 500}{100}\} = 1 - P\{Z \leq \dfrac{a - 500}{100}\} = \dfrac{2}{5}$

得$1 - \Phi(\dfrac{a - 500}{100}) = 0.4 \to \Phi(\dfrac{a - 500}{100}) = 0.6$，查表得$\Phi(0.25) \approx 0.6$，故$a \approx 525$

随机变量函数的分布

离散型随机变量

例1：随机变量X的分布律如下，求：1、$Y = (X - 1)^2$的分布律 2、Y的分布函数$F_Y(y)$。

X	-2	-1	0	1	2
p_k	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{15}$	$\frac{11}{30}$

解：1、

X	-2	-1	0	1	2
Y	9	4	1	0	1

Y	0	1	4	9
P	$\frac{1}{15}$	$\frac{17}{30}$	$\frac{1}{6}$	$\frac{1}{5}$

2、$F_Y(y) = \begin{cases} 0, y < 0 \\ \frac{1}{15}, 0 \le y < 1 \\ \frac{19}{30}, 1 \le y < 4 \\ \frac{24}{30}, 4 \le y < 9 \\ 1, y \ge 9 \end{cases}$

连续型随机变量

例2：设连续型随机变量X的分布函数为$F(x) = \begin{cases} A + Be^{-2x}, x > 0 \\ 0, x \le 0 \end{cases}$，求：

1、A、B的值 2、$P(-1 < x < 1)$ 3、概率密度函数$f_X(x)$ 4、若$Y = 3X + 1$,求概率密度函数$f_Y(y)$

解：

1. $1 = \lim\limits_{x \to +\infty} F(x) = \lim\limits_{x \to +\infty} (A + Be^{-2x}) = A \to A = 1$

   $\lim\limits_{x \to 0^+} F(x) = F(0) = 0 = \lim\limits_{x \to 0^+} (A + Be^{-2x}) = 1 + B \to B = -1$

   $F(x) = \begin{cases} 1 - e^{-2x}, x > 0 \\ 0, x \le 0 \end{cases}$
2. $P(-1 < x < 1) = F(1) - F(-1) = 1 - e^{-2} - 0 = 1 - e^{-2}$
3. $f_X(x) = \frac{dF(x)}{dx} = \begin{cases} \frac{d(1 - e^{-2x})}{dx}, x > 0 \\ 0, x \le 0 \end{cases} = \begin{cases} 2e^{-2x}, x > 0 \\ 0, x \le 0 \end{cases}$
4. 由$Y = 3X + 1$，首先确定Y的取值范围：$x > 0 \to y > 1$；$x \le 0 \to y \le 1$

   $F_Y(y) = P\{Y \le y\} = P\{3x + 1 \le y\} = P\{X \le \frac{y-1}{3}\} = F_X(\frac{y-1}{3})$

   由此可用x表示y的分布函数 $\to F_Y(y) = \begin{cases} 1 - e^{-\frac{2}{3}(y-1)}, y > 1 \\ 0, y \le 1 \end{cases}$，求导得概率密度函数

   $f_Y(y) = \begin{cases} \frac{2}{3}e^{-\frac{2}{3}(y-1)}, y > 1 \\ 0, y \le 1 \end{cases}$

二维离散型随机变量

引言

例：学校研究学生阿巴和小东收快递，用X表示阿巴每周收快递数量，Y表示小东每周收快递数量，两人每周快递数量均分布在18、19、20三个数量上，得到下面的联合分布律：

X\Y	18	19	20
18	0.05	0.15	0.20
19	0.07	0.11	0.22
20	0.04	0.07	0.09

联合分布率

$P\{X - x_i, Y = y_j\} = P_{ij}(i, j = 1, 2, 3, \ldots)$

性质

1. $P_{ij} \ge 0$
2. $\sum\limits_i \sum\limits_j P_{ij} = 1$

联合分布函数

一维：$F(x) = P\{X \le x\}$

二维：$F(x, y) = P\{X \le x, Y \le y\}$

以引言为例，$F(19, 20) = P\{X \le 19, Y \le 20\} = 0.05 + 0.15 + 0.20 + 0.07 + 0.11 + 0.22 = 0.8$

边缘分布

单独考虑X、Y随机变量的分布。

X的边缘分布率：$P_i=\sum_j P_{ij},(j=1,2,\ldots)$，分布函数：$F_X(x)=P\{X\leq x\}$。

Y的边缘分布率：$P_i=\sum_j P_{ij},(j=1,2,\ldots)$，分布函数：$F_Y(y)=P\{Y\leq y\}$。

以引言为例，

X\Y	18	19	20	$P\{X=X_i\}$
18	0.05	0.15	0.20	0.40
19	0.07	0.11	0.22	0.40
20	0.04	0.07	0.09	0.20
$P\{Y=Y_i\}$	0.16	0.33	0.51	1.00

$P\{X=X_i\}$、$P\{Y=Y_i\}$即为X、Y的边缘分布。

条件分布

条件概率：$P(B|A)=\dfrac{P(AB)}{P(A)}$

相似地，条件概率 $P\{X=x|Y=y\}=\dfrac{P\{X=x,Y=y\}}{P\{Y=y\}}$

独立性

独立$\Leftrightarrow P(AB)=P(A)\cdot P(B)$

令：$A=\{X=x_i\},B=\{Y=y_j\}$, 则$P\{X=x_i,Y=y_j\}=P(X=x_i)\cdot P(Y=y_j)$。$P(X=x_i)=P_i,P(Y=y_j)=P_j$

综上，X、Y相互独立$\Leftrightarrow P_{ij}=P_i\cdot P_j,(i,j=1,2,\ldots)$

例1：已知二维离散型随机变量(X,Y)的联合分布率和边缘分布率满足下表：

X\Y	0	1	2	$P\{X=X_i\}$
-1	0.1		0.0	0.3
0	0.2	0.0		
1		0.1	0.1	
$P\{Y=Y_i\}$	0.4			

(1)将上表填写完整

(2)判断(X,Y)是否独立并说明理由

(3)写出$U=X+Y$的分布律

解：(1)

X\P\Y	0	1	2	$P\{X=X_i\}$
-1	**0.1**	0.2	**0.0**	**0.3**
0	**0.2**	**0.0**	0.2	0.4
1	0.1	**0.1**	0.1	0.3
$P\{Y=Y_i\}$	**0.4**	0.3	0.3	1.0

(2)$P_{11}=0.1\neq P_{1\cdot}\times P_{\cdot 1}=0.3\times 0.4=0.12$，故$(X,Y)$不独立

(3)

X(U\P)\Y	0	1	2
-1	-1\0.1	0\0.2	1\0.0
0	0\0.2	1\0.0	2\0.2
1	1\0.1	2\0.1	3\0.1

U	-1	0	1	2	3
P	0.1	0.4	0.1	0.3	0.1

例2：设二维随机变量(X,Y)的分布律为：

X\P\Y	1	2	3	$P\{X=X_i\}$
1		$\frac{1}{8}$		
2	$\frac{1}{8}$			
$P\{Y=Y_i\}$	$\frac{1}{6}$			1.0

若X、Y相互独立，

(1)填写上表空白部分

(2)求$U=\max\{X,Y\}$的分布律

(3)求$P(X>Y),P(X<Y)$

解：(1)

X\P\Y	1	2	3	$P\{X=X_i\}$
1	$\frac{1}{24}$	$\frac{1}{8}$	$\frac{1}{12}$	$\frac{1}{4}$
2	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{3}{4}$
$P\{Y=Y_i\}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$	1

(2)

X(P\U)\Y	1	2	3	$P\{X=X_i\}$
1	$\frac{1}{24}$\1	$\frac{1}{8}$\2	$\frac{1}{12}$\3	$\frac{1}{4}$
2	$\frac{1}{8}$\2	$\frac{3}{8}$\2	$\frac{1}{4}$\3	$\frac{3}{4}$
$P\{Y=Y_i\}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$	1

U	1	2	3
P	$\frac{1}{24}$	$\frac{5}{8}$	$\frac{1}{3}$

(3)$P(X>Y)=P\{X=2,Y=1\}=\dfrac{1}{8}$

$P(X<Y)=P\{X=1,Y=2\}+P\{X=1,Y=3\}+P\{X=2,Y=3\}=\dfrac{1}{8}+\dfrac{1}{12}+\dfrac{1}{4}=\dfrac{11}{24}$

二维连续型随机变量

联合分布函数

一维：$F(x)=P(X\le x)=\int_{-\infty}^{x}f(t)dt$

性质：

1. $P\{a<x<b\}=\int_{a}^{b}f(x)dx$
2. $P\{-\infty<x<+\infty\}=\int_{-\infty}^{+\infty}f(x)dx=1$

二维：$F(x,y)=P\{X\le x,Y\le y\}=\int_{-\infty}^{y}\int_{-\infty}^{x}f(u,v)dudv$

性质：

1. $P\{a\le x<b,c\le y<d\}=\int_{c}^{d}\int_{a}^{b}f(x,y)dxdy$
2. $P\{-\infty<x<+\infty,-\infty<y<+\infty\}=F(-\infty,+\infty)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)dxdy=1$

边缘分布

只考虑X,Y各自的分布

X的边缘分布函数

$$F_X(x)=P\{X\le x,-\infty<y<+\infty\}=\int_{-\infty}^{x}[\int_{-\infty}^{+\infty}f(x,y)dy]dx$$

X的概率密度函数

$$f_X(x)=\frac{dF_X(x)}{dx}=\int_{-\infty}^{+\infty}f(x,y)dy\to \text{一个关于}x\text{的函数}$$

Y的边缘分布函数

$$F_Y(y)=P\{Y\le y,-\infty<x<+\infty\}=\int_{-\infty}^{y}[\int_{-\infty}^{+\infty}f(x,y)dx]dy$$

Y的概率密度函数

$$f_Y(y)=\frac{dF_Y(y)}{dy}=\int_{-\infty}^{+\infty}f(x,y)dx\to \text{一个关于}y\text{的函数}$$

条件分布

1. 对于固定的$X=x$，若$f_X(x)>0$，则$f(y|x)=f_{Y|X}(y|x)=\dfrac{f(x,y)}{f_X(x)}$
2. 对于固定的$Y=y$，若$f_Y(y)>0$，则$f(x|y)=f_{X|Y}(x|y)=\dfrac{f(x,y)}{f_Y(y)}$

例1：设二维随机变量(X,Y)的概率密度函数为：

$$f(x,y)=\begin{cases}cx^2y,x^2\le y\le 1\\0,else\end{cases}$$

求：(1)常数c　(2)边缘概率密度$f_X(x)$　(3)$f_{Y|X}(y|x);f_{Y|X}(y|x=\dfrac{1}{3})$

解：(1)$P\{-1<x<1,x^2<y<1\}=\displaystyle\int_{x^2}^{1}\int_{-1}^{1}cx^2ydxdy=\frac{4c}{21}=1\to c=\frac{21}{4}$

(2)$f_X(x)=\displaystyle\int_{-\infty}^{+\infty}f(x,y)dy=\begin{cases}\displaystyle\int_{x^2}^{1}cx^2ydy,-1\le x\le 1\\0,else\end{cases}=\begin{cases}\dfrac{21}{8}x^2(1-x^4),-1\le x\le 1\\0,else\end{cases}$

(3)由$f_X(x)>0\Rightarrow x\ne 0,x\ne\pm1.$　$f_{Y|X}(y|x)=\dfrac{f(x,y)}{f_X(x)}=\begin{cases}\dfrac{2y}{1-x^4},x^2<y<1\\0,else\end{cases}$

把$x=\dfrac{1}{3}$代入，得$f_{Y|X}(y|x=\dfrac{1}{3})=\begin{cases}\dfrac{81y}{40},\dfrac{1}{9}<y<1\\0,else\end{cases}$

独立性

$f(x,y)=f_X(x)\cdot f_Y(y)\Leftrightarrow X$和$Y$相互独立

例2：设二维随机变量(X,Y)得联合概率密度为：$f(x,y)=\begin{cases}ke^{-\frac{y}{3}},0<x<3,y>0\\0,else\end{cases}$

(1)求参数k的值　(2)判断随机变量X,Y是否相互独立

解：(1)$\displaystyle\int_{0}^{3}dx\int_{0}^{+\infty}ke^{-\frac{y}{3}}dy=k\int_{0}^{3}-3e^{-\frac{y}{3}}|_{0}^{+\infty}dx=k\int_{0}^{3}3dx=9k=1\to k=\frac{1}{9}$

$$f(x,y)=\begin{cases}\frac{1}{9}e^{-\frac{y}{3}},0<x<3,y>0\\0,else\end{cases}$$

(2)$f(x,y)=f_X(x)\cdot f_Y(y)\Leftrightarrow X$和$Y$是相互独立

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy=\begin{cases}\displaystyle\int_{0}^{+\infty}\frac{1}{9}e^{-\dfrac{y}{3}},0<x<3\\0,else\end{cases}=\begin{cases}\frac{1}{3},0<x<3\\0,else\end{cases}$$

$$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)dx=\begin{cases}\displaystyle\int_{0}^{3}\frac{1}{9}e^{-\frac{y}{3}}dx,y>0\\0,else\end{cases}=\begin{cases}\frac{1}{3}e^{-\frac{y}{3}},y>0\\0,else\end{cases}$$

$f_X(x)\cdot f_Y(y)==\begin{cases}\dfrac{1}{9}e^{-\frac{y}{3}},0<x<3,y>0\\0,else\end{cases}=f(x,y),$故随机变量$X,Y$相互独立

两个随机变量的函数 $Z=X+Y$

已知二维随机变量(X,Y)的概率密度函数为$f(x,y)$，则$Z=X+Y$的概率密度函数为$f_Z(z)=\int_{-\infty}^{+\infty}f(x,z-x)dx$或$f_Z(z)=\int_{-\infty}^{+\infty}f(y,z-y)dy$

例3：已知(X,Y)的联合概率密度为：$f(x,y)=\begin{cases}x+y,0<x<1,<y<1\\0,else\end{cases}$

求$Z=X+Y$的概率密度函数

解：$f_Z(z)=\int_{-\infty}^{+\infty}f(x,z-x)dx$

$$f(x,z-x)\neq 0\Leftrightarrow\begin{cases}0<x<1\\0<z-x<1\end{cases}\Rightarrow\begin{cases}0<x<1\\z-1<x<z\end{cases}$$

$$f_Z(z)=\int_{-\infty}^{+\infty}f(x,z-x)dx=\begin{cases}\int_0^z zdx,0<z<1\\\int_{z-1}^1 zdx,1\leq z<2\\0,else\end{cases}=\begin{cases}z^2,0<z<1\\z(z-3,1\leq z<2\\0,else\end{cases}$$

期望与方差

数学期望 $E(X)$

1. 离散型

   $X=P\{X=x_k\}=P_k,k=1,2,\ldots$

   $E(X)=x_1P_1+x_2P_2+\cdots=\sum_{k=1}^{\infty}x_kP_k$

2. 连续型

   对于随机变量x，其概率密度函数为$f(x)$，则$E(X)=\int_{-\infty}^{+\infty}xf(x)dx$

数学期望的性质

1. $E(C)=C,C\ \ is\ \ constant$
2. $E(X+C)=E(X)+C,C\ \ is\ \ constant$
3. $E(CX)=CE(X),C\ \ is\ \ constant$
4. $E(X+Y)=E(X)+E(Y)$
5. 若X,Y相互独立，则$E(XY)=E(X)\cdot E(Y)$

方差$D(X)$

偏差的平方的期望，用于描述数据与平均值得偏离程度，为$E[(X-E(X))^2]$

1. 离散型

   $D(X)=(x_1-E(X))^2\cdot P_1+(x_2-E(X))^2\cdot P_2+\cdots+(x_n-E(X))^2\cdot P_n=\sum_{k=1}^{n}[x_k-E(X)]^2\cdot P_k$

2. 连续型

   $D(X)=\int_{-\infty}^{+\infty}[X-E(X)]^2\cdot f(x)dx$

注意：$D(X)=E(X^2)-[E(X)]^2$

方差的性质

1. $D(C)=0,C\ \ is\ \ constant$
2. $D(X+C)=D(X),,C\ \ is\ \ constant$
3. $D(CX)=C^2\cdot D(X),C\ \ is\ \ constant$
4. $D(X\pm Y)=D(X)+D(Y)\pm 2E[(X-E(X))(Y-E(Y))]$
5. 若X,Y相互独立，则$D(X\pm Y)=D(X)\pm D(Y)$

常用分布的期望与方差

分布	参数	分布律或概率密度	数学期望	方差
0-1分布	p	$P\{x=k\}=p^k(1-p)^{1-k},k=0,1$	p	$p(1-p)$
二项分布 $B(n,p)$	n,p	$P\{x=k\}=C_n^k p^k(1-p)^{n-k}$	np	$np(1-p)$
泊松分布 $P(\lambda)$ $\pi(\lambda)$	λ	$P\{x=k\}=\dfrac{\lambda^k e{-}\lambda}{k!}$	λ	λ
均匀分布 $U(a,b)$	a,b	$f(x)=\dfrac{1}{b-a},a<x<b$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$
正态分布 $N(\mu,\sigma^2)$	μ,σ	$f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\dfrac{(x-\mu)^2}{2\sigma^2}},-\infty<x<+\infty$	μ	σ^2
指数分布 $e(\lambda)$	λ	$f(x)=\begin{cases}\lambda e^{-\lambda x},x>0\\0,x\le 0\end{cases}$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$

例1：设X表示某学生10次考试不及格的次数，假设这10次考试相互独立，每次不及格概率为0.4，求$E(X^2),D(5X+3)$

解：由题意，采用二项分布$B(n,p)$，$X\sim B(10,0.4)$

$E(X)=np=10\times0.4=4,D(X)=np(1-p)=10\times0.4\times0.6=2.4$。

$D(X)=E(X^2)-[E(X)]^2=2.4\Rightarrow E(X^2)=D(X)+[E(X)]^2=2.4+4^2=18.4$

$D(5X+3)=5^2D(X)=25\times18.4=460$

例2：假设某宝宝去幼儿园，迟到一次概率为0.2，若一周5天都没迟到，则可以获得十朵小红花；若迟到一次，则获得五朵小红花；若迟到两次，则只能获得一朵小红花；迟到三次及以上没有小红花。问该宝宝一周内获得小红花数量的期望。

解：设Y表示该宝宝获得小红花的数量，X表示该宝宝迟到的天数。

由题意得，采用二项分布$B(n,p),X\sim B(5,0.2)$

$P\{Y=10\}=P\{X=0\}=C_5^0 0.2^0 0.8^5=0.328$

$P\{Y=5\}=P\{X=1\}=C_5^1 0.2^1 0.8^4=0.41$

$P\{Y=1\}=P\{X=2\}=C_5^2 0.2^2 0.8^3=0.205$

$P\{Y=0\}=P\{X\ge3\}=P\{X=3\}+P\{X=4\}+P\{X=5\}=C_5^3 0.2^3 0.8^2+C_5^4 0.2^4 0.8^1+C_5^5 0.2^5 0.8^0=0.057$

Y	0	1	5	10
P	0.057	0.205	0.41	0.328

$$E(Y)=0\times0.057+1\times0.205+5\times0.41+10\times0.328=5.535$$

协方差与相关系数

描述两个随机变量的关系(线性关系)，线性关系有正相关、负相关和不相关三种关系。

协方差 $Cov(X,Y)$

$Cov(X,Y)=E\{[X-E(X)][Y-E(Y)]\}$

1. 可正可负可零
2. $\begin{aligned}Cov(X,Y)&=E\{[X-E(X)][Y-E(Y)]\}=E\{XY-XE(Y)-YE(X)+E(X)E(Y)\}\\&=E(XY)-E(Y)E(X)-E(X)E(Y)+E(X)E(Y)=E(XY)-E(X)E(Y)\end{aligned}$

协方差的性质

1. $Cov(X,Y)=Cov(Y,X)$
2. $Cov(X,a)=Cov(b,Y)=0,a、b\ \ are\ \ constants$
3. $Cov(aX,bY)=abCov(X,Y),a、b\ \ are\ \ constants$
4. $Cov(X_1+X_2,Y)=Cov(X_1,Y)+Cov(X_2,Y)$
5. $Cov(X_1+X_2,Y_1+Y_2)=Cov(X_1,Y_1)+Cov(X_1,Y_2)+Cov(X_2,Y_1)+Cov(X_2,Y_2)$
6. 若X,Y相互独立，则$Cov(X,Y)=0$
7. 方差和协方差的关系：$D(X\pm Y)=D(X)+D(Y)\pm2Cov(X,Y)$
8. 协方差受到量纲的影响

对协方差标准化

$$\hat{X}=\frac{X-E(X)}{\sqrt{D(X)}},\hat{Y}=\frac{Y-E(Y)}{\sqrt{D(Y)}}$$

$$Cov(\hat{X},\hat{Y})=Cov[\frac{X-E(X)}{\sqrt{D(X)}},\frac{Y-E(Y)}{\sqrt{D(Y)}}]=\frac{Cov[X-E(X),Y-E(Y)]}{\sqrt{D(X)}\sqrt{X(Y)}}$$
$$=\frac{Cov(X,Y)-Cov(X,E(Y))-Cov(E(X),Y)+Cov(E(X),E(Y))}{\sqrt{D(X)}\sqrt{X(Y)}}=\frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

相关系数

用于表示两个随机变量之间的线性关系

$$\rho_{xy}=\frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

1. $-1\le\rho_{xy}\le 1$
2. $\rho_{xy}\begin{cases}=1,\text{完全正相关}\\>0,\text{正相关}\\0,\text{完全无关}\\<0,\text{负相关}\\=-1,\text{完全负相关}\end{cases}$
3. $|\rho_{xy}|$越大，相关性越强；$|\rho_{xy}|$越接近0，相关性越弱

相关系数的性质

1. X,Y不相关，有$\rho_{xy}=0\Leftrightarrow Cov(X,Y)=0\Leftrightarrow E(XY)=E(X)E(Y)\Leftrightarrow D(X\pm Y)=D(X)+D(Y)$
2. X,Y互相独立$\Rightarrow$ X,Y不相关

例1：设X,Y的分布律为：

X\Y	0	1	2	3	$P(X=x_i)$
1	0	3/8	3/8	0	3/4
3	1/8	0	0	1/8	1/4
$P(Y=y_j)$	1/8	3/8	3/8	1/8	1

求$Cov(X,Y),\rho_{xy}$并判断X与Y是否独立。

解：$Cov(X,Y)=E(XY)-E(X)E(Y),\rho_{xy}=\frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$

X	1	3	Y	0	1	2	3
P	3/4	1/4		1/8	3/8	3/8	1/8
X^2	1	9	Y^2	0	1	4	9

$$E(X)=\frac{3}{2},E(X^2)=3,E(Y)=\frac{3}{2},E(Y^2)=3\Rightarrow D(X)=\frac{3}{4},D(Y)=\frac{3}{4}$$

XY	0	1	2	3	6	9
P	1/8	3/8	3/8	0	0	1/8

$$E(XY)=\frac{9}{4}$$

$$Cov(X,Y)=E(XY)-E(X)E(Y)=0,\rho_{xy}=\frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=0$$

因为：$P_{1,1}=0\neq P_{1\cdot}\times P_{\cdot 1}=\frac{3}{4}\cdot\frac{1}{8}=\frac{3}{32}$，所以$X,Y$不独立。

例2：设二维随机变量(X,Y)的概率密度为：

$$f(x,y)=\begin{cases}2,x>0,y>0,x+y<1\\0,else\end{cases}$$

求$Cov(X,Y)$和ρ_{xy}

解：$Cov(X,Y)=E(XY)-E(X)E(Y),\rho_{xy}=\frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$

$$E(X)=\int_{-\infty}^{+\infty}xf_X(x)dx,E(Y)=\int_{-\infty}^{+\infty}yf_Y(y)dx,E(XY)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyf_{XY}(xy)dxdy$$

$$D(X)=E(X^2)-[E(X)]^2,D(Y)=E(Y^2)-[E(Y)]^2$$

$$E(X^2)=\int_{-\infty}^{+\infty}x^2f_X(x)dx,E(Y^2)=\int_{-\infty}^{+\infty}y^2f_Y(y)dx$$

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy=\begin{cases}\int_0^{1-x}2dy,0<x<1\\0,else\end{cases}=\begin{cases}2-2x,0<x<1\\0,else\end{cases}$$

$$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)dy=\begin{cases}\int_0^{1-y}2dx,0<y<1\\0,else\end{cases}=\begin{cases}2-2y,0<y<1\\0,else\end{cases}$$

$$E(X)=\int_{-\infty}^{+\infty}xf_X(x)dx=\int_0^1x(2-2x)dx=\frac{1}{3}$$

$$E(Y)=\int_{-\infty}^{+\infty}xf_Y(y)dx=\int_0^1y(2-2y)dy=\frac{1}{3}$$

$$E(XY)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyf_{XY}(xy)dxdy=\int_0^1\int_0^{1-x}2xydydx=\int_0^1xdx\int_0^{1-x}2ydy=\int_0^1x\cdot y^2|_0^{1-x}dx=\int_0^1x(1-x)^2dx$$

$$=\int_0^1x^3-2x^2+xdx=\frac{x^4}{4}|_0^1-2\frac{x^3}{3}|_0^1+\frac{x^2}{2}|_0^1=\frac{1}{4}-2\times\frac{1}{3}+\frac{1}{2}=\frac{1}{12}$$

$$E(X^2)=\int_{-\infty}^{+\infty}x^2f_X(x)dx=\int_0^1x^2(2-2x)dx=\frac{1}{6}$$

$$E(Y^2)=\int_{-\infty}^{+\infty}y^2f_Y(y)dx=\int_0^1y^2(2-2y)dy=\frac{1}{6}$$

$$D(X)=E(X^2)-[E(X)]^2=\frac{1}{6}-\frac{1}{3}^2=\frac{1}{18}$$

$$D(Y)=E(Y^2)-[E(Y)]^2=\frac{1}{6}-\frac{1}{3}^2=\frac{1}{18}$$

$$Cov(X,Y)=E(XY)-E(X)E(Y)=\frac{1}{12}-\frac{1}{3}\times\frac{1}{3}=-\frac{1}{36}$$

$$\rho_{xy}=\frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=\frac{-\frac{1}{36}}{\sqrt{\frac{1}{18}}\sqrt{\frac{1}{18}}}=-\frac{1}{2}$$

大数定律

频率：做n次试验，其中事件A发生了n_A次，则$f_A=\dfrac{n_A}{n}$

随着n的增大，f_A逐渐稳定于某个值，这个值就是概率P

伯努利大数定律

做n次试验，其中事件A发生了n_A次，$f_A=\dfrac{n_A}{n}$表示n次试验A发生的频率，P是事件A在每次试验中发生的概率，则对于任意的$\varepsilon>0$，有：

$$\lim_{n\to\infty}P\{|f_A-P|<\varepsilon\}=1$$

则$f_A\xrightarrow{P}P$（f_A依概率收敛于P），n足够大时，可以用f_A近似P

辛钦大数定律

设随机变量$X_1,X_2,\cdots,X_n,\cdots$独立且同分布，则它们具有相同的期望$E(X_k)=\mu,k=1,2,\ldots$。作前$n$个变量的算术平均值$\overline{X}$，$\overline{X}=\dfrac{1}{n}\sum\limits_{k=1}^{n}x_k$，则对于任意$\varepsilon>0$，有：

$$\lim_{n\to\infty}P\{|(\overline{x})-\mu|<\varepsilon\}=1$$

即$\overline{X}\xrightarrow{P}\mu$，说明了当$n$足够大时，可以用$\overline{X}$近似$\mu$

中心极限定理

林德伯格-莱维中心极限定理

设随机变量$X_1,X_2,\ldots$独立同分布，且存在数学期望与方差，即$E(X_k)=\mu,D(X_k)=\sigma^2\neq0(k=1,2,\ldots)$，则对于任意的$x\in(-\infty,+\infty)$，有$\lim\limits_{n\to\infty}P\{\dfrac{\sum_{k=1}^{n}}{\sqrt{n}\sigma}<x\}=\dfrac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}dt}$

解释：

当n足够大时，有$\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \overset{近似}{\sim} N(0,1)$，$\sum_{k-1}^{n} X_k N \overset{近似}{\sim} (n\mu, n\sigma)$

棣莫夫-拉普拉斯中心极限定理

设随机变量$X_1, X_2, \ldots$独立同分布，且$P\{X_k = 1\} = p, P\{X_k = 0\} = 1 - p, (0 < p < 1, k = 1, 2, \ldots)$，则对于任意的$x \in (-\infty, +\infty)$，有$\lim_{n \to \infty} P\{\frac{\sum_{k=1}^{n} X_k - np}{\sqrt{np(1-p)}} < x\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}$

解释：

$X = \sum_{k=1}^{n} X_k \sim B(n, p)$

当n足够大时，有：

$\frac{X - np}{\sqrt{np(1-p)}} \overset{近似}{\sim} N(0,1)$，$X \overset{近似}{\sim} N[np, np(1-p)]$

即二项分布的极限分布是正态分布

例：某高校图书馆阅览室共有880个座位，该校共有学生12000人，已知每晚每个学生到阅览室仔细地概率为8%。

(1)求阅览室每天晚上座位不够用的概率

(2)若要以80%的概率保证晚上去阅览室自习的学生都有座位，阅览室还需要增加多少座位？

解：(1)设随机变量X表示去阅览室的人数，$X \sim B(12000, 0.08)$

$\because 12000足够大 \therefore X \overset{近似}{\sim} N(960, 883.2)$

$P(880 < X \leq 12000) = P\{\frac{880-960}{\sqrt{883.2}} < \frac{X-960}{\sqrt{883.2}} \leq \frac{12000-960}{\sqrt{883.2}}\} \approx \Phi(\frac{12000-960}{\sqrt{883.2}}) - \Phi(\frac{880-960}{\sqrt{883.2}}) \approx 1 - (1 - 0.9964) = 0.99$

(2)假设要增加a个座位，使得$P\{X \leq 880 + a\} \geq 0.8$

$P\{X \leq 880 + a\} = P\{\frac{X-960}{\sqrt{883.2}} \leq \frac{880+a-960}{\sqrt{883.2}}\} \approx \Phi(\frac{880+a-960}{\sqrt{883.2}}) \geq 0.8$

查表得$\Phi(0.842) = 0.8 \Rightarrow \frac{880+a-960}{\sqrt{883.2}} \Rightarrow a \leq 105.02$

$\therefore$要增加至少106张座位

样本的相关概念（数理统计部分）

“抽样调查”基本概念

1. 总体：试验的全部可能观察值
2. 个体：每一个可能观察值
3. 总体容量：总体中包含的个体数量
4. 抽样调查：从总体中抽取一部分个体进行观测
   1. 总体X种进行n次独立的重复观察，得到观察结果用$X_1, X_2, \ldots, X_n$表示。随机变量$X_1, X_2, \ldots, X_n$来自同一个总体，且独立同分布。那么称$X_1, X_2, \ldots, X_n$是来自同一总体的一组简单随机样本(简称样本)、
   2. 抽象调查中得到一组数值$x_1, x_2, \ldots, x_n$依次是$X_1, X_2, \ldots, X_n$的观察值，称为样本值
5. 样本量：抽取样本中个体的数量
6. 总体期望(μ)，总体方差(σ^2)，样本均值($\overline{X}$)，样本方差(s^2)

统计量

简单随机样本$X_1, X_2, \ldots, X_n$的函数$g(X_1, X_2, \ldots, X_n)$，**不包含未知参数**的**随机变量**。

$\overline{X} = \frac{1}{n}(X_1, X_2, \ldots, X_n)$

例1：设总体$X \sim N(\mu, \sigma^2)$，其中μ已知，σ^2未知。$X_1, X_2, \ldots, X_n$是来自总体X的样本，$x_1, x_2, \ldots, x_n$是抽样得到的样本观察值，问下列哪个不是统计量？（2,3）

(1)$\frac{1}{n}\sum_{i=1}^{n} X_i$　(2)$\sum_{i=1}^{n}(\frac{X_i - \mu}{\sigma})^2$　(3)$\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$　(4)$\max\{X_1, X_2, \ldots, X_n\}$

常用统计量：

1. 样本均值：$\overline{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$
2. 样本方差：$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$

3. 样本标准差：$s=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2}$

4. 样本k阶原点矩：$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k(k=1,2,\ldots)$，用于表示随机变量离原点距离的平均值

5. 样本k阶中心矩：$B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k(k=1,2,\ldots)$，用于表示随机变量离均值距离的平均值

6. 峰度：$Kurt(X)=\frac{B_4}{\sigma^4}$，用于表示分布的陡峭情况，正态分布的峰度为3

7. 偏度：$Skew(X)=\frac{B_3}{\sigma^3}$，用于表示分布的偏离情况，正态分布的偏度为0

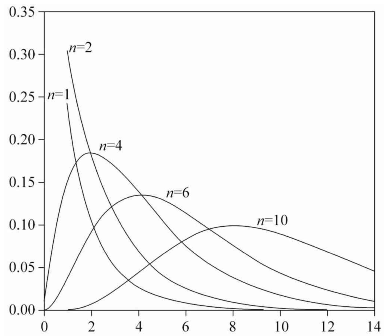
例2：抽样调查某小区年龄分布，随机抽取5个人，得到观察值8,29,15,8,21，计算样本均值和样本方差。

解：$\overline{X}=\frac{1}{5}\sum_{i=1}^{5}X_i=\frac{8+29+15+8+21}{5}=16.2$

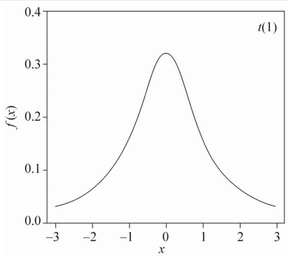
$$s^2=\frac{1}{5-1}\sum_{i=1}^{5}(X_i-\overline{X})^2=\frac{(8-16.2)^2+(29-16.2)^2+(15-16.2)^2+(8-16.2)^2+(21-16.2)^2}{4}=80.7$$

抽样分布

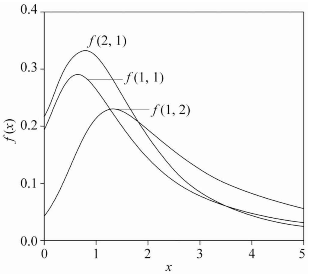
$\chi^2(n)$分布

定义	$\chi^2=X_1^2+X_2^2+\cdots+X_n^2$服从自由度为$n$的$\chi^2$分布，记作$\chi^2\sim\chi^2(n)$ 其中$X_1,X_2,\cdots,X_n$相互独立，且$X_i\sim N(0,1),i=1,2,\ldots$
图像	
期望与方差	$E(\chi^2)=n\,D(\chi^2)=2n$
性质	若$X\sim\chi_1^2(n_1),Y\sim\chi_2^2(n_2)$，$X,Y$相互独立，则$X+Y\sim\chi^2(n_1+n_2)$

$t(n)$分布

定义	设$X\sim N(0,1),Y\sim\chi^2(n)$，$X,Y$互相独立，则称$T=\frac{X}{\sqrt{\frac{Y}{n}}}$服从自由度为$n$的$t$分布，记作$T\sim t(n)$
图像	
性质	当$n>45$时，有$T=\frac{X}{\sqrt{\frac{Y}{n}}}\overset{近似}{\sim}N(0,1)$

$F(n_1,n_2)$**分布**

定义	设$U\sim\chi_1^2(n_1),V\sim\chi_2^2(n_2)$，$U,V$相互独立，则称$F=\dfrac{U/n_1}{V/n_2}$服从自由度为$(n_1,n_2)$的$F$分布 记作$F\sim F(n_1,n_2)$，其中$n_1$被称为第一自由度，$n_2$被称为第二自由度
图像	
性质	$F\sim F(n_1,n_2)$，则$\dfrac{1}{F}\sim F(n_2,n_1)$

正态分布样本均值和样本方差的分布

设$X_1,X_2,...,X_n$来自正态总体$X\sim N(\mu,\sigma^2)$，$\overline{X}$表示样本均值，s^2表示样本方差，则：

1. $\overline{X}\sim N(\mu,\dfrac{\sigma^2}{n})$
2. $\dfrac{(n-1)s^2}{\sigma^2}\sim\chi^2(n-1)$
3. $\overline{X}$与s^2相互独立
4. $\dfrac{\overline{X}-\mu}{\frac{s}{\sqrt{n}}}=\dfrac{\overline{X}-\mu}{\sqrt{\frac{s^2}{n}}}\sim t(n-1)$

例1：设随机变量$X\sim N(1,2^2)$，$X_1,X_2,...,X_{100}$是取自总体X的样本，$\overline{X}$为样本均值，已知$Y=a\overline{X}+b\sim N(0,1)$，求$a$和$b$的值

解：$\overline{X}\sim N(1,\dfrac{1}{25})$，$E(\overline{X})=1$，$D(\overline{X})=\dfrac{1}{25}$

$E(Y)=E(a\overline{X}+b)=aE(\overline{X})+b=0\Rightarrow a+b=0$

$D(Y)=D(a\overline{X}+b)=a^2D(\overline{X})=1\Rightarrow a^2=25$

$$\begin{cases}a+b=0\\a^2=25\end{cases}\Rightarrow\begin{cases}a=5\\b=-5\end{cases}\begin{cases}a=-5\\b=5\end{cases}$$

例2：设总体$X\sim N(0,9)$，$X_1,X_2,...,X_{30}$是取自总体X的样本，写出下列统计量$Y=\dfrac{1}{9}\sum\limits_{i=1}^{15}X_i^2$，$Z=\dfrac{4X_{17}}{\sqrt{\sum_{i=1}^{16}X_i^2}}$，$W=\dfrac{\sum_{i=1}^{20}X_i^2}{2\sum_{i=21}^{30}X_i^2}$的分布，并计算$Y$的期望与方差

解：$X\sim N(0,9)$ $\dfrac{X-\mu}{\sigma}=\dfrac{X}{3}\sim N(0,1)$

$Y=\dfrac{1}{9}\cdot 9\sum_{i=1}^{15}(\dfrac{X_i}{3})^2=\sum_{i=1}^{15}(\dfrac{X_i}{3})^2\sim\chi^2(15),E(Y)=15,D(Y)=30$

$$Z=\frac{4\cdot 3\cdot\frac{X_{17}}{3}}{4\cdot 3\cdot\sqrt{\sum_{i=1}^{16}(\frac{X_i}{3})^2/16}}=\frac{\frac{X_{17}}{3}}{\sqrt{\sum_{i=1}^{16}(\frac{X_i}{3})^2/16}}\sim t(16)$$

$$W=\frac{9\cdot 20\cdot\sum_{i=1}^{20}(\frac{X_i}{3})^2/20}{2\cdot 9\cdot 10\cdot\sum_{i=21}^{30}(\frac{X_i}{3})^2/10}=\frac{\sum_{i=1}^{20}(\frac{X_i}{3})^2/20}{\sum_{i=21}^{30}(\frac{X_i}{3})^2/10}\sim F(20,10)$$

点估计

估计：通过样本信息推测总体信息

点估计：构造合适的统计量，估计总体的未知参数

矩估计

用样本矩来估计总体矩

$A_k=\dfrac{1}{n}\sum_{i=1}^{n}X_i^k$用$\overline{X}=\dfrac{1}{n}\sum_{i=1}^{n}X_i$估计总体$\mu$，记作$\hat{\mu}=\overline{X}$

计算未知参数θ的矩估计方法

1. $\mu = E(X) = g(\theta)$
2. 令$\mu = g(\theta) = \overline{X}$，反解出$\hat{\theta} = h(\overline{X})$，称为$\theta$的估计量
3. 用样本观察值求出$\overline{X}$的值，代入$h(\overline{X})$种，得到θ的估计值

例1：设总体X的概率分布如下：

X	0	1	2	3
P	θ^2	$2\theta(1-\theta)$	θ^2	$1-2\theta$

其中$\theta(0 < \theta < \dfrac{1}{2})$是未知参数，利用总体$X$的如下样本值：

$$3, 1, 3, 0, 3, 1, 2, 3$$

求θ的矩估计值。

解：$\mu = E(X) = 0 \times \theta^2 + 1 \times 2\theta(1-\theta) + 2 \times \theta^2 + 3(1-2\theta) = 3 - 4\theta$

令$\mu = 3 - 4\theta = \overline{X}, 反解得\hat{\theta} = \dfrac{3-\overline{X}}{4}$

$\overline{X} = \dfrac{3+1+3+0+3+1+2+3}{8} = 2, 带入\hat{\theta} = \dfrac{3-\overline{X}}{4} = \dfrac{1}{4}$

例2：设总体X的概率密度为$f(x) = \begin{cases} e^{-(x-\theta)}, x \geq 0 \\ 0, x < 0 \end{cases}$，而$X_1, X_2, ..., X_n$是来自总体$X$的简单随机样本，求未知参数$\theta$的矩估计量？

解：$\mu = E(X) = \displaystyle\int_{-\infty}^{+\infty} xf(x)dx = \int_0^{+\infty} xe^{-(x-\theta)}dx \overset{分部积分}{\Longrightarrow} e^\theta$

令$\mu = e^\theta = \overline{X}$，反解出$\hat{\theta} = \ln\overline{X} = \ln\left(\displaystyle\sum_{i=1}^n X_i\right), i = 1, 2, ..., n$

极大似然估计

一个试验中，得到了一组样本观察值，则有理由相信这个观察值出现的概率最大。因此，考察未知参数θ取值时，这组样本观察值出现的概率最大，那么就用这个值座位θ的极大似然估计值。

求极大似然估计的方法

1. 写出似然函数$L(\theta)$，即当前样本所对应的概率
2. 为方便求导，对于$L(\theta)$取对数
3. 对于取对数后的$L(\theta)$求导并令$\dfrac{dlnL(\theta)}{d\theta} = 0$
4. 反解出θ，该θ即为估计值$\hat{\theta}$
5. 若所要求为估计量，则使用随机变量的表达形式

例3：设总体X的概率分布如下：

X	0	1	2	3
P	θ^2	$2\theta(1-\theta)$	θ^2	$1-2\theta$

其中$\theta(0 < \theta < \dfrac{1}{2})$是未知参数，利用总体$X$的如下样本值：

$$3, 1, 3, 0, 3, 1, 2, 3$$

求θ的极大似然估计值。

解：$L(\theta) = \theta^2 \times [2\theta(1-\theta)]^2 \times \theta^2 \times (1-2\theta)^4 = 4\theta^6(1-\theta)^2(1-2\theta)^4$

$\ln L(\theta) = \ln 4\theta^6(1-\theta)^2(1-2\theta)^4 = \ln 4 + \ln\theta^6 + \ln{(1-\theta)}^2 + \ln{(1-2\theta)}^4 = \ln 4 + 6\ln\theta + 2\ln{(1-\theta)} + 4\ln{(1-2\theta)}$

$\dfrac{d\ln L(\theta)}{d\theta} = 0 + \dfrac{6}{\theta} - \dfrac{2}{1-\theta} - \dfrac{8}{1-2\theta} = \dfrac{2(24\theta^2 - 14\theta + 3)}{\theta(1-\theta)(1-2\theta)}$

令$\dfrac{d\ln L(\theta)}{d\theta} = 0$，解得$\hat{\theta}_1 = \dfrac{7-\sqrt{13}}{13}, \hat{\theta}_2 = \dfrac{7+\sqrt{13}}{13}(> \dfrac{1}{2}, 与题设不符，舍去)$

$\therefore \hat{\theta} = \dfrac{7-\sqrt{13}}{13}就是\theta的极大似然估计值$

例4：设总体X的概率密度为$f(x) = \begin{cases} (\theta+1)x^\theta, 0 < x < 1 \\ 0, else \end{cases}$，其中$\theta > -1$时未知参数，$x_1, x_2 ... x_n$时来自总体$X$的一个容量为$n$的简单随机样本，求$\theta$的极大似然估计量。

解：$L(\theta)=(\theta+1)^n\prod_{i=1}^{n}x_i^\theta$

$\ln L(\theta)=n\ln(\theta+1)+\theta\sum_{i=1}^{n}\ln(x_i)$

$\frac{dlnL(\theta)}{d\theta}=\frac{n}{\theta+1}+\sum_{i=1}^{n}\ln(x_i)$

令$\frac{d\ln L(\theta)}{d\theta}=0$，解得$\hat{\theta}=-\frac{n}{\sum_{i=1}^{n}\ln(x_i)}-1$

$\therefore\hat{\theta}=-\frac{n}{\sum_{i=1}^{n}\ln(X_i)}-1$就是$\theta$的极大似然估计量

区间估计与假设检验

区间估计 $N(\mu,\sigma^2)$

区间估计是参数估计的一种形式，通常也是对均值或方差进行估计。主体思想是通过从总体中抽取的样本，根据一定的正确度与准确度的要求，构造出适当的区间，以作为总体参数真值所在范围的估计。点估计估计出的是一个具体的值没区间估计估计得到的是一个区间。

计算区间$(\underline{\theta},\overline{\theta})$使得$P\{\underline{\theta}<\mu<\overline{\theta}\}=95\%$

置信区间：$(\underline{\theta},\overline{\theta})$

置信上限：$\overline{\theta}$

置信下限：$\underline{\theta}$

置信度/置信水平：$1-\alpha$

设总体X的分布函数是$F(x,\theta)$，其中θ是未知参数。对于同一给定的值$\alpha(0<\alpha<1)$，若有样本$X_1,X_2,\ldots,X_n$确定的两个统计量$\underline{\theta}(X_1,X_2,\ldots,X_n)$和$\overline{\theta}(X_1,X_2,\ldots,X_n)$满足$P\{\underline{\theta}<\theta<\overline{\theta}\}\geq1-\alpha$，则称随机区间$(\underline{\theta},\overline{\theta})$是参数$\theta$的置信度为$1-\alpha$的（双侧）置信区间。

若统计量$\underline{\theta}=\underline{\theta}(X_1,X_2,\ldots,X_n)$满足$P\{\theta>\underline{\theta}\}\geq1-\alpha$，则称随机区间$(\underline{\theta},+\infty)$是$\theta$的置信度为$1-\alpha$的单侧置信区间，并称$\underline{\theta}$为单侧置信下限。

若统计量$\overline{\theta}=\overline{\theta}(X_1,X_2,\ldots,X_n)$满足$P\{\theta<\overline{\theta}\}\geq1-\alpha$，则称随机区间$(-\infty,\overline{\theta},)$是$\theta$的置信度为$1-\alpha$的单侧置信区间，并称$\overline{\theta}$为单侧置信上限。

置信区间的意义：样本容量固定为n，加入对总体进行100次抽样，就得到了100个置信区间。这些区间有的包含θ的真实值，有的不包含。但是假设当置信度$1-\alpha=95\%$时，这100个区间中大约有95个包含了θ的真实值。

估计μ，σ^2已知时

用$\overline{X}$估计μ　　$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i\sim N(\mu,\sigma^2),\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\sim N(0,1)$

$P\{-U_{0.025}<\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}<U_{0.025}\}=95\%\Longrightarrow P\{\overline{X}-\frac{\sigma}{\sqrt{n}}U_{0.025}<\mu<\overline{X}+\frac{\sigma}{\sqrt{n}}U_{0.025}\}=95\%$

μ落在$(\overline{X}-\frac{\sigma}{\sqrt{n}}U_{0.025},\overline{X}+\frac{\sigma}{\sqrt{n}}U_{0.025})$内的概率为$0.95\Longrightarrow\mu$落在$(\overline{X}-\frac{\sigma}{\sqrt{n}}U_{\frac{\alpha}{2}},\overline{X}+\frac{\sigma}{\sqrt{n}}U_{\frac{\alpha}{2}})$内的概率为$0.95$

$\underline{\mu}$落在$\overline{X}-\frac{\sigma}{\sqrt{n}}U_\alpha$，$\overline{\mu}$落在$\overline{X}+\frac{\sigma}{\sqrt{n}}U_\alpha$内的概率为$1-\alpha$

估计μ，σ^2未知时

用$\overline{X}$估计μ，用s代替σ

$\frac{\overline{X}-\mu}{\frac{s}{\sqrt{n}}}\sim t(n-1)\Longrightarrow\mu$落在$(\overline{X}-\frac{s}{\sqrt{n}}t_{\frac{\alpha}{2}}(n-1),\overline{X}+\frac{s}{\sqrt{n}}t_{\frac{\alpha}{2}}(n-1))$内的概率为$1-\alpha$

$\underline{\mu}$落在$\overline{X}-\frac{s}{\sqrt{n}}t_\alpha(n-1)$，$\overline{\mu}$落在$\overline{X}+\frac{s}{\sqrt{n}}t_\alpha(n-1)$内的概率为$1-\alpha$

估计σ^2，μ未知时

σ^2落在$(\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}(n-1)},\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)})$内的概率为$1-\alpha$

$\underline{\sigma}^2$落在$\frac{(n-1)s^2}{\chi^2_\alpha(n-1)}$，$\overline{\sigma}^2$落在$\frac{(n-1)s^2}{\chi^2_{1-\alpha}(n-1)}$内的概率为$1-\alpha$

例1：某商店每天每百元骰子的利润率$X\sim N(\mu,1)$服从正态分布，均值为μ，长期以来方差σ^2稳定为1，现随机抽取的100天的利润，样本均值$\overline{x}=5$，试求μ的置信水平为95%的置信区间。($t_{0.05}(100)=1.99,\Phi(1.96)=0.975$)

解：$\dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\sim N(0,1)\quad P\{-U_{0.025}<\dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}<U_{0.025}\}=95\%$

$$\Longrightarrow \text{置信区间为}(\overline{X}-\frac{\sigma}{\sqrt{n}}U_{0.025},\overline{X}+\frac{\sigma}{\sqrt{n}}U_{0.025})$$

$$\overline{x}=5,\sigma=1,n=100,U_{0.025}=1.96\Rightarrow \text{置信区间为}(5-\frac{1}{10}\times 1.96,5+\frac{1}{10}\times 1.96)\Rightarrow(4.804,5.796)$$

例2：某大学中教授的年龄$X\sim N(\mu,\sigma^2)$，μ,σ^2均未知。现随即了解到5位教授的年龄如下：39,54,61,72,59。试求均值μ的置信度为0.95的置信区间。($t_{0.025}(4)=2.7764$)

解：$\dfrac{\overline{X}-\mu}{\frac{s}{\sqrt{n}}}\sim t(n-1)$

$$\Longrightarrow \mu\text{落在}(\overline{X}-\frac{s}{\sqrt{n}}t_{0.025}(n-1),\overline{X}+\frac{s}{\sqrt{n}}t_{0.025}(n-1))\text{内的概率为}95\%$$

$$\overline{X}=\frac{1}{5}(39+54+61+72+59)=57,s=\sqrt{\frac{1}{n-1}\sum_{i=1}^{5}(X_i-\overline{X})^2}=12,n=5$$

$$\Rightarrow \text{置信区间为}(57-\frac{12}{\sqrt{5}}\times 2.7764,57+\frac{12}{\sqrt{5}}\times 2.7764)\Rightarrow(42.13,71.87)$$

例3：设(θ_1,θ_2)是参数θ的置信度为$1-\alpha$的区间估计，则以下结论正确的是(C)

A. 参数θ落在区间(θ_1,θ_2)之内的概率为$1-\alpha$

B. 参数θ落在区间(θ_1,θ_2)之外的概率为α

C. 区间(θ_1,θ_2)包含参数θ的概率为$1-\alpha$

D. 对不同的样本观测值，区间(θ_1,θ_2)的长度相同

假设检验

对于假设检验问题，提出关于总体得一个假设，称为原假设，记作H_0；与原假设相对立的假设，称为备择假设，记作H_1。假设检验中用到的统计量，称为检验统计量。检验统计量吧样本空间分为两个区域，使得H_0被拒绝的样本观察值所组成的区域内为拒绝域。此时，检验统计量落入拒绝域中的概率是给定的小概率α，α被称为显著水平。

设θ为总体的位置参数，θ_0是已知参数，关于θ_0的假设检验类型有：

类型	**方向**	H_0	H_1
双边检验	双边	$\theta=\theta_0$	$\theta\neq\theta_0$
单边检验	右边	$\theta\leq\theta_0$	$\theta>\theta_0$
单边检验	左边	$\theta\geq\theta_0$	$\theta<\theta_0$

假设检验中的两类错误

类型	**含义**	**犯错的概率**	**说明**
第一类错误	原假设H_0为真，却拒绝H_0，即为**弃真错误**	$\alpha=P\{\text{拒绝}H_0\|H_0\text{为真}\}$	(1)仅控制犯第一类错误的概率的检验称为显著性检验，α称为显著性水平
第二类错误	原假设H_0不真，却接受H_0，即为**去伪错误**	$\beta=P\{\text{接受}H_0\|H_0\text{不真}\}$	(2)当样本容量固定时，α和β种任意一个减小，另一个必然增大；如果使得α和β同时增大，那么只能增大样本容量

假设检验的步骤

1. 根据题意写出原假设H_0和备择假设H_1
2. 选择检验方法，写出检验统计量及其分布
3. 根据给定的显著性水平确定拒绝域
4. 计算检验统计量的值，做出推断

检验μ，σ^2已知时

原假设与备择假设:

$H_0:\mu=\mu_0,H_1:\mu\neq\mu_0$

$H_0:\mu\leq\mu_0,H_1:\mu>\mu_0$

$H_0:\mu\geq\mu_0,H_1:\mu<\mu_0$

检验法及检验统计量:

$$U\text{检验}\ U=\frac{\overline{X}-\mu}{\sigma_0/\sqrt{n}}\sim N(0,1)$$

拒绝域:

$|u|\geq u_{\frac{\alpha}{2}}$

$u\geq u_\alpha$

$u\leq -u_\alpha$

检验μ，σ^2未知时

原假设与备择假设:

$H_0:\mu=\mu_0,H_1:\mu\neq\mu_0$

$H_0:\mu\leq\mu_0,H_1:\mu>\mu_0$

$H_0:\mu\geq\mu_0,H_1:\mu<\mu_0$

检验法及检验统计量:

$$T\text{检验}\ T=\frac{\overline{X}-\mu}{s/\sqrt{n}}\sim t(n-1)$$

拒绝域:

$|t|\geq t_{\frac{\alpha}{2}}(n-1)$

$t\geq t_\alpha(n-1)$

$t\leq t_\alpha(n-1)$

检验σ^2，μ已知时

原假设与备择假设:

$H_0:\sigma^2=\sigma_0^2,H_1:\sigma^2\neq\sigma_0^2$

$H_0:\sigma^2\leq\sigma_0^2,H_1:\sigma^2>\sigma_0^2$

$H_0:\sigma^2\geq\sigma_0^2,H_1:\sigma^2<\sigma_0^2$

检验法及检验统计量:

$$\chi^2\text{检验}\ \chi^2=\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2\sim\chi^2(n)$$

拒绝域:

$\chi^2\leq\chi^2_{1-\frac{\alpha}{2}}(n)\text{或}\chi^2\geq\chi^2_{\frac{\alpha}{2}}(n)$

$\chi^2\geq\chi^2_\alpha(n)$

$\chi^2\leq\chi^2_{1-\alpha}(n)$

检验σ^2，μ未知时

原假设与备择假设:

$H_0:\sigma^2=\sigma_0^2,H_1:\sigma^2\neq\sigma_0^2$

$H_0:\sigma^2\leq\sigma_0^2,H_1:\sigma^2>\sigma_0^2$

$H_0:\sigma^2\geq\sigma_0^2,H_1:\sigma^2<\sigma_0^2$

检验法及检验统计量:

$$\chi^2\text{检验}\,\chi^2=\frac{(n-1)s^2}{\sigma^2}\sim\chi^2(n-1)$$

拒绝域:

$\chi^2\leq\chi^2_{1-\frac{\alpha}{2}}(n-1)\text{或}\chi^2\geq\chi^2_{\frac{\alpha}{2}}(n-1)$

$\chi^2 \geq \chi^2_\alpha(n-1)$

$\chi^2 \leq \chi^2_{1-\alpha}(n-1)$

例4：某次考试的成绩服从正态分布，随机抽取了36位考生的成绩，算得平均分为66.5分，标准差为15分，在显著性水平0.05下，是否可以认为这次考试的平均分为70分？

解：$\overline{X}=66.5, s=15, \mu=70, n=36, \alpha=0.05$

提出假设：$H_0:\mu=70, H_1:\mu\neq 70$

$T=\dfrac{\overline{X}-\mu}{\dfrac{s}{\sqrt{n}}}\sim t(n-1)$符合题设条件

给定显著性水平0.05，写出拒绝域$|T|<-t_{\frac{\alpha}{2}}(n-1)$

计算统计量作出推断$\left|\dfrac{\overline{X}-\mu}{\dfrac{s}{\sqrt{n}}}\right|=\left|\dfrac{66.5-70}{\dfrac{15}{\sqrt{36}}}\right|=1.4$

查表得$t_{0.025}(35)=2.0301$

$\because 1.4<2.0301 \therefore$ 不在拒绝域内，接受原假设

例5：设$X_1,X_2,\ldots,X_n$是来自正态总体$X\sim N(\mu,\sigma^2)$的简单随机样本，其中参数μ,σ^2未知，记$\overline{X}=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i, Q^2=\sum\limits_{i=1}^{n}(X_i-\overline{X})^2$，则假设$H_0:\mu=0$的$t$检验使用统计量$t$为？

解：由题意，采用$\dfrac{\overline{X}-\mu}{s/\sqrt{n}}\sim t(n-1)$，$\mu=0$

$s^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2=\dfrac{1}{n-1}Q^2\Rightarrow s=\dfrac{Q}{\sqrt{n-1}}$

$t=\dfrac{\overline{X}-\mu_0}{s/\sqrt{n}}=\dfrac{\overline{X}}{\dfrac{Q}{\sqrt{n(n-1)}}}=\dfrac{\overline{X}\sqrt{n(n-1)}}{Q}$

例6：设样本$X_1,X_2,\ldots,X_n$抽自总体$X\sim N(\mu,\sigma^2)$，μ,σ^2均未知。要对μ作假设检验，统计假设为$H_0:\mu=\mu_0(\mu_0\text{已知}), H_1:\mu\neq\mu_0$,则要用检验统计量为$\underline{\quad\dfrac{\overline{X}-\mu}{s/\sqrt{n}}\quad}$，给定显著水平$\alpha$，则检验的拒绝区间为$\underline{\quad(-\infty,-t_{\frac{\alpha}{2}}(n-1))\cup(t_{\frac{\alpha}{2}}(n-1),+\infty)\quad}$。

解：由题意，采用$\dfrac{\overline{X}-\mu}{s/\sqrt{n}}\sim t(n-1)$，$\mu=\mu_0$，本题纯概念