概率统计(理工) 第二版 阅读笔记

第一章 基础概率论

1.1 样本空间与随机事件

随机试验满足：

1. 可在相同条件下重复进行
2. 实验结果不止一个，实验前可知一切结果
3. 实验前不知道出现哪个结果，试验后能确定结果

随机试验用E表示

记Ω为一个试验所有可能结果的集合，则Ω称为**样本空间。**试验的任何一个可能结果称为**样本点**

样本空间的一个子集为一个**随机事件**，简称**事件。**Ω包含所有样本点，是**必然事件**；$\varnothing$是**不可能事件**；仅包含一个样本点的是**基本事件**

事件B**包含**事件A表示为$A \subset B$，表示若A发生则B一定发生。若同时$B \subset A$，则A和B事件**相等**

$A \cup B$表示**并事件**，表示A与B至少发生一个的事件；$A \cap B$表示**交事件**，表示A与B同时发生的一个事件

$A - B$称为事件A与B的**差**，表示发生A而不发生B的事件，满足：

$$P(A-B) = P(A) - P(AB)$$

若$A \cap B = \varnothing$，则称A与B**互斥**；若同时满足$A \cup B = \Omega$，则称A与B**互逆。**若$A_1, A_2, \cdots, A_n$两两互斥，且$\cup_{i=1}^{n} A_i = \Omega$，则称$A_1, A_2, \cdots, A_n$是一个**完备事件组**

事件的运算满足**交换律**、**结合律**、**分配率**和**对偶律**，其中对偶律可表示为：

$$\begin{aligned} \overline{A \cup B} &= \overline{A} \cap \overline{B} \\ \overline{AB} &= \overline{A} \cup \overline{B} \end{aligned}$$

1.2 事件发生的概率

在n次重复试验中，若事件A发生了k次，则称k为事件A发生的**频数**，称$\frac{k}{n}$为**频率**，记为：

$$f_n(A) = \frac{k}{n}$$

频率满足：

$$(1)\quad 0 \le f_n(A) \le 1$$
$$(2)\quad f_n(\Omega)=1, f_n(\varnothing)=0$$

且在A_i互斥时满足：

$$f_n(\cup_{i=1}^{r} A_i)=\sum_{i=1}^{r} f_n(A_i)$$

> 互斥事件的频率和等于互斥事件和(并集)的频率

设Ω为样本空间，对于事件A，对应一个$P(A)$，若$P(A)$满足：

1. 非负性 $P(A)>0$
2. 规范性 $P(\Omega)=1$
3. 可列可加性 即对互斥事件$P(\cup_{i=1}^{\infty} A_i)=\sum_{i=1}^{\infty} P(A_i)$
   则$P(A)$是事件A的**概率**

满足：

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

1.3 等可能概型

等可能概型指的是一个试验中，所有样本点都等可能出现的概率模型

1.3.1 古典概型

满足:

1. 试验仅有限个结果
2. 每个结果在试验中出现的可能性相等

的是**古典概型**

设A为Ω中的事件，包含样本点ω_{ik}，则：

$$P(A)=\frac{k}{n}=\frac{\sum_{ij,j=1}^{k}\omega_{ij}}{\sum_{i=1}^{n}\omega_i}$$

记为A的**古典概率**

古典概率的第一个摸球模型：从N个编号球中抽r个球，若放回，可能情况有N^r中；若不放回，可能情况有$P_N^r=N(N-1)\cdots(N-r+1)$种

古典概率的第二个摸球模型：从N个无编号且仅有黑白二色的球中抽取n个，恰摸到k个白球的概率为$p_k=\frac{C_k^m C_{N-k}^{N-m}}{C_N^n}$，这个概率被称为**超几何概率**

1.3.2 几何概型

若事件的结果等可能地出现在一个有界的欧式区域Ω内，这个试验的概率模型就是**几何概型**，计算公式：

$$P(A)=\frac{m(A)}{m(\Omega)}$$

其中$m()$表示几何度量

> 几何概型都可以用作图解释

通过模拟大量的随机试验，根据实验结果确定一个定值的方法叫做**蒙特卡洛法**

1.4 条件概率

设A、B为同一试验的两事件，且$P(B)>1$，称：

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

为B发生条件下，A发生的**条件概率**，满足：

$$P(A|B)\geq 0P(\Omega|B)=1$$

同样在互斥条件下条件概率可加，满足其他概率性质

由条件概率得到：

$$P(AB)=P(A|B)P(B)$$

这被称为**乘法公式**，推广后：

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

由乘法公式可以推得**全概率公式**：

$$P(B)=\sum_{i=1}^{n}P(A_i)P(B|A_i)$$

表示B可能由各种事件A_i发生后引起的概率。**贝叶斯公式**：

$$P(A_i|B)=\frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n}P(A_j)P(B|A_j)}$$

1.5 事件的独立性

对于同一试验的两事件A、B，若：

$$P(AB)=P(A)P(B)$$

则称A与B**相互独立**

> 互斥事件必不独立，因为一个出现会使另一个出现的概率为0
> 由上面的推断可知，独立事件一定不互斥

独立事件满足：独立事件中的两事件及其逆事件相互独立

设n个事件A_i，其中A_i和A_j是其中的两个事件，若对于任意的i,j，都有：

$$P(A_iA_j)=P(A_i)P(A_j)$$

则称这n个事件**两两独立**。在此基础上，若对于任意k个事件都有：

$$P(A_{i1}A_{i2}\cdots A_{ik})=P(A_{i1})P(A_{i2})\cdots P(A_{ik})$$

则称这n个事件**相互独立**

> 实际应用中一般通过实际意义判断独立性

在概率论中，概率小于5%的的事件一般被认为是**小概率事件**，有两个特点：

1. 在一次试验中几乎不可能发生
2. 在大量重复试验中几乎必定发生至少一次

将随机试验E重复进行n次，若每次试验结果不会相互影响，这样的试验被称为 **n重伯努利试验**。特别的，若每次试验结果只有两个，即A与$\overline{A}$，这样的试验被称为 **n重伯努利试验**，相应的数学模型叫**伯努利概型**。

在n重伯努利试验中，若A在每次试验中发生的概率为p，则在n次实验中，A发生k次的概率为：

$$P_n(k)=C_n^kp^k(1-p)^{n-k}$$

在n重独立试验中，每次试验结果可能是$A_1,A_2,\cdots,A_k$，且它们构成一个完备事件组。则在n次试验中，A_i各自发生r_i次的概率为：

$$p=\frac{n!}{r_1!r_2!\cdots r_k!}p_1^{r_1}p_2^{r_2}\cdots p_k^{r_k}$$

其中$\sum_{i=1}^k r_i=n$，上式被称为**多项概率公式**

> r表示选择次数，p表示各自概率，可用组合数公式推导

第二章 随机变量及其分布

2.1 随机变量和分布函数

设Ω为一个样本空间。若对于每个样本点$\omega\in\Omega$，规定一个实数$X(\omega)$，这样就定义了一个定义域为Ω的实值函数$X=X(\omega)$，称X为**随机变量**

设X是一随机变量，对任意实数x，定义：

$$F(x)=P(X\le x),\quad x\in R$$

称$F(x)$是随机变量X的**分布函数**

> 注意x一定要取遍实数轴，画出分布函数为0和1的区域

设$F(x)$为一个随机变量X的分布函数，则：

1. 当$x_1 < x_2$时，$F(x_1) \leq F(x_2)$，即$F(x)$单调不减
2. $F(-\infty) = P(X \leq -\infty) = \lim_{x \to -\infty} F(x) = 0$，同样地$F(\infty) = 1$
3. $F(x)$是右连续的，$F(x) = \lim_{x \to 0^+} F(x + 0)$
4. 对任意x_0，$P(x_0) = F(x_0) - F(x_0 - 0)$

2.2 离散型随机变量及其分布

若随机变量X的所有可能取值为有限个或可数无穷多个值，则称X为**离散型随机变量**

设离散型随机变量X的取值$x_1, x_2, \cdots, x_n$，且X取值概率为：

$$p_k = P(X = x_k), \quad k = 1, 2, \cdots, n$$

则称上式为X的**概率分布**、**概率函数**或**分布律**，记作：

$$X \sim \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ p_1 & p_2 & \cdots & p_n \end{pmatrix}$$

因为事件$(X = x_k)$是两两互斥的，则：

$$F(x) = P(X \leq x) = P(\cup_{x_k \leq x}(X = x_k)) = \sum_{x_k \leq x} p_k$$

离散型随机变量常见的分布包括：

- 几何分布
- 超几何分布
- 二项分布
- 泊松分布

几何分布

在n重伯努利试验中，若试验科一直重复，叫做**可列重伯努利试验**

若随机变量X取值为$1, 2, \cdots$，且：

$$p_k = p(X = k) = p(1 - p)^{k-1}$$

则称X服从参数p的**几何分布**，记为$X \sim G(p)$，其中p表示成功率

几何分布可以用来表示n重伯努利实验中，首次成功需要实验次数的概率

超几何分布

设N、n、m为正整数，且$n \leq N, m \leq N$，若随机变量有分布律：

$$p_k = P(X=k) = \frac{C_m^k C_{N-m}^{n-k}}{C_N^n}$$

则称X服从**超几何分布**，记为$X \sim H(n,m,N)$

二项分布

若X取值为有限自然数，且：

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

其中p表示成功率，则称X服从参数为n、p的**二项分布**，记为$X \sim B(n,p)$。特别地，当X取值为0或1时，表示**0-1分布**

> 当超几何分布的N足够大时，可以近似为二项分布，其中$p = \frac{m}{N}$

泊松分布

设随机变量$X_n \sim B(n,p_n)$，满足$np_n = \lambda$，则有：

$$\lim_{n\to\infty} P(X_n = k) = \lim_{n\to\infty} C_n^k p_n^k (1-p_n)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}$$

这个定理称为**泊松定理**

若随机变量X的可能取值为自然数，且：

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

其中λ为常数，则称X服从参数为λ的泊松分布，记为$X \sim P(\lambda)$。由泊松定理知，当p较小且n较大时，二项分布的概率函数近似为泊松分布，其中$\lambda = np$

2.3 连续型随机变量及其分布

设随机变量X的分布函数为$F(x)$，若存在一个非负函数$f(x)$，使得对任意的实数x，都有：

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

则称X为**连续型随机变量**，$f(x)$为X的**概率密度函数**，简称为**密度函数**或**密度**

密度函数的性质包括：

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

设X为连续型随机变量，则有：

- 对任意常数$a < b$，有

$$P(a < X \leq b) = \int_a^b f(x)dx$$

- $F(x)$是连续函数，在$f(x)$连续点，有：

$$F'(x) = f(x)$$

- 对任意常数C，有$P(X = C) = 0$

均匀分布

设随机变量X的密度函数为：

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

则称X在区间$[a, b]$上满足**均匀分布**，记为$X \sim U(a, b)$

> 一般地，把密度函数大于0的区间叫做连续型随机变量的取值区间

指数分布

设随机变量X的密度函数为：

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

则称X服从参数λ的**指数分布**，记为$X \sim e(\lambda)$。不难得到分布函数：

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

指数分布具有**无记忆性**。即对于$\forall t > 0, \Delta t > 0$，对于一个指数分布的$X$，有：

$$P(X > t + \Delta t | X > t) = P(X > \Delta t)$$

指数分布是唯一具有无记忆性的连续型随机变量分布

Γ分布

关于$\alpha > 0$的含参积分：

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha - 1} e^{-x} dx$$

叫做Γ积分，也叫 Γ**函数**

Γ函数的性质有：

- $\Gamma(\alpha + 1) = \alpha \Gamma(\alpha), \quad \alpha > 0$
- $\Gamma(1) = 1, \Gamma(\frac{1}{2}) = \sqrt{\pi}$
- 对正整数n，$\Gamma(n) = (n - 1)!$

设随机变量X的密度函数为：

$$f(x)=\begin{cases}\frac{\beta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}, & x>0\\ 0, & x\leq 0\end{cases}$$

其中$\alpha>0$，$\beta>0$为常数，则称X服从参数为α,β的 Γ**分布**，记为$X\sim\Gamma(\alpha,\beta)$

> 当$\alpha=1$，时，Γ分布退化为指数分布

2.4 随机变量函数的分布

离散型随机变量函数

离散型随机变量X的函数$Y=g(X)$的概率分布相当于X有概率分布
$p_k=P(X=x_k),k=1,2,\cdots$时：

$$P(Y=y_j)=\sum_{g(x_k)=y_j}p_k,\quad j=1,2,\cdots$$

连续型随机变量函数

设有连续型随机变量X，密度函数$f_X(x)$，连续型随机变量Y有$Y=g(X)$，

1. 由X取值区间，得到Y的值域$R(Y)$
2. 求Y的分布函数，对$\forall y\in R(Y)$，

$$\begin{aligned}F_Y(y)&=P(Y\leq y)=P(g(X)\leq y)=P(X\in G(y))\\&=\int_{G(y)}f_X(x)dx\end{aligned}$$

其中$X\in G(y)$由X关于y的函数解出，对$y\notin R(Y)$，$F(y)=0$ 或 $F(y)=1$

3. 最后$f_Y(y)=F_Y'(y),y\in R(Y)$或$f_Y(y)=0,y\notin R(Y)$

第三章 多元随机变量及其分布

3.1 二维随机变量及其分布函数

设X和Y是定义在同一样本空间Ω上的两个随机变量，则称(X,Y)是**二维随机变量(向量)。**对$\forall$实数a,b，称：

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

为**二维分布函数**，或称为X与Y的**联合分布函数**

> 联合分布函数表示几个随机变量同时发生的概率

二维分布的性质有：

- $F(x,y)$分别关于x,y单调不减

- $F(-\infty,\infty)=1$，$F(-\infty,-\infty)=F(x,-\infty)=F(-\infty,y)=0$
- $F(x,y)$关于x,y都右连续
- 对$\forall x_1<x_2,y_1<y_2$，有：

$$F(x_2,y_2)+F(x_1,y_1)-F(x_1,y_2)-F(x_2,y_1)\geq 0$$

离散型二维随机变量

设二维离散型随机变量的$\forall$取值为x_i,y_j，记：

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

则称上式为(X,Y)的**二维概率分布**或**分布律**，或称**联合概率分布**

满足的性质有：

- $p_{ij}\geq 0$
- $\sum_{i,j}p_{ij}=1$

由二维离散型随机变量可以推到**三项分布。**即在一个n重独立实验中，每次试验都有A_1,A_2,A_3三个可能结果，其中A_1,A_2,A_3的发生概率为p_1,p_2,p_3，令随机变量X和Y分别表示n次实验中A_1和A_2的发生次数，可知联合分布：

$$P(X=k_1,Y=k_2)=\frac{n!}{k_1!k2!(n-k_1-k_2)!}p_1^{k_1}p_2^{k_2}p_3^{n-k_1-k_2}$$

称(X,Y)服从参数为p_1,p_2,n的三项分布，记为$(X,Y)\sim T(n;p_1,p_2)$

二维连续型随机变量

对二维随机变量(X,Y)，如果存在二元非负函数$f(x,y)$，使得$\forall x,y\in R$，有：

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

则称(X,Y)是**二维连续型随机变量**，$f(u,v)$是其**二维概率密度函数**，简称为**密度**

满足：

- $f(x,y)\geq 0$
- $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)dxdy=1$
- 若$F(x,y)$连续，且在$f(x,y)$的连续点(x,y)，有：

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

- 对平面上任意区域G，若$f(x,y)$在G上可积，有：

$$P((X,Y)\in G)=\iint\limits_{G}f(x,y)dxdy$$

- 对平面上任意一条曲线L，$P((X,Y)\in L)=0$

设G是平面上的一个有界区域，若二维随机变量有密度函数：

$$f(x,y)=\begin{cases}\frac{1}{m(G)},(x,y)\in G\\0,else\end{cases}$$

则称(X,Y)在G上**均匀分布**

3.2 边缘分布与独立性

设$F(x,y)$是二维随机变量的分布函数，则X与Y的**边缘分布函数**：

$$F_X(x,y)=F(x,+\infty)F_Y(x,y)=F(+\infty,y)$$

对任意的x,y，若二位分布和边缘分布满足：

$$F(x,y)=F_X(x)F_Y(y)$$

则称X与Y**相互独立**。当它们相互独立时，若$g(x),h(y)$是x,y的连续函数，则新的分布$g(X)$和$h(Y)$也是连续函数

离散型变量的边缘分布与独立性

设(X,Y)是二维离散型随机变量，二维概率分布为：

$$p_{ij}=P(X=x_i,Y=y_i)$$

此时X,Y各自的分布律：

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

称为**边缘概率分布**，简称为**边缘分布**

在上面的记号下，边缘分布可求：

$$p_{i\cdot}=\sum_j p_{ij}p_{\cdot j}=\sum_i p_{ij}$$

可推得：

$$X,Y独立\Leftrightarrow p_{ij}=p_{i\cdot}p_{\cdot j}$$

> 求离散型边缘分布的过程，相当于用全概率公式求一个概率，即表示为子概率之和

连续型变量的边缘分布与独立性

设(X,Y)是二维离散型随机变量，x,y的边缘分布分别为$F_X(x),F_Y(y)$，则：

$$F_X(x)=\int_{-\infty}^{\infty}f(x,y)dy\ F_Y(y)=\int_{-\infty}^{\infty}f(x,y)dx$$

> 边缘密度积分符号中的无穷表示在实数轴上寻找d后面符号可以被积分的区域进行积分；这个符号受到另一个符号取值的限制(如dx受到y变动的限制)，因此需要分另一个符号变动区间来构建边缘密度

此时可推得：

$$X,Y独立 \Leftrightarrow F(x,y)=F_X(x)F_Y(y)$$

在三个密度函数的公共连续点上成立

3.3 条件分布与条件密度

当X,Y不独立时，有条件概率：

$$F_{X|Y}(x|y)=P(X\le x|Y=y),x\in R$$

称为$Y=y$条件下X的**条件分布函数**

离散型条件分布

对任意给定的y_j，$P(Y=y_j)=p_{\cdot j}>0$，称：

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

为$Y=y_j$条件下X的**条件概率分布**，满足：

- $P(X=x_i|Y=y_j)\ge 0$
- $\sum_i P(X=x_i|Y=y_j)=\sum_i \frac{p_{ij}}{p_{\cdot j}}=1$

> 看成小样本除以大样本

由此条件概率可得**条件概率分布函数**：

$$F_{X|Y}(x|y_j)=P(X\le x|Y=y_j)=\sum_{x_i\le x}\frac{p_{ij}}{p_{\cdot j}}$$

由边缘分布和条件分布可得二维概率分布：

$$P(X=x_i,Y=y_j)=P(Y=y_j)P(X=x_i|Y=y_i)$$

连续型条件分布

对一个二维随机变量，若对$\forall\epsilon>0$，有$P(y-\epsilon<Y\le y+\epsilon)>0$，则：

$$\lim_{\epsilon\to 0^+}P(X\le x|y-\epsilon<Y\le y+\epsilon)$$

定义为$Y=y$条件下，X的**条件分布函数**，且可由边缘密度得到：

$$F_{X|Y}(x|y)=\int_{-\infty}^{x}\frac{f(u,y)}{f_Y(y)}du,x\in R$$

同时推出**条件密度函数**：

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

> 条件密度的范围：当条件变量取值时(两端都是实数)，条件密度中另一个变量取值(这个取值范围包括条件变量，因为随着条件变量变动而变动)

3.4 二维随机变量函数的分布

设(X,Y)是二维离散型随机变量，有联合分布律：

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

则$Z=g(X,Y)$有分布律：

$$P(Z=z_k)=\sum_{g(x_i,y_j)}p_{ij}$$

设(X,Y)是二维连续型随机变量，则$Z=g(X,Y)$是一维连续型随机变量，求密度$f_Z(z)$的步骤：

1. 确定Z的值域$R(Z)$
2. 对任意的$z\in R(Z)$，求分布函数$F_Z(z)$，即：

$$\begin{aligned}F_Z(z)&=P(Z\le z)=P(g(X,Y)\le z)\\&=P((X,Y)\in G(z))=\iint_{G(z)}f(x,y)dxdy\end{aligned}$$

注意$(X,Y)\notin R(Z)$时，有$F_Z(z)=0$或1
3. 求导解$f_Z(z)$

> 笔者总结为：先画有效区域，再画y关于x的含参z的函数，最后求z变化时，$y(x)$与有效区域的交集并积分

对于函数$Z=X+Y$的函数，我们有**卷积公式**：

1. 当X与Y不独立时，

$$\begin{aligned}f_Z(z)&=\int_{-\infty}^{\infty}f(x,z-x)dx\\&=\int_{-\infty}^{\infty}f(z-y,y)dy\end{aligned}$$

2. 当X与Y独立时，

$$\begin{aligned}f_Z(z)&=\int_{-\infty}^{\infty}f_X(x)f_Y(z-x)dx\\&=\int_{-\infty}^{\infty}f_X(z-y)f_Y(y)dy\end{aligned}$$

> 运用此公式，需要注意由z的变动范围求$f(z)$的范围，首先应当求出z关于x的不等式

3.5 *多维随机变量

设$X\sim B(n,p),Y\sim B(m,p)$，且$X$与$Y$相互独立，则：

$$Z=X+Y\sim B(n+m,p)$$

且有：

$$P(Z=k)=\sum_{r=0}^{k}P(X=r)P(Y=k-r)$$

这个公式叫做**离散卷积公式**

> 上面的结论可以推到可数个伯努利分布相加

设$X\sim P(\lambda_1),Y\sim P(\lambda_2)$，且$X$与$Y$相互独立，则：

$$Z=X+Y\sim P(\lambda_1+\lambda_2)$$

> 上面的结论可以推到可数个泊松分布相加

设$X_i\sim \Gamma(\alpha_i)$，且X_i相互独立，则：

$$Z=\sum_{i=1}^{n}X_i\sim \Gamma(\sum_{i=1}^{n}\alpha_i,\beta)$$

极值分布函数分为极大值分布函数和极小值分布函数。设X_i相互独立，记$M=max\{X_i\}$，$N=min\{X_i\}$，则M与N分布：

$$F_M(x)=\prod_{i=1}^{n}F_i(x)F_N(x)=1-\prod_{i=1}^{n}[1-F_i(x)]$$

同分布时：

$$F_M(x)=F^n(x)F_N(x)=1-[1-F(x)]^n$$

第四章 随机变量的数字特征

4.1 数学期望

设离散型随机变量X的概率分布：

$$P(X=x_k)=p_k$$

若级数$\sum_{k=1}^{\infty}x_kp_k$绝对收敛，则称这个级数是随机变量$X$的**数学期望**，简称**期望**，记为：

$$E(X)=\sum_{k=1}^{\infty}x_kp_k$$

数学期望是x_i的加权平均，也被称为**均值**。若上述级数不收敛，则称没有数学期望

设连续型随机变量X的密度$f(x)$，若反常积分$\int_{-\infty}^{\infty} xf(x)dx$绝对收敛，则称这个积分是$X$的数学期望记为：

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

可以证明：

- 若$X \sim B(n,p)$，则$E(X) = p$
- 若$X \sim \Gamma(\alpha,\beta)$，则$E(X) = \frac{\alpha}{\beta}$。特别地，若$X \sim e(\lambda)$，则$E(X) = \frac{1}{\lambda}$

对于离散型随机变量，设$Y = g(X)$，则：

$$E(Y) = E(g(X)) = \sum_{i=1}^{\infty} g(x_i)pi$$

对于连续型随机变量：

$$E(Y) = \int_{-\infty}^{\infty} g(x)f(x)dx$$

> 此时$E(Y)$是关于x取值变化的分段积分
> 对于二维离散型随机变量：
>
> $$E(Z) = E(g(X,Y)) = \sum_i \sum_j g(x_i,y_j)p_{ij}$$

对于二维连续型随机变量：

$$E(Z) = E(g(X,Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy$$

数学期望的性质有：

- $E(C) = C$
- 线性性质 $E(\sum_{i=1}^{n} C_iX_i + b) = \sum_{i=1}^{n} C_iE(X_i) + b$
- 若X与Y相互独立，则：

$$E(XY) = E(X)E(Y)$$

> 这个性质可以推广到可数个随机变量

4.2 方差

若期望$E(X - E(X))^2$存在，则称为X的**方差**：

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

且称$\sqrt{D(X)}$称为X的**均方差**或**标准差**

对于离散型随机变量：

$$D(X)=\sum_{k=1}^{\infty}[x_k-E(X)]^2p_k$$

对于连续型随机变量：

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

实际计算时，常用的公式为：

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

特别地，当$E(X)=0$时，$D(X)=E(X^2)$

可以证明：

- 若$X\sim B(n,p)$，则$D(X)=np(1-p)$
- 若$X\sim P(\lambda)$，则$D(X)=\lambda$
- 若$X\sim \Gamma(\alpha,\beta)$，则$D(X)=\frac{\alpha}{\beta^2}$。特别地，若$X\sim e(\lambda)$，则$D(X)=\frac{1}{\lambda^2}$

方差的性质有：

- $D(C)=0$
- 若X,Y独立，则$D(X\pm Y)=D(X)+D(Y)$
- $D(\sum_{i=1}^{n}C_iX_i+b)=\sum_{i=1}^{n}C_i^2D(X_i)$
- $D(X)=0$的充要条件为，存在一个C使得$P(X=C)=1$，且$C=E(X)$。当$P(X=C)=1$时，称X服从**退化分布**

变异系数、矩和中心距

若随机变量的期望和方差均存在且期望不为零，则称：

$$C_v=\frac{\sqrt{D(X)}}{|E(X)|}$$

为X的**变异系数**，反映X在均值附近的相对集中程度

若随机变量X对非负整数k有下列期望存在：

$$m_k=E(X^k)$$

称m_k为X的k阶**原点矩**，期望是一阶原点矩

若有：

$$\mu_k=E[X-E(X)]^k$$

称μ_k为X的k阶**中心距**

4.3 协方差和相关系数

对于二维随机变量(X,Y)，若$E[[X-E(X)][Y-E(Y)]]$存在，则称其为X与Y的**协方差**，记为：

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

特别地，$Cov(X,X)=D(X)$，且：

$$Cov(X,Y)=E(XY)-E(X)E(Y)$$

协方差的性质有：

- $Cov(X,Y)=Cov(Y,X)$
- $Cov(X,a)=0$
- $Cov(aX,bY)=abCov(X,Y)$
- $Cov(X+Y,X)=Cov(X,Z)+Cov(Y,Z)$
- $D(X\pm Y)=D(X)+D(Y)\pm 2Cov(X,Y)$
- $D(aX\pm bY)=a^2D(X)+b^2D(Y)+2abCov(X,Y)$ 这个公式是方差的线性组合公式
- 若X与Y相互独立，则$Cov(X,Y)=0$

对于二维随机变量，称向量$(E(X),E(Y))$为其**均值向量**，称矩阵：

$$\mathbf{V}=\begin{pmatrix}D(X) & Cov(X,Y)\\ Cov(Y,X) & D(Y)\end{pmatrix}$$

为二维变量的**协方差阵**。推广到n维后，可以发现协方差阵是一个对阵矩阵

设随机变量X与Y的期望和方差存在，且方差均大于0，则称：

$$X^*=\frac{X-E(X)}{\sqrt{D(X)}}$$

为X的**标准化随机变量**，称：

$$Cov(X^*,Y^*)=R(X,Y)$$

为X,Y的**相关系数**

相关系数的性质有：

- $E(X^*)=0$
- $D(X^*)=1$
- $|R(X,Y)|=1$的充要条件是存在常数a, b, $a\neq 0$，使得$P(Y=aX+b)=1$

当$R(X,Y)=1$时，称二者**完全线性相关**。对于上述的性质3可知，a的正负反映了二者的正负相关性；$|R(X,Y)|$越大，说明相关性越强；特别地当$R(X,Y)=0$时，称二者**不相关**

> 不相关和不独立时两个概念。独立的两个随机变量一定不相关，不相关的两个随机变量不一定独立，因为相关性反映的是二者的线性关系

第五章 正态分布

5.1 正态分布及其密度与分布

若随机变量X的密度：

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},x\in R$$

则称X服从**标准正态分布**，记为$X\sim N(0,1)$，其分布函数记为：

$$\Phi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$$

$\Phi(x)$的性质包括：

- 是偶函数，$\Phi(-x)=1-\Phi(x)$
- $\Phi(0)=\frac{1}{2}$

> $\Phi(x)$不能表达为初等函数

设$\mu,\sigma>0$是任意常数，若随机变量X满足：

$$Z=\frac{X-\mu}{\sigma}\sim N(0,1)$$

则称X服从参数μ,σ^2的**正态分布**，记为$X\sim N(\mu,\sigma^2)$

$X\sim N(\mu,\sigma^2)$满足：

- 分布函数

$$F(x)=\Phi(\frac{x-\mu}{\sigma})$$

- 计算概率

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

- 概率密度函数

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

注意到概率密度函数的图像满足：曲线对称与$x=\mu$，顶点$max\{f(x)\}=\frac{1}{\sqrt{2\pi}\sigma}$，曲线的渐近线是$x$轴且拐点为$\mu\pm\sigma$。因此，称$\mu$为**位置参数**，确定对称轴位置；称$\sigma$为**刻度参数**，$\sigma$越小，曲线越高瘦

> 实际应用中，常把区间$(\mu-3\sigma,\mu+\sigma)$为X的实际取值区间，这就是**3σ原则**

对于一个正态分布的随机变量，其期望和方差分别为：

$$E(X)=\mu D(X)=\sigma^2$$

注意到当$b\neq 0$时：

$$Y=a+bX\sim N(a+b\mu,b^2\sigma^2)$$

且对于独立的$X\sim N(\mu_1,\sigma_1^2),Y\sim N(\mu_2,\sigma_2^2)$，有：

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

由此得到正态分布的**线性性质**，设X_i相互独立，$X_i\sim N(\mu_i,\sigma_i^2)$，有：

$$Z=\sum_{i=1}^{n}C_iX_i\sim N(\sum_{i=1}^{n}C_i\mu_i,\sum_{i=1}^{n}C_i^2\sigma_i^2)$$

这个性质叫做**正态分布可加性**

> 此性质常用于求正态分布的线性函数

5.2 二维正态分布

设随机变量(X,Y)有二维密度：

$$\begin{aligned}f(x,y)=&\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}exp\{-\frac{1}{2(1-r^2)}[\frac{(x-\mu_1)^2}{\sigma_1^2}-\\&2r\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}]\},\quad(x,y)\in R\end{aligned}$$

则称其服从二维正态分布，记为$(X,Y)\sim N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;r)$，其中$|r|<1$等为分布参数

对于一个二维正态分布，求边缘分布：

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

且我们可以得到：

$$\mu_1=E(X),\mu_2=E(Y)\sigma_1^2=D(X),\sigma_2^2=D(Y)r=R(X,Y)$$

在二维正态分布中，X与Y独立的充要条件是$r=0$

> 在两个一维分布中，$r=0$不能推出这两个分布独立
> 此结论常结合协方差公式求解

二维正态分布的条件分布也是正态分布：

$$f_{X|Y}(x|y)\sim N(\mu_1+r\frac{\sigma_1}{\sigma_2}(y-\mu_2),\sigma_1^2(1-r^2))$$

二维随机变量服从二维正态分布的充要条件是X与Y的任意非零线性组合$Z=aX+bY$服从一维正态分布

*5.3 自然指数分布族

若存在$H \in R$上的实值函数$\varphi(\theta)$，以及不依赖于θ的函数$h(x)$，非退化的随机变量X有概率分布或概率密度函数：

$$\varphi(x,\theta) = exp\{\theta x - \varphi(\theta\}h(x), x \in G, \theta \in H$$

则称X服从**自然指数分布族分布**，其中θ叫做**自然参数**，H叫做**自然参数空间**，$\varphi(\theta)$叫做**累积量母函数**，G叫做**支撑集**且不依赖于θ

若X服从自然指数分布族分布，则：

$$E(X) = \varphi'(\theta), D(X) = \varphi''(\theta), \theta \in H$$

注意到记$E(X) = m = \varphi'(\theta)$，此时$m$与$\theta$有一一对应关系，则记$m$为**均值参数**，其取值为**均值空间**，记为M；于是$D(X)$也是m的函数，记为$D(X) = \varphi''(\theta)|_{(\theta=\theta(m)} = V(m)$

若X_i独立同服从于一个自然指数分布族分布，其方差函数$V(m)$，则：

$$Y = \sum_{i=1}^{n} X_i$$

也服从同一个自然指数分布族分布，且：

$$E(Y) = nm, D(Y) = nV(m)$$

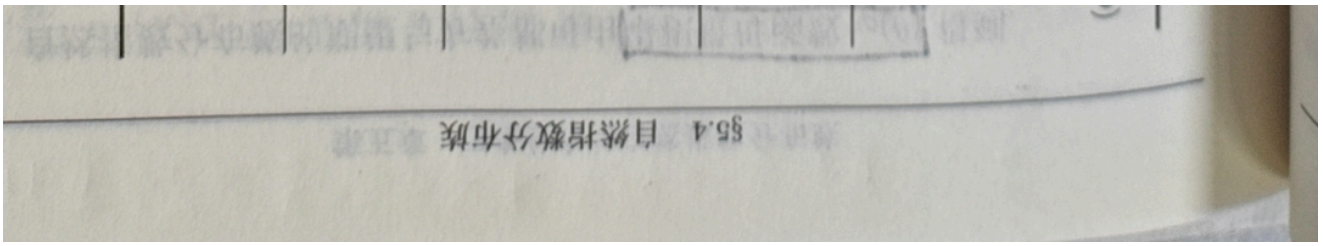
其中$m = E(X_i)$

表 5.1　常见自然指数分布族分布

	正态分布 $N(\mu,\sigma^2)$	泊松分布 $P(\lambda)$	Γ 分布 $\Gamma(\alpha,\beta)$	二项分布 $B(n,p)$	负二项分布 $NB(r,p$
密度函数或概率分布	$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $-\infty<x<+\infty$ $-\infty<\mu<+\infty$	$\frac{\lambda^x e^{-\lambda}}{x!}$ $x=0,1,2,\cdots$ $\lambda>0$	$\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$ $x>0$ $\beta>0$	$\binom{n}{x}p^x q^{n-x}$ $x=0,1,\cdots,n$ $0<p<1$	$\binom{r-1}{x-1}p^r q^{x-r}$, $x=r,r+1,\cdots$ $0<p<1$
自然参数 θ	$\frac{\mu}{\sigma^2}$	$\ln\lambda$	$-\beta$	$\ln\frac{p}{q}$	$\ln q$
累积量母函数 $\varphi(\theta)$	$\frac{\sigma^2\theta^2}{2}$	e^θ	$-\alpha\ln(-\theta)$ $=-\alpha\ln\beta$	$n\ln(1+e^\theta)$ $=-n\ln q$	$r\ln\left(\frac{e^\theta}{1-e^\theta}\right)$ $=r\ln\frac{q}{p}$
均值参数 $m=\varphi'(\theta)$	$m=\mu=\theta\sigma^2$	$m=\lambda=e^\theta$	$m=\frac{\alpha}{\beta}=-\frac{\alpha}{\theta}$	$m=np=\frac{n}{1+e^{-\theta}}$	$m=\frac{r}{p}=\frac{r}{1-e^\theta}$
方差函数 $V(m)=\varphi''(\theta)$	$\sigma^2=\sigma^2 m^0$	$\lambda=m$	$\frac{\alpha}{\beta^2}=\frac{m^2}{\alpha}$	$npq=-\frac{m^2}{n}+m$	$\frac{rq}{p^2}=\frac{m^2}{r}-m$

注: 1. 作为自然指数分布族分布, 正态分布的参数 σ^2, Γ 分布的参数 α 是作为已知的, 且 $\alpha=1$ 时是指数分布 $e(\beta)$.

2. 负二项分布又叫做帕斯卡分布, 当 $r=1$ 时, 是几何分布 $G(p)$.



第六章 极限定理

6.1 大数律

随机变量的方差$D(X)$刻画了随机变量的取值集中在$E(X)$附近的程度，因此对$\forall \varepsilon > 0$，$P(|X - E(X)| < \varepsilon)$与$D(X)$有关。可以发现$D(X)$越小，这个概率就越大

切比雪夫不等式：设随机变量X的数学期望和方差都存在，对$\forall \varepsilon > 0$，有：

$$P(|X - E(X)| < \varepsilon) \geq 1 - \frac{D(X)}{\varepsilon^2}$$

或有：

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{D(X)}{\varepsilon^2}$$

这两个等价的不等式被称为切比雪夫不等式

设$\{X_n\}$是随机变量序列(一系列随机变量取值的集合)，a为常数，若对$\forall \varepsilon > 0$，有：

$$\lim_{n \to \infty} P(|X_n - a| \geq \varepsilon) = 0$$

或

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

则称X_i**依概率收敛**于a，记为$X_n \xrightarrow{P} a$

由切比雪夫不等式可知，在随机变量序列中，若$E(X_n) = \mu_n$，$D(X_n) = \sigma_n^2$存在，且当$n \to \infty$时，有$\sigma_n^2 \to 0$，则：

$$X_n - \mu_n \xrightarrow{P} 0$$

在一个随机变量序列中,若其中任意有限个随机变量都相互独立，则称其为一个**独立的随机变量序列**

> 可以看做是每次试验都是独立的

对一个随机变量序列$\{X_k\}$，记$\overline{X} = \frac{1}{n}\sum X_k$，若$\overline{X} - \frac{1}{n}\sum X_k \xrightarrow{P} 0$，则称此随机变量序列服从**大数律**

切比雪夫大数律：对独立随机变量序列，若其均值和方差均存在，且有常数C，使得$D(X_k) \leq C$，则有：

$$\frac{1}{n}\sum_{k=1}^{n}X_k-\frac{1}{n}\sum_{k=1}^{n}E(X_k)\xrightarrow{P}0$$

独立同分布大数律：设一个随机变量序列$\{X_k\}$是独立同分布的，且$E(X_k)=\mu, D(X_k)=\sigma^n$，则：

$$\overline{X}\xrightarrow{P}\mu$$

> 此大数律说明多次重复试验下，总体期望可以用样本期望估计

伯努利大数律：在n次伯努利试验下，事件A发生的频率为$f_n(A)=\frac{n_A}{n}$。设其发生的概率为$p=p(A)$，则：

$$f_n(A)\xrightarrow{P}p=P(A)$$

> 此大数律说明频率在实验次数足够多时可以用来估计概率

6.2 中心极限定理

林德伯格-列维中心极限定理：设随机变量序列独立同分布，且$E(X_k)=\mu, D(X_k)=\sigma^n$，记：

$$Y_n=\frac{\sum_{k=1}^{n}X_k-n\mu}{\sqrt{n}\sigma}=\frac{\frac{1}{n}\sum_{k=1}^{n}X_k-\mu}{\sigma/\sqrt{n}}$$

则对$\forall x\in R$，有：

$$\lim_{n\to\infty}F_n(x)=\lim_{n\to\infty}P(Y_n\le x)=\Phi(x)$$

注意到：

$$E(\frac{1}{n}\sum_{k=1}^{n}X_k)=\frac{1}{n}E(\sum_{k=1}^{n}X_k)=n\mu D(\frac{1}{n}\sum_{k=1}^{n}X_k)=\frac{1}{n^2}D(\sum_{k=1}^{n}X_k)=\frac{\sigma^2}{n}$$

因此Y_n是$\frac{1}{n}\sum_{k=1}^{n}X_k$的标准化随机变量

> 上述定理又称为**独立同分布的中心极限定理**，表明Y_n的分布函数的极限函数是标准正态分布函数

独立同分布的中心极限定理的应用形式是，在n充分大时，近似地：

$$\frac{1}{n}\sum_{k=1}^{n}X_k\sim N(\mu,\frac{\sigma^2}{n})$$

> 此应用形式说明，在统计量足够大时，多次试验的均值是一个正态分布，且方差为$\frac{\sigma^2}{n}$

棣莫弗-拉普拉斯中心极限定理：设随机变量序列$\{X_n\}$中，$X_n\sim B(n,p)$，则$\forall x\in R$，有：

$$\lim_{n\to\infty}P(\frac{X_n-np}{\sqrt{p(1-p)}}\le x)=\Phi(x)$$

> 此定理说明，当n足够大时，二项分布可用正态分布近似

应用形式为n充分大时，有：

$$X_n \sim N(np, np(1-p))$$

> 由应用形式可以求X介于a, b之间的概率，使用正态分密度公式

注意：

- 二项分布的泊松分布需要$p \leq 0.1$，但正态近似不需要
- n充分大被认为是$n \geq 50$
- 当n充分大时，忽略$P(X=a)$和$P(X=b)$，因此用正态分布密度计算时不考虑端点

第七章 数理统计基础

7.1 总体与样本

总体：全体被研究的对象
样本：每个被研究的对象

设随机变量X_i与总体X同分布，则称$X_1, X_2, \cdots, X_n$为来自筒体的样本容量为n的**简单随机样本**，简称**样本。X_i的取值x_i被称为样本观测值

当样本是离散型随机变量时，分布律：

$$P(X_1=x_1, X_2=x_2, \cdots, X_n=x_n) = p(x_1)p(x_2)\cdots p(x_n) = \prod_{i=1}^{n} p(x_i)$$

当样本是连续型随机变量时，密度函数：

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

7.2 样本的分布

7.2.1 χ^2分布

设随机变量序列X_i独立同分布于标准正态分布$N(0,1)$，称随机变量：

$$\chi^2 = \sum_{i=1}^{n} X_i^2$$

所服从的分布为自由度为n的 χ^2**分布**，记为$\chi^2 \sim \chi^2(n)$

注意到，χ^2分布也是$\alpha = \frac{n}{2}, \beta = \frac{1}{2}$的$\Gamma(\frac{n}{2}, \frac{1}{2})$分布，所以$\chi^2$分布的密度函数：

$$f(x,n) = \begin{cases} \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

因此，$E(\chi^2) = n, D(\chi^2) = 2n$

设相互独立的$\chi_1^2 \sim \chi^2(n), \chi_2^2 \sim \chi^2(m)$，则：

$$\chi_1^2 + \chi_2^2 \sim \chi^2(n+m)$$

> 这是χ^2分布的线性性质

7.2.2 t分布

设随机变量$X \sim N(0,1)$，$Y \sim \chi^2(n)$，且X, Y相互独立，称随机变量：

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

服从的分布是自由度为n的**t分布**，记为$t \sim t(n)$，t分布又称学生氏分布。它的分布函数为：

$$t(x,n) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(1+\frac{x^2}{n})^{-\frac{n+1}{2}}$$

可以证明：

$$\lim_{n\to\infty} t(x,n) = \Phi(x)$$

其中$\Phi(x)$是标准正态分布的密度函数。事实上，当$n \geq 45$时，t分布就接近正态分布了

7.2.3 F分布

设随机变量$X \sim \chi^2(n), Y \sim \chi^2(m)$，且$X, Y$独立，称随机变量：

$$F = \frac{X/n}{Y/m}$$

服从的分布为自由度为(n,m)的**F分布**，记为$F \sim F(n,m)$，其中n和m分别为第一、第二自由度

由定义得F分布的密度：

$$f(x,n,m) = \begin{cases} \frac{\Gamma(\frac{n+m}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})}\frac{n}{m}^{\frac{n}{2}}x^{\frac{n}{2}-1}(1+\frac{n}{m}x)^{-\frac{n+m}{2}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

> 注意，所有的分布都要求随机变量之间独立，否则不能构成此分布

7.2.4 分位点

设随机变量X，概率$0 < p < 1$，若实数a_p满足：

$$F(a_p) = P(X \leq a_p) = p$$

则称a_p为X的**p分位点**，当$p = 1/2$时，$a_{\frac{1}{2}}$为**中位数**

> a_p是一个数量，是用于统计的一个实数

当$n > 45$时，χ^2分布的p分位点有近似：

$$\chi_p^2(n) \approx \frac{1}{2}(u_p + \sqrt{2n-1})^2$$

其中u_p是标准正态分布的p分位数

对于$t \sim t(n)$，其分位点在$n > 45$时有：

$$t_p(n) \approx u_p$$

对于$F(n,m)$，满足：

$$F_p(n,m) = \frac{1}{F_{1-p}(m,n)}$$

7.3 统计量和抽样分布定理

7.3.1 统计量

样本X_i的一个连续函数$g(X_i)$称为**样本函数**，若其不含任何未知参数，则称其为一个**统计量**，代入样本观测值后的$g(x_i)$称为**统计量的观测值**

常用的统计量有：

- 样本均值

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- 样本方差

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

- 样本k阶原点矩

$$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

- 样本k阶中心距

$$B_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k$$

7.3.2 抽样分布定理

一个正态总体下

1. 定理一
   样本X_i来自正态总体$N(\mu, \sigma^2)$，则

$$\overline{X} \sim N(\mu, \frac{\sigma^2}{n}) \ \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

由此推出对于来自任意总体的样本X_i，有：

$$E(\overline{X}) = E(X) \ D(\overline{X}) = \frac{D(X)}{n}$$

2. 定理二
   样本X_i来自正态总体$N(\mu, \sigma^2)$，则

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1) \ \overline{X}与S^2相互独立$$

   其中$\frac{(n-1)S^2}{\sigma^2}$满足：

$$\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\overline{X})^2 = \frac{nB_2}{\sigma^2}$$

3. 定理三
   样本X_i来自正态总体$N(\mu, \sigma^2)$，则：

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

由于t分布的极限是标准正态分布，因此n充分大时此定理可看做上式t满足标准正态分布

4. 定理四

对任何总体X，记$E(X)=\mu$，$D(X)=\sigma^2>0$，设样本X_i，当n充分大时，近似有：

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

> 上式1式是中心极限定理的等价描述

两个正态总体下

下面，设X与Y是两个总体，其两个独立样本X_i和Y_i的样本均值和方差表示为$\overline{X}, \overline{Y}$和$S_1^2, S_2^2$

1. 定理一
   设两个总体$X \sim N(\mu_1, \sigma_1^2)$，$Y \sim N(\mu_2, \sigma_2^2)$，则统计量：

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

   从而统计量U满足：

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

2. 定理二
   设两个总体满足各自的正态分布，则：

$$T=\frac{\overline{X}-\overline{Y}-(\mu_1,-\mu_2)}{S_\omega\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$$

   其中的：

$$S_\omega^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

3. 定理三

设两个总体满足各自的正态分布，则：

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

第八章 参数估计

8.1 点估计

设总体X的分布函数为$F(x,\theta)$， 其中的θ是未知参数。从总体抽取的样本$X_1,X_2,\cdots,X_n$的观测值$x_1,x_2,\cdots,x_n$可用于构造某个统计量$\hat{\theta}(X_1,X_2,\cdots,X_n)$，观测值为$\hat{\theta}(x_1,x_2,\cdots,x_n)$可估计参数$\theta$。称：

$$\hat{\theta}(x_1,x_2,\cdots,x_n)$$

为θ的**估计值**，称：

$$\hat{\theta}(X_1,X_2,\cdots,X_n)$$

为θ的**估计量**

上面的定义中，估计值和估计量都是θ的**点估计**。对于有k个未知量的分布函数，记$\theta=(\theta_1,\theta_2,\cdots,\theta_k)$为一个$k$维向量，可构造$k$个统计量来进行点估计，这$k$个统计量分别是$\theta_1,\theta_2,\cdots,\theta_k$的**点估计量**

8.1.1 矩估计法

设总体 X 的分布函数为 $F(x,\theta)$，其中的 θ 是一维未知参，若 $\exists E(X)$，则 $m=E(X)$ 一般是关于 θ 的函数，即 $m=m(\theta)$。由此，反解 $\theta=g(m)$，再用样本均值 $\overline{X}$ 代替 m，可得到 θ 的一个估计量 $g(\overline{X})$。这个方法就是**矩估计**，$\hat{\theta}=g(\overline{X})$ 是 θ 的**矩估计量**

对于k维未知量$\theta=(\theta_1,\theta_2,\cdots,\theta_k)$，若其$k$阶原点矩存在，则可用$k$阶原点矩构造$k$个矩估计量，最后解一个$k$元未知数方程得到$\theta$

当$k=2$时，不难算出：

$$\begin{cases}\mu=m_1\\ \sigma^2=m_2-m_1^2\end{cases}$$

若用样本前二阶原点矩代替总体原点矩，得到矩估计量：

$$\begin{cases}\hat{\mu}=\overline{X}\\ \hat{\sigma}^2=A_2-\overline{X}^2=B_2\end{cases}$$

> 上式中，估计σ^2的是样本的二阶中心矩不是样本方差。因此，这是一个有偏估计量

参数θ的矩估计量可以是不同的；在样本容量n较小时，矩估计的误差较大

8.1.2 极大似然估计法

设总体X是离散型随机变量，对于样本$X_1,X_2,\cdots,X_n$，得到一组观测值$x_1,x_2.\cdots,x_n$。由独立同分布性，获得这组观测值的概率是：

$$\begin{aligned}&P(X_1=x_1,X_2=x_2,\cdots,X_n=x_n)\\=&P(X_1=x_1)P(X_2=x_2)\cdots P(X_n=x_n)\\=&\prod_{i=1}^{n}p(x_i,\theta)=L(\theta)\end{aligned}$$

称这个函数$L(\theta)$是**似然函数**

根据小概率事件原理，我们认为出现上式结果的概率应当做大，因此因该选择一个估计值θ，使得：

$$L(\hat{\theta})=max\{L(\theta)\}$$

当X是连续型随机变量时，同样有似然函数：

$$L(\theta)=\prod_{i=1}^{n}f(x_i,\theta)$$

设仅含一个未知数θ的总体X，在已知分布律或密度的前提下，设观测值x_i，若存在θ的一个值$\hat{\theta}(x_1,x_2,\cdots,x_n)=\hat{\theta}$，使得：

$$L(\hat{\theta})=max\{L(\hat{\theta})\}$$

则称$\hat{\theta}$是θ的**极大似然估计值**，统计量$\hat{\theta}(X_1,X_2,\cdots,X_n)$是$\theta$的**极大似然估计量**

由定义知，极大似然估计值可由方程：

$$\frac{dL(\theta)}{d\theta}=0$$

或

$$\frac{d(\ln L(\theta))}{d\theta}=0$$

取得。当上式无解时，可以从定义考虑求极大似然解，即从$\hat{\theta}$中选择一个值使得上式成立

可以证明，当X服从单峰分布时，若上式有解，则这个解就是θ的极大似然估计值

> 除均匀分布，常见的分布都是单峰分布

计算得出，正态总体的极大似然估计为：

- 均值 $\hat{\mu}=\overline{X}$
- 方差 $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2=B_2$

若X服从自然指数分布族分布，则均值$m=E(X)$的极大似然估计量就是样本均值$\overline{X}$；除二项分布外，X的方差函数$V(m)$的极大似然估计量都是$V(\overline{X})$
当当
若X服从自然指数分布族分布，则均值$m=E(X)$的极大似然估计量就是样本均值$\overline{X}$；除二项分布外，X的方差函数$V(m)$的极大似然估计量都是$V(\overline{X})$，θ含有多个未知参数，可以考虑对每个未知参数求偏导数，再解方程取得极大似然估计量

8.2 统计量选择标准

统计量的评估标准包括：

- 无偏性标准
- 有效性标准
- 一致性标准

8.2.1 无偏性标准

若未知参数θ的估计量$\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$，有：

$$E(\hat{\theta})=\theta$$

则称$\hat{\theta}$是θ的**无偏估计量**，否则是**有偏估计量**

对有偏估计量，称：

$$b(\hat{\theta})=E(\hat{\theta})-\theta$$

为$\hat{\theta}$的**偏差**。若样本容量$n\to\infty$时，有$b(\hat{\theta})\to 0$，则称$\hat{\theta}$是θ的**渐进无偏估计量**

8.2.2 有效性标准

对于两个无偏估计量$\hat{\theta}_1$和$\hat{\theta}_2$，若：

$$D(\hat{\theta}_1) \leq D(\hat{\theta})$$

则称θ_1比θ_2更有效

> 可以证明，样本均值比其他任何一个线性无偏估计都有效

8.2.3 一致性标准

若$\hat{\theta}$是θ的估计量，若：

$$\hat{\theta}_n \to \theta$$

则称$\hat{\theta}_n$是θ的**一致估计量**

> 样本方差是总体方差的一致估计量

8.2.4 均方误差标准

估计量$\hat{\theta}$的一个统计量：

$$M(\hat{\theta}) = E(\hat{\theta} - \theta)^2$$

称为$\hat{\theta}$关于θ的**均方误差**。若两个统计量满足：

$$M(\hat{\theta}_1) \leq M(\hat{\theta}_2)$$

则称$\hat{\theta}_1$在均方误差下比$\hat{\theta}_2$更有效

不难发现：

$$M(\hat{\theta}) = D(\hat{\theta}) + (E(\hat{\theta}) - \theta)^2$$

因此有效性标准是均方误差标准的特殊情况。均方误差标准用于衡量两个有偏估计量或一个无偏一个有偏估计量之间的比较

8.3 区间估计

8.3.1 置信区间

设一个总体X的分布函数$F(X, \theta)$，其中X_i是一个样本，$\hat{\theta}_1$和$\hat{\theta}_2$是样本构造的两个统计量。若对于给定的概率$1-\alpha$，有：

$$P(\hat{\theta}_1 < \theta < \theta_2) = 1 - \alpha$$

则称随机区间$\hat{\theta}_1, \hat{\theta}_2$是参数$\theta$的置信度为$1-\alpha$的**置信区间**。$\hat{\theta}_1$和$\hat{\theta}_2$分别称为**置信下限**和**置信上限**，$1-\alpha$称为**置信度**或**置信水平**

求置信区间的步骤：

1. 找到一个总体的样本，取θ的一个无偏估计量(习惯上)

2. 通过$\hat{\theta}$出发，找一个样本函数$W=W(\hat{\theta})$，且分布已知，只含有一个参数θ，找到W的分位点
3. 查表，找到$\frac{\alpha}{2}$和$1-\frac{\alpha}{2}$分位点W_1和W_2，使得$P(W_{\frac{\alpha}{2}}<W<W_{1-\frac{\alpha}{2}})=1-\alpha$
4. 从上面的不等式解出等价不等式$\hat{\theta}_1<\theta<\hat{\theta}_2$，此时就能找到置信区间
5. 代入样本观测值，得到实数置信区间

这样求得的置信区间是**双侧置信区间**，还可以求**单侧置信区间**

8.3.2 一个正态总体下参数置信区间

设总体$X\sim N(\mu,\sigma^2)$，样本$X_1,X_2,\cdots,X_n$

8.3.2.1 已知$\sigma^2=\sigma_0^2$，求均值μ的置信区间

因为$\overline{X}$是μ的无偏估计，取样本函数$U=\frac{\overline{X}-\mu}{\sigma_0/\sqrt{n}}$，且：

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

可以求得对于给定的置信度，有置信区间：

$$(\overline{X}-u_{1-\frac{\alpha}{2}}\frac{\sigma_0}{\sqrt{n}},\quad \overline{X}+u_{1-\frac{\alpha}{2}}\frac{\sigma_0}{\sqrt{n}})$$

8.3.2.2 σ^2未知，求均值μ的置信区间

不知道总体方差σ^2，我们考虑用样本方差S^2代替，此时构造：

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

此时得到置信区间：

$$(\overline{X}-t_{1-\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}},\quad \overline{X}+t_{1-\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}})$$

8.3.2.3 μ未知，求方差σ^2置信区间

因为S^2是σ^2的无偏估计，考虑：

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

因此得到σ^2的置信区间：

$$(\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)},\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)})$$

> 求σ的置信区间，应该开根号

8.3.3 两个正态总体下参数的置信区间

下面的假设基于两个正态分布总体X,Y，他们各自取独立样本，并有均值和方差

8.3.3.1 已知σ_1^2,σ_2^2，求均值差$\mu_1-\mu_2$的置信区间

此时，$\overline{X}-\overline{Y}$是$\mu_1-\mu_2$的无偏估计，考虑：

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

得到置信区间：

$$(\overline{X}-\overline{Y}-u_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}},\quad \overline{X}-\overline{Y}+u_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}})$$

8.3.3.2 σ_1^2,σ_2^2未知但相等，求$\mu_1-\mu_2$的置信区间

考虑：

$$T=\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$$

其中：

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

此时得到置信区间为：

$$(\overline{X}-\overline{Y}-t_{1-\frac{\alpha}{2}}S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}},\quad \overline{X}-\overline{Y}+t_{1-\frac{\alpha}{2}}S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}})$$

8.3.3.3 两个总体的μ和σ^2都未知，求方差比$\frac{\sigma_1^2}{\sigma_2^2}$的置信区间

由抽样分布定理知：

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

得到置信区间：

$$(\frac{S_1^2}{S_2^2F_{1-\frac{\alpha}{2}}},\quad \frac{S_1^2}{S_2^2F_{\frac{\alpha}{2}}})$$

8.3.4 自然指数分布族均值参数的置信区间

设服从某个分布$F(X)$的总体随机变量X，均值$E(X)=m(\theta)$，方差$D(X)=\sigma^2(\theta)$，由中心极限定理可得($n>50$)：

$$\begin{aligned}U&=\frac{\sum_{i=1}^{n}X_i-nm(\theta)}{\sqrt{n}\sigma(\theta)}\\&=\frac{\overline{X}-m(\theta)}{\sigma(\theta)/\sqrt{n}}\end{aligned}$$

当这个分布服从自然指数分布族分布时，置信区间：

$$P(|U|<u_{1-\frac{\alpha}{2}})=P(|\frac{\overline{X}-m(\theta)}{\sigma(\theta)/\sqrt{n}}|<u_{1-\frac{\alpha}{2}})\approx 1-\alpha$$

上式不好解，用数学证明当n充分大时，$\frac{\overline{X}-m(\theta)}{V(\overline{X})/\sqrt{n}}$服从标准正态分布，据此构造得到置信区间：

$$(\overline{X}-u_{1-\frac{\alpha}{2}}\sqrt{\frac{V(\overline{X})}{n}},\overline{X}+u_{1-\frac{\alpha}{2}}\sqrt{\frac{V(\overline{X})}{n}})$$

8.3.5 单侧置信区间

我们有时只关心未知参数是否高于或低于某个阈值，此时构造的置信区间是**单侧置信区间**，其中：

$$P(\theta>\overline{\theta}_1)=1-\alpha$$

是**单侧置信下限**，而：

$$P(\theta<\overline{\theta}_2)=1-\alpha$$

是**单侧置信上限**

第九章 假设检验

9.1 假设检验的基本概念

统计学中的一些问题常常需要对总体进行抽样得到样本，然后通过分析样本的某些统计量，判断总体的这个统计量是否满足某个特征。我们可以用**假设检验**的方法来完成

> 假设检验就像是数学中的反证法

提出的假设：

$$H_0:\theta=\theta_0,H_1:\theta\neq\theta_0$$

其中H_0被称为**原假设**，H_1被称为**备择假设**

现在，我们需要构造一个统计量，使得我们接受或拒绝原假设。可以考虑构造这样的统计量，使得它满足某个分布。若在一个小概率下，这个统计量没有发生，我们认为原假设成立，否则

不成立。这个小概率被称为**显著性水平**

假设检验又叫**显著性检验**，其中U是**检验统计量**，$u_{1-\alpha}$是**临界值**，$W=|U|>u_{1-\alpha}$是检验的**拒绝域**

> 当我们的构造统计量落在拒绝域，说明小概率事件发生了，应当拒绝H_0

当拒绝域位于原假设θ_0两端时，这种检验叫做**双侧检验**；相对应地，有**左侧检验**和**右侧检验**

当我们进行假设检验时，会出现两类错误。第一类错误叫**弃真**，即H_0正确时，否定了H_0而选择H_1；第二类错误叫**取伪**，即H_0错误时，没有拒绝H_0。两类错误不能同时减少，必须根据实际选择优先减少哪类错误

假设检验的基本步骤：

1. 选取合适统计量，这个统计量满足某个分布
2. 选定显著性水平，获取在此分布下的显著性水平分位数
3. 计算统计量，与这个显著性水平的分位数比较，看是否落在拒绝域

9.2 正态总体下的假设检验

进行假设检验时，我们的思路同反证法。根据H_1中的不等号来决定统计量的不等号

9.2.1 一个正态总体的参数检验

9.2.1.1 $\sigma^2=\sigma_0^2$，μ的检验

此时构造：

$$U=\frac{\overline{X}-\mu_0}{\sigma_0/\sqrt{n}}\sim N(0,1)$$

称为**u检验**

9.2.1.2 σ^2未知，μ的检验

此时构造：

$$t=\frac{\overline{X}-\mu_0}{S/\sqrt{n}}\sim t(n-1)$$

称为 t**检验**

9.2.1.3 μ未知，σ^2的检验

此时构造：

$$\chi^2=\frac{(n-1)S^2}{\sigma_0^2}\sim \chi^2(n-1)$$

称为 χ^2**检验**

9.2.2 两个正态总体下的假设检验

依然根据H_1中不等号的方向选择检验的方向

9.2.2.1 σ_1^2,σ_2^2已知，$\mu_1=\mu_2$的检验

构造：

$$U=\frac{\overline{X}-\overline{Y}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1)$$

9.2.2.2 $\sigma_1^2=\sigma^2=\sigma^2$，$\sigma^2$未知，$\mu_1=\mu_2$的检验

构造：

$$T=\frac{\overline{X}-\overline{Y}}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$$

其中：

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

9.2.2.3 μ_1,μ_2未知，$\sigma_1^2=\sigma_2^2$的检验

构造：

$$F=\frac{S_1^2}{S_2^2}\sim F(n_1-1,n_2-1)$$

> 注意此时$H_1:\sigma_1^2>\sigma_2^2$时，拒绝域：$F>F_{1-\alpha}(n_1-1,n_2-1)$；$H_1:\sigma_1^2<\sigma_2^2$时，拒绝域：$F<F_{\alpha}(n_1-1,n_2-1)$。二者下标不同

9.3 自然指数分布族的检验

由中心极限定理，构造：

$$U=\frac{\overline{X}-m_0}{\sqrt{V(m_0)}/n}\sim N(0,1)$$

9.4 总体分布的χ^2拟合优度检验

当总体分布情况未知时，我们希望检验这个总体是否符合某个分布，这时可以考虑 χ^2**拟合优度检验**

χ^2拟合优度检验的思想：

1. 将数轴作划分，一般取$5 \leq k \leq 16$：
   $$
   - \infty = a_0 < a_1 < a_2 < \cdots < a_k = \infty
   $$
2. H_0成立时，X落入I_i的概率为($I_i = (a_{i-1}, a_i]$)：
   $$
   p_i = P(a_{i=1} < X \leq a_i) = F_0(a_i) - F_0(a_{i-1})
   $$
   并计算样本落入I_i的概率$\frac{n_i}{n}$
3. 由伯努利大数律知，当H_0成立且n较大时，n_i/n与p_i应当充分接近，即$|\frac{n_i}{n} - p_i|$应当充分小，也即$(\frac{n_i}{n} - p_i)^2$充分小。于是拟合优度与$(\frac{n_i}{n} - p_i)^2$的加权平方和有关，考虑：
   $$
   \chi^2 = \sum_{i=1}^{k} \frac{n}{p_i}(\frac{n_i}{n} - p_i)^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} \sim \chi^2(k-r-1)
   $$
   作为统计量，其中k为区间数，r为未知参数个数

注意计算p_i时，若$F_0(x)$中含有未知参数θ则要事先用极大似然法估计

这时候，由统计量，我们得到的拒绝域：

$$
W = \{\chi^2 > \chi^2_{1-\alpha}(k-r-1)\}
$$

其中α为显著性水平

第十章 线性回归

10.1 线性回归模型

设x是普通变量，随机变量Y，且：

$$
Y = \alpha + \beta x + \epsilon, \quad \epsilon \sim N(0, \sigma^2)
$$

其中α, β, σ^2是不依赖于x的未知参数，称此模型为**一元线性回归模型**。此时：

$$
Y \sim N(\alpha + \beta x, \sigma^2)
$$

称Y的期望：

$$
\tilde{Y} = E(Y) = \alpha + \beta
$$

是Y关于x的**线性回归函数**，称α和β为**回归系数**，x为**回归变量**。在上面的模型中，Y与x存在线性相关关系，称为**线性回归关系**

对于样本观测值对(x_i, y_i)，若能得到α和β的点估计$\tilde{\alpha}$和$\tilde{\beta}$，则称：

$$
\tilde{Y} = \tilde{\alpha} + \tilde{\beta} x
$$

为**线性回归方程**，它所代表的直线叫做**线性回归直线**

利用极大似然法，估计出α, β, σ^2：

$$\begin{cases} \hat{\alpha} = \overline{Y} - \hat{\beta}\overline{x} \\ \hat{\beta} = \frac{s_x Y}{s_{xx}} \\ \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}[Y_i - (\hat{\alpha} + \hat{\beta}x)^2] \end{cases}$$

> 这里的估计方差σ^2是残差平方和的均值，也就是均方误差

下面是关于$\hat{\alpha}, \hat{\beta}, \hat{\sigma}^2$的性质：

1. $\overline{Y}, \hat{\beta}, \hat{\sigma}^2$相互独立
2. $(\hat{\alpha}, \hat{\beta})$满足一个二维正态分布
3. $\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-2)$

10.2 线性回归模型的显著性检验

对于线性回归模型，我们的显著性检验相当于：

$$H_0 : \beta = 0 H_1 : \beta \neq 0$$

根据上面提到的性质，构建一个检验统计量：

$$t = \frac{\hat{\beta}\sqrt{s_{xx}}}{\sqrt{\frac{n}{n-2}\hat{\sigma}^2}} \sim t(n-2)$$

当H_1成立时，$|t|$偏大，得到检验拒绝域：

$$W = \{|t| > t_{1-\alpha/2}(n-2)\}$$

通过数学推到我们得到一次估计值与实际值之间：

$$\hat{Y}_0 - Y_0 \sim N\left(0, \sigma^2\left[1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{s_{xx}}\right]\right)$$

因此考虑上面的性质3，构造：

$$t = \frac{\hat{Y}_0 - Y_0}{\sqrt{\frac{n\hat{\sigma}^2}{n-2}\left[1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{s_{xx}}\right]}} \sim t(n-2)$$

由上式可以得到Y_0的**预测区间**：

$$(\hat{Y}_0 - \delta(x_0), \hat{Y}_0 + \delta(x_0))$$

其中：

$$\delta(x_0) = t_{1-\frac{\alpha}{2}}\sqrt{\frac{n}{n-2}\hat{\sigma}^2\left[1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{s_{xx}}\right]}$$

#概率统计
#math