第一章　数理统计的基本知识

第一节　引论

第一节引论

1.数理统计学是数学的一个分支学科

2.数理统计学研究怎样有效地收集、整理和分析带有随机性的数据，以对所考察的问题作出推断和预测，直至为采取一定的决策和行动提供依据和建议。

3.数理统计的基本内容

$$\text{(1)数据收集}\begin{cases}\text{全面观测}\\\text{抽样技术}\\\text{试验设计}\end{cases}$$

$$\text{(2)统计推断}\begin{cases}\text{参数估计}\\\text{假设检验}\end{cases}$$

第一节引论

1.数理统计学是数学的一个分支学科

2.数理统计学研究怎样有效地收集、整理和分析带有随机性的数据，以对所考察的问题作出推断和预测，直至为采取一定的决策和行动提供依据和建议。

3.数理统计的基本内容

$$(1)\text{数据收集}\begin{cases}\text{全面观测}\\\text{抽样技术}\\\text{试验设计}\end{cases}$$

$$(2)\text{统计推断}\begin{cases}\text{参数估计}\\\text{假设检验}\end{cases}$$

第一节引论

1.数理统计学是数学的一个分支学科

2.数理统计学研究怎样有效地收集、整理和分析带有随机性的数据，以对所考察的问题作出推断和预测，直至为采取一定的决策和行动提供依据和建议。

3.数理统计的基本内容

$$(1)\text{数据收集}\begin{cases}\text{全面观测}\\ \text{抽样技术}\\ \text{试验设计}\end{cases}$$

$$(2)\text{统计推断}\begin{cases}\text{参数估计}\\ \text{假设检验}\end{cases}$$

第一节引论

例：某钢厂日产某型号钢筋10000根，质量检查员每天抽查50根的强度，于是有

（1）如何从仅有的50根钢筋强度数据估计整批钢筋强度的平均值？又如何估计整批钢筋强度偏离平均值的离散程度？——参数估计

（2）若规定了这种型号的钢筋强度，从抽查的50个强度数据如何判断整批钢筋的平均强度与规定标准有无差异？——假设检验

（3）若采用不同工艺，抽样的50个强度数据有大有小，那么强度呈现的差异是由工艺不同造成的，还是仅仅由随机因素造成的？——方差分析

（4）若钢筋强度与某种原料成分的含量有关，那么从抽查50根钢筋得到的强度与该成分含量的对应数据，如何表达整批钢筋强度与该成分含量之间的关系？——回归分析

第一节引论

4.应用：工业，农业，医药卫生，生物，环境，管理，金融，保险等等

- 《草木缘情：中国古典文学中的植物世界》–潘富俊
  植物会解决文学史上的公案？它们说《红楼梦》不是曹雪芹一个人写的。《红楼梦》前80回平均每回出现植物11种，后40回每回3.8种。
- 《魔鬼经济学(Freakonomics)》–史蒂芬・列维特，史蒂芬・都伯纳

第一节引论

5.统计计算是一门包含数理统计、计算数学及计算机科学的交叉学科。

6.软件：**Matlab, Splus, SAS, SPSS, R**等等

第二节　数理统计的基本概念

第二节数理统计的基本概念

一.总体和样本

1.总体： 所研究的全部元素组成的集合（母体）

2.个体： 总体中的每个元素称为个体

例：研究某批灯泡的寿命，则该批灯泡寿命数据的集合就构成一个总体，其中每个灯泡的寿命就是一个个体。

注:
*(1)*常用大写字母X, Y, Z表示总体。

(2) X是一随机变量，若X的分布函数为$F(x)$,则总体X为具有分布函数$F(x)$的总体。

第二节数理统计的基本概念

3.样本：从总体抽取的部分个体，称为样本，用$(X_1, X_2, ..., X_n)$表示，n为样本中所含的个体数，称为样本大小或样本容量。

注:(1)$(X_1, \cdots, X_n)$可以看成一个n维随机向量。

(2)每次抽样观测得到$(X_1, \cdots, X_n)$的一组确定值$(x_1, \cdots, x_n)$称作样本观测值。

(3)样本$(X_1, \cdots, X_n)$可能取值的全体称为样本空间(Sample Space),记为$\mathcal{X}$.即$(x_1, \cdots, x_n) \in \mathcal{X}$

4.简单随机样本：满足以下两条性质

(1)代表性：$X_1, \cdots, X_n$与总体X有相同的分布.

(2)独立性：$X_1, \cdots, X_n$是相互独立的随机变量.

第二节数理统计的基本概念

设总体$X \sim F(x)$，则样本$(X_1, \cdots, X_n)$的联合分布函数为$\prod_{i=1}^{n} F(x_i)$,

- 若连续型总体X的密度函数为$f(x)$, 则样本$(X_1, \cdots, X_n)$的联合密度函数为$f_n(x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i)$.
- 若总体X为离散型, 则样本$(X_1, \cdots, X_n)$的联合分布为$P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i)$.

例1：设某批产品共有N个，其中次品数为M，其次品率为$p=M/N$, 从这批产品中任取一件，用X 来描述其质量，$X=\begin{cases}1, & 所取产品为次品\\ 0, & 所取产品为正品\end{cases}$ 则X的分布为
$P\{X=x\}=p^x(1-p)^{1-x}, x=0,1.$
随机抽取样本$(X_1,\cdots,X_n)$，其联合分布为
$P\{X_1=x_1,...,X_n=x_n\}=p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}, x_i=0,1.$

例2：某城市居民收入服从正态分布$N(\mu,\sigma^2)$,概率密度函数为$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}, -\infty<x<\infty$
随机抽取样本$(X_1,\cdots,X_n)$，其联合密度为$L(x_1,\cdots,x_n)=(2\pi\sigma^2)^{-n/2}\exp\{-\frac{\sum_{i=1}^n(x_i-\mu)^2}{2\sigma^2}\}$

例1：设某批产品共有N个，其中次品数为M，其次品率为$p = M/N$, 从这批产品中任取一件，用X 来描述其质量，$X = \begin{cases} 1, & 所取产品为次品 \\ 0, & 所取产品为正品 \end{cases}$ 则X的分布为

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

随机抽取样本$(X_1, \cdots, X_n)$，其联合分布为

$P\{X_1 = x_1, ..., X_n = x_n\} = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}, x_i = 0, 1.$

例2：某城市居民收入服从正态分布$N(\mu, \sigma^2)$,概率密度函数为$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}, -\infty < x < \infty$

随机抽取样本$(X_1, \cdots, X_n)$，其联合密度为$L(x_1, \cdots, x_n) = (2\pi\sigma^2)^{-n/2} \exp\{-\frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^2}\}$

例1：设某批产品共有N个，其中次品数为M，其次品率为$p=M/N$, 从这批产品中任取一件，用X来描述其质量，$X=\begin{cases}1, & \text{所取产品为次品}\\ 0, & \text{所取产品为正品}\end{cases}$ 则X的分布为

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

随机抽取样本$(X_1,\cdots,X_n)$，其联合分布为

$P\{X_1=x_1,...,X_n=x_n\}=p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}, x_i=0,1.$

例2：某城市居民收入服从正态分布$N(\mu,\sigma^2)$,概率密度函数为$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}, -\infty<x<\infty$

随机抽取样本$(X_1,\cdots,X_n)$，其联合密度为$L(x_1,\cdots,x_n)=(2\pi\sigma^2)^{-n/2}\exp\{-\frac{\sum_{i=1}^n(x_i-\mu)^2}{2\sigma^2}\}$

例1：设某批产品共有N个，其中次品数为M，其次品率为$p = M/N$, 从这批产品中任取一件，用X 来描述其质量，$X = \begin{cases} 1, & 所取产品为次品 \\ 0, & 所取产品为正品 \end{cases}$ 则X的分布为
$P\{X = x\} = p^x(1-p)^{1-x}, x = 0, 1.$
随机抽取样本$(X_1, \cdots, X_n)$，其联合分布为
$P\{X_1 = x_1, ..., X_n = x_n\} = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}, x_i = 0, 1.$

例2：某城市居民收入服从正态分布$N(\mu, \sigma^2)$,概率密度函数为$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}, -\infty < x < \infty$
随机抽取样本$(X_1, \cdots, X_n)$，其联合密度为$L(x_1, \cdots, x_n) = (2\pi\sigma^2)^{-n/2} \exp\{-\frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^2}\}$

例1：设某批产品共有N个，其中次品数为M，其次品率为$p = M/N$, 从这批产品中任取一件，用X 来描述其质量，$X = \begin{cases} 1, & \text{所取产品为次品} \\ 0, & \text{所取产品为正品} \end{cases}$ 则X的分布为
$P\{X = x\} = p^x(1-p)^{1-x}, x = 0, 1.$
随机抽取样本$(X_1, \cdots, X_n)$，其联合分布为
$P\{X_1 = x_1, ..., X_n = x_n\} = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}, x_i = 0, 1.$

例2：某城市居民收入服从正态分布$N(\mu, \sigma^2)$,概率密度函数为$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}, -\infty < x < \infty$
随机抽取样本$(X_1, \cdots, X_n)$，其联合密度为$L(x_1, \cdots, x_n) = (2\pi\sigma^2)^{-n/2} \exp\{-\frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^2}\}$

例1：设某批产品共有N个，其中次品数为M，其次品率为$p = M/N$，从这批产品中任取一件，用X 来描述其质量，$X = \begin{cases} 1, & \text{所取产品为次品} \\ 0, & \text{所取产品为正品} \end{cases}$ 则X的分布为
$P\{X = x\} = p^x(1-p)^{1-x}, x = 0, 1.$
随机抽取样本$(X_1, \cdots, X_n)$，其联合分布为
$P\{X_1 = x_1, ..., X_n = x_n\} = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}, x_i = 0, 1.$

例2：某城市居民收入服从正态分布$N(\mu, \sigma^2)$,概率密度函数为$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}, -\infty < x < \infty$
随机抽取样本$(X_1, \cdots, X_n)$，其联合密度为$L(x_1, \cdots, x_n) = (2\pi\sigma^2)^{-n/2} \exp\{-\frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^2}\}$

第二节数理统计的基本概念

二.直方图–分布密度函数曲线的近似

1.找出样本观测值的最小值和最大值,并把包含它们的区间$[a,\ b]$分成m等分.

注：一般分为7–18组，以$m=1+3.32\log n$作为组数,h为组距

$$a=c_0<c_1<\cdots<c_m=b, h=c_i-c_{i-1}=\frac{b-a}{m}, \qquad i=1,\cdots,m,$$

2.数出样本观测值落在各区间$(c_{i-1},\ c_i]$中的个数n_i,称为第i组的组频数,$f_i=\frac{n_i}{n}$称为第i组的组频率。

3.同一组的数据都看成是相同的,它们都等于组中值$\frac{c_{i-1}+c_i}{2}$,分组整理。

4.在x轴上标出点$c_i,\ i=0,1,\cdots,m$,以各区间$(c_{i-1},\ c_i]$为底,组频率与组距之比$y_i=\frac{f_i}{h}=\frac{n_i}{nh}$为高作矩形,这种图称为频率直方图。

例3：表中的125 个数据表示某锅炉所炼生铁中锰的含量,每天测得5 个数据，作频率直方图。

1.40	1.28	1.36	1.38	1.44	1.40	1.34	1.54	1.44	1.46
1.80*	1.44	1.46	1.50	1.38	1.54	1.50	1.48	1.52	1.58
1.52	1.46	1.42	1.58	1.70	1.62	1.58	1.62	1.76	1.68
1.68	1.66	1.62	1.72	1.60	1.62	1.46	1.38	1.42	1.38
1.60	1.44	1.46	1.38	1.34	1.38	1.34	1.36	1.58	1.38
1.34	1.28	1.08	1.08	1.36	1.50	1.46	1.28	1.18	1.28
1.26	1.50	1.52	1.38	1.50	1.52	1.50	1.46	1.34	1.40
1.50	1.42	1.38	1.36	1.38	1.42	1.34	1.48	1.36	1.36
1.32	1.40	1.40	1.26	1.26	1.16	1.34	1.40	1.16	1.54
1.24	1.22	1.20	1.30	1.36	1.30	1.48	1.28	1.18	1.28
1.30	1.52	1.76	1.16	1.28	1.48	1.46	1.48	1.42	1.36
1.32	1.22	1.72	1.18	1.36	1.44	1.28	1.10	1.06*	1.10
1.16	1.22	1.24	1.22	1.34					

第二节数理统计的基本概念

表1.2 生铁含锰量频数表

各组分点	组中值	频数	各组分点	组中值	频数
0.99–1.09	1.04	3	1.39–1.49	1.44	29
1.09–1.19	1.14	9	1.49–1.59	1.54	19
1.19–1.29	1.24	18	1.59–1.69	1.64	9
1.29–1.39	1.34	32	1.69–1.79	1.74	5
			1.79–1.89	1.84	1

注:(1)频率直方图面积为1,

(2)在R中命令为hist;

第二节数理统计的基本概念

众数：数据最集中的取值，也就是最大频率所对应的组中值；

算术平均：$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$；

中位数：将数据按大小顺序排列，居于中间的那个数值；

极差：最大观测值与最小观测值之差。

三 统计量

定义1：设$(X_1, \cdots, X_n)$是来自总体X的一个样本，
$T = T(x_1, \cdots, x_n)$ 是样本空间$\mathcal{X}$上的实值函数，
若$T(X_1, \cdots, X_n)$也是随机变量,且不依赖于任何未知参数,则
称$T(X_1, \cdots, X_n)$为**统计量**(Statistics).

例4.设X_1, X_2是从总体$N(\mu, \sigma^2)$中抽取的一个样本，其中参数μ已知，σ^2未知，判定以下各式哪个是统计量？

$$X_1 \cdot X_2 - 3\sigma^2, \quad X_1^2 + X_2^2 + 5\sigma, \quad X_1, \quad X_1 + 5X_2^2, \quad X_1/X_2 + 3\mu X_1^2$$

注：
(1)借助统计量，可以把子样所含信息进行数学加工，使其浓缩，从而使问题解决。
(2)统计量是一个随机变量。

三 统计量

定义2: 设$(X_1,\cdots,X_n)$是取自总体X的大小为n的样本,则

(1)$\overline{X}=\frac{1}{n}\Sigma_{i=1}^{n}X_i$称为样本均值;

(2)$S^2=\frac{1}{n-1}\Sigma_{i=1}^{n}(X_i-\overline{X})^2$称为样本方差,

而$S=\sqrt{\frac{1}{n-1}\Sigma_{i=1}^{n}(X_i-\overline{X})^2}$为样本均方差或样本标准差;

(3)$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$称为样本的$k$阶原点矩;

$B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k$为样本的$k$阶中心距.

注:二阶中心矩B_2有时记为$\tilde{S}^2$,即$\tilde{S}^2=\frac{1}{n}\Sigma_{i=1}^{n}(X_i-\overline{X})^2$.

特别地,$A_1=\overline{X},\ B_2=\tilde{S}^2=\frac{n-1}{n}S^2$.

(4)$b_s=\frac{B_3}{B_2^{3/2}}$,称为样本的偏度;

(5)$b_k=\frac{B_4}{B_2^2}-3$称为样本峰度;

(6)$V=\frac{S}{\overline{X}}$称为样本的变异系数。

三 统计量

记$E(X) \triangleq \mu, \quad D(X) \triangleq \sigma^2, \quad E(X^h) \triangleq \alpha_h, \quad E(X-\mu)^h \triangleq \mu_h$

定理1.1:设总体X的分布函数$F(x)$具有二阶矩,$(X_1, \cdots, X_n)$是取自这个总体的一个样本, 则对样本均值$\overline{X}$,有

$$E(\overline{X}) = \mu, \qquad D(\overline{X}) = \frac{\sigma^2}{n}.$$

定理1.2:设总体X的分布函数$F(x)$具有二阶矩,$(X_1, \cdots, X_n)$是取自这个总体的一个样本,则对样本方差S^2,有

$$E(S^2) \quad = \quad \sigma^2,$$

三 统计量

定理1.3: 设总体X的分布函数$F(x)$具有$2k$阶矩,$(X_1, \cdots, X_n)$是取自这个总体的一个样本,则对k 阶样本矩A_k,有

$$E(A_k) = \alpha_k, \quad D(A_k) = \frac{\alpha_{2k} - \alpha_k^2}{n}$$

第三节　次序统计量及其分布

第三节 次序统计量及其分布

一.次序统计量及经验分布函数

1.**定义**: 设$(X_1,\cdots,X_n)$是取自总体X的一个样本,将它们按大小排列为$X_{(1)}\le X_{(2)}\le\cdots\le X_{(n)}$, 则称$(X_{(1)},X_{(2)},\cdots,X_{(n)})$为**次序统计量**,称$X_{(i)}$为第$i$ 个次序统计量.

特别:$X_{(1)}=\min\{X_1,\cdots,X_n\},X_{(n)}=\max\{X_1,\cdots,X_n\}$分别称为最小次序统计量和最大次序统计量。他们的取值分别称为极小值和极大值.

2.**定义**: 设$(X_1,\cdots,X_n)$是取自分布函数为$F(x)$的总体X的一个样本,以$\nu_n(x)$ 表示$(X_1,\cdots,X_n)$ 中不超过x的观测值的个数,则称$F_n(x)=\frac{\nu_n(x)}{n}$为**经验分布函数**,简记为EDF.

第三节 次序统计量及其分布

注:$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$,则经验分布函数为

$$F_n(x) = \begin{cases} 0, & x < x_{(1)}; \\ \frac{k}{n}, & x_{(k)} \leq x < x_{(k+1)}, \ k = 1, \cdots, n-1; \\ 1, & x_{(n)} \leq x, \end{cases}$$

第三节 次序统计量及其分布

3.性质:(1)$0 \le F_n(x) \le 1$;

(2)单调不减的右连续函数;

(3)在$x = x_{(k)}$有间断点,在每个间断点上有跃度$\frac{1}{n}$.

例：从总体X中抽取容量为7的样本，其观测值为1, 3, 2.5, 2, 2.5, 3, 2.5，求X的经验分布函数。

解：将样本由小到大排序,

有$1 < 2 < 2.5 = 2.5 = 2.5 < 3 = 3$,由定义得经验分布函数为

$$F_7(x) = \begin{cases} 0, & x < 1; \\ \frac{1}{7}, & 1 \le x < 2 \\ \frac{2}{7}, & 2 \le x < 2.5 \\ \frac{5}{7}, & 2.5 \le x < 3 \\ 1, & x \ge 3, \end{cases}$$

第三节 次序统计量及其分布

3.性质:(1)$0 \le F_n(x) \le 1$;

(2)单调不减的右连续函数;

(3)在$x = x_{(k)}$有间断点,在每个间断点上有跃度$\frac{1}{n}$.

例：从总体X中抽取容量为7的样本，其观测值为1, 3, 2.5, 2, 2.5, 3, 2.5，求X的经验分布函数。

解：将样本由小到大排序，

有$1 < 2 < 2.5 = 2.5 = 2.5 < 3 = 3$,由定义得经验分布函数为

$$F_7(x) = \begin{cases} 0, & x < 1; \\ \frac{1}{7}, & 1 \le x < 2 \\ \frac{2}{7}, & 2 \le x < 2.5 \\ \frac{5}{7}, & 2.5 \le x < 3 \\ 1, & x \ge 3, \end{cases}$$

第三节 次序统计量及其分布

3.性质:(1)$0 \leq F_n(x) \leq 1$;

(2)单调不减的右连续函数;

(3)在$x = x_{(k)}$有间断点,在每个间断点上有跃度$\frac{1}{n}$.

例：从总体X中抽取容量为7的样本，其观测值为1, 3, 2.5, 2, 2.5, 3, 2.5，求X的经验分布函数。

解：将样本由小到大排序，

有$1 < 2 < 2.5 = 2.5 = 2.5 < 3 = 3$,由定义得经验分布函数为

$$F_7(x) = \begin{cases} 0, & x < 1; \\ \frac{1}{7}, & 1 \leq x < 2 \\ \frac{2}{7}, & 2 \leq x < 2.5 \\ \frac{5}{7}, & 2.5 \leq x < 3 \\ 1, & x \geq 3, \end{cases}$$

第三节 次序统计量及其分布

4. 记$F(x)=P\{X\leq x\}$, $F_n(x)=\frac{\nu_n(x)}{n}$表示$\{X\leq x\}$的频率，固定$x$, $\nu_n(x), F_n(x)$是$(X_1,\ldots,X_n)$的函数，是随机变量.

$$P\{\nu_n(x)=k\} = P\{n\text{次独立试验中恰有 }k\text{次}X\leq x\}$$

则$\nu_n(x)\sim b(n,F(x))$.

$$= \binom{n}{k}[F(x)]^k[1-F(x)]^{n-k}.$$

由二项分布知$E[\nu_n(x)]=nF(x)$,则$E[F_n(x)]=F(x)$.

第三节 次序统计量及其分布

4. 记$F(x)=P\{X\leq x\}$, $F_n(x)=\frac{\nu_n(x)}{n}$表示$\{X\leq x\}$的频率, 固定$x$, $\nu_n(x), F_n(x)$是$(X_1,...,X_n)$的函数, 是随机变量.

$$P\{\nu_n(x)=k\} = P\{n\text{次独立试验中恰有 }k\text{次}X\leq x\}$$

则$\nu_n(x)\sim b(n,F(x))$.

$$= \binom{n}{k}[F(x)]^k[1-F(x)]^{n-k}.$$

由二项分布知$E[\nu_n(x)]=nF(x)$,则$E[F_n(x)]=F(x)$.

第三节 次序统计量及其分布

4. 记$F(x) = P\{X \leq x\}$, $F_n(x) = \frac{\nu_n(x)}{n}$表示$\{X \leq x\}$的频率，固定$x$, $\nu_n(x), F_n(x)$是$(X_1, ..., X_n)$的函数，是随机变量.

$$P\{\nu_n(x) = k\} = P\{n\text{次独立试验中恰有 }k\text{次}X \leq x\}$$

则$\nu_n(x) \sim b(n, F(x))$.

$$= \binom{n}{k}[F(x)]^k[1 - F(x)]^{n-k}.$$

由二项分布知$E[\nu_n(x)] = nF(x)$,则$E[F_n(x)] = F(x)$.

第三节 次序统计量及其分布

4. 记$F(x)=P\{X\le x\}$, $F_n(x)=\frac{\nu_n(x)}{n}$表示$\{X\le x\}$的频率，固定$x$, $\nu_n(x), F_n(x)$是$(X_1,\ldots,X_n)$的函数，是随机变量.

$$P\{\nu_n(x)=k\} = P\{n\text{次独立试验中恰有 }k\text{次}X\le x\}$$

则$\nu_n(x)\sim b(n,F(x))$.

$$= \binom{n}{k}[F(x)]^k[1-F(x)]^{n-k}.$$

由二项分布知$E[\nu_n(x)]=nF(x)$,则$E[F_n(x)]=F(x)$.

第三节 次序统计量及其分布

4. 记$F(x)=P\{X\leq x\}$, $F_n(x)=\frac{\nu_n(x)}{n}$表示$\{X\leq x\}$的频率，固定$x$, $\nu_n(x), F_n(x)$是$(X_1,\dots,X_n)$的函数，是随机变量.

$$P\{\nu_n(x)=k\} = P\{n\text{次独立试验中恰有 }k\text{次}X\leq x\}$$

则$\nu_n(x)\sim b(n,F(x))$.

$$= \binom{n}{k}[F(x)]^k[1-F(x)]^{n-k}.$$

由二项分布知$E[\nu_n(x)]=nF(x)$,则$E[F_n(x)]=F(x)$.

第三节 次序统计量及其分布

5.**Glivenko定理**: 设总体X的分布函数为$F(x)$,经验分布函数为$F_n(x)$,记

$$D_n = \sup_{-\infty<x<+\infty} \mid F_n(x) - F(x) \mid,$$

则有$P\{\lim_{n\to\infty} D_n = 0\} = 1.$

注:(1)当n很大时，由每一组样本观测值得到的经验分布函数$F_n(x)$都是总体分布$F(x)$的一个良好近似。(样本推断总体)

(2)两种近似方法:频率直方图(密度)；经验分布函数(分布)

第三节　次序统计量及其分布

二.次序统计量的分布

(一)单个次序统计量由于$\nu_n(x) \sim b(n, F(x))$，易知

$$\begin{aligned}
\{\nu_n(x) = 0\} &= \{X_{(1)} > x\}, \quad \{\nu_n(x) = n\} = \{X_{(n)} \le x\}, & (1)\\
\{\nu_n(x) = k\} &= \{X_{(k)} \le x < X_{(k+1)}\},\ 1 \le k \le n-1, & (2)\\
\{\nu_n(x) \ge k\} &= \{X_{(k)} \le x\},\ 1 \le k \le n-1. & (3)
\end{aligned}$$

若总体X有分布函数$F(x)$, 密度函数$f(x)$, 则$X_{(i)}$的分布函数为

$$\begin{aligned}
F_i(x) &= P\{X_{(i)} \le x\} = P\{\nu_n(x) \ge i\}\\
&= \sum_{k=i}^{n} \binom{n}{k} [F(x)]^k [1-F(x)]^{n-k} \qquad (4)
\end{aligned}$$

第三节 次序统计量及其分布

二.次序统计量的分布

(一)单个次序统计量由于$\nu_n(x) \sim b(n, F(x))$，易知

$$\begin{aligned}
\{\nu_n(x)=0\} &= \{X_{(1)}>x\}, \quad \{\nu_n(x)=n\}=\{X_{(n)}\le x\}, && (1)\\
\{\nu_n(x)=k\} &= \{X_{(k)}\le x<X_{(k+1)}\},\ 1\le k\le n-1, && (2)\\
\{\nu_n(x)\ge k\} &= \{X_{(k)}\le x\},\ 1\le k\le n-1. && (3)
\end{aligned}$$

若总体X有分布函数$F(x)$, 密度函数$f(x)$, 则$X_{(i)}$的分布函数为

$$\begin{aligned}
F_i(x) &= P\{X_{(i)}\le x\}=P\{\nu_n(x)\ge i\}\\
&= \sum_{k=i}^{n}\binom{n}{k}[F(x)]^k[1-F(x)]^{n-k} && (4)
\end{aligned}$$

第三节 次序统计量及其分布

注: (1)最大次序统计量$X_{(n)}$的分布函数为$F_n(x)=[F(x)]^n$,
分布密度函数为$f_n(x)=n[F(x)]^{n-1}f(x)$.

(2)最小次序统计量$X_{(1)}$的分布函数为$F_1(x)=1-[1-F(x)]^n$,
分布密度函数为$f_1(x)=n[1-F(x)]^{n-1}f(x)$.

例1: 设总体X的密度函数为$f(x)=\begin{cases}2x, & 0\le x<1\\ 0, & 其他\end{cases}$，$(X_1,\cdots,X_4)$为取自总体的一个样本，求$X_{(3)}$的分布函数，并计算$P(X_{(3)}>\frac{1}{2})$.

第三节　次序统计量及其分布

注: (1)最大次序统计量$X_{(n)}$的分布函数为$F_n(x)=[F(x)]^n$,
分布密度函数为$f_n(x)=n[F(x)]^{n-1}f(x)$.

(2)最小次序统计量$X_{(1)}$的分布函数为$F_1(x)=1-[1-F(x)]^n$,
分布密度函数为$f_1(x)=n[1-F(x)]^{n-1}f(x)$.

例1: 设总体X的密度函数
为$f(x)=\begin{cases} 2x, & 0\le x<1 \\ 0, & 其他 \end{cases}$，$(X_1,\cdots,X_4)$为取自总体的一个样本，求$X_{(3)}$的分布函数，并计算$P(X_{(3)}>\frac{1}{2})$.

第三节 次序统计量及其分布

注: (1)最大次序统计量$X_{(n)}$的分布函数为$F_n(x) = [F(x)]^n$，分布密度函数为$f_n(x) = n[F(x)]^{n-1}f(x)$.

(2)最小次序统计量$X_{(1)}$的分布函数为$F_1(x) = 1 - [1 - F(x)]^n$，分布密度函数为$f_1(x) = n[1 - F(x)]^{n-1}f(x)$.

例1: 设总体X的密度函数为$f(x) = \begin{cases} 2x, & 0 \le x < 1 \\ 0, & 其他 \end{cases}$，$(X_1, \cdots, X_4)$为取自总体的一个样本，求$X_{(3)}$的分布函数，并计算$P(X_{(3)} > \frac{1}{2})$.

第三节 次序统计量及其分布

注: (1)最大次序统计量$X_{(n)}$的分布函数为$F_n(x) = [F(x)]^n$，分布密度函数为$f_n(x) = n[F(x)]^{n-1}f(x)$.

(2)最小次序统计量$X_{(1)}$的分布函数为$F_1(x) = 1 - [1 - F(x)]^n$，分布密度函数为$f_1(x) = n[1 - F(x)]^{n-1}f(x)$.

例1: 设总体X的密度函数为$f(x) = \begin{cases} 2x, & 0 \le x < 1 \\ 0, & 其他 \end{cases}$，$(X_1, \cdots, X_4)$为取自总体的一个样本，求$X_{(3)}$的分布函数，并计算$P(X_{(3)} > \frac{1}{2})$.

第三节 次序统计量及其分布

注: (1)最大次序统计量$X_{(n)}$的分布函数为$F_n(x)=[F(x)]^n$, 分布密度函数为$f_n(x)=n[F(x)]^{n-1}f(x)$.

(2)最小次序统计量$X_{(1)}$的分布函数为$F_1(x)=1-[1-F(x)]^n$, 分布密度函数为$f_1(x)=n[1-F(x)]^{n-1}f(x)$.

例1: 设总体X的密度函数为$f(x)=\begin{cases} 2x, & 0\le x<1 \\ 0, & \text{其他} \end{cases}$，$(X_1,\cdots,X_4)$为取自总体的一个样本，求$X_{(3)}$的分布函数，并计算$P(X_{(3)}>\frac{1}{2})$.

第三节　次序统计量及其分布

解:因为$F(x) = \begin{cases} 0, & x < 0 \\ x^2, & 0 \le x < 1 \\ 1, & x \ge 1 \end{cases}$,

而

$$\begin{aligned} F_3(x) &= \sum_{k=3}^{4} \binom{4}{k} [F(x)]^k [1 - F(x)]^{4-k} \\ &= 4[F(x)]^3[1 - F(x)] + [F(x)]^4 \\ &= 4[F(x)]^3 - 3[F(x)]^4 \\ &= \begin{cases} 0, & x < 0 \\ 4x^6 - 3x^8, & 0 \le x < 1 \\ 1, & x \ge 1 \end{cases} \end{aligned}$$

所以

$$\begin{aligned} P(X_{(3)} > \tfrac{1}{2}) &= 1 - P(X_{(3)} \le \tfrac{1}{2}) = 1 - F_3(\tfrac{1}{2}) \\ &= 1 - (4(\tfrac{1}{2})^6 - 3(\tfrac{1}{2})^8) = 1 - \frac{13}{256} = \frac{243}{256} \end{aligned}$$

第三节 次序统计量及其分布

解:因为$F(x)=\begin{cases} 0, & x<0 \\ x^2, & 0\le x<1 \\ 1, & x\ge 1 \end{cases}$,

而

$$\begin{aligned} F_3(x) &= \sum_{k=3}^{4}\binom{4}{k}[F(x)]^k[1-F(x)]^{4-k} \\ &= 4[F(x)]^3[1-F(x)]+[F(x)]^4 \\ &= 4[F(x)]^3-3[F(x)]^4 \\ &= \begin{cases} 0, & x<0 \\ 4x^6-3x^8, & 0\le x<1 \\ 1, & x\ge 1 \end{cases} \end{aligned}$$

所以

$$\begin{aligned} P(X_{(3)}>\tfrac{1}{2}) &= 1-P(X_{(3)}\le\tfrac{1}{2})=1-F_3(\tfrac{1}{2}) \\ &= 1-(4(\tfrac{1}{2})^6-3(\tfrac{1}{2})^8)=1-\frac{13}{256}=\frac{243}{256} \end{aligned}$$

第三节 次序统计量及其分布

解:因为$F(x)=\begin{cases}0, & x<0\\ x^2, & 0\le x<1\\ 1, & x\ge 1\end{cases}$,

而

$$\begin{aligned}F_3(x) &= \sum_{k=3}^{4}\binom{4}{k}[F(x)]^k[1-F(x)]^{4-k}\\ &= 4[F(x)]^3[1-F(x)]+[F(x)]^4\\ &= 4[F(x)]^3-3[F(x)]^4\\ &= \begin{cases}0, & x<0\\ 4x^6-3x^8, & 0\le x<1\\ 1, & x\ge 1\end{cases}\end{aligned}$$

所以

$$\begin{aligned}P(X_{(3)}>\tfrac{1}{2}) &= 1-P(X_{(3)}\le\tfrac{1}{2})=1-F_3(\tfrac{1}{2})\\ &= 1-(4(\tfrac{1}{2})^6-3(\tfrac{1}{2})^8)=1-\frac{13}{256}=\frac{243}{256}\end{aligned}$$

第三节 次序统计量及其分布

三. 分位数

1.定义:若随机变量X的分布函数为$F(x)$,对任$0<p<1$,称

$$F^{\leftarrow}(p)=inf\{x:F(x)\geq p\}$$

为随机变量X或分布$F(x)$的分位数函数.记$x_p=F^{\leftarrow}(p)$，称x_p为随机变量X或分布$F(x)$的p分位数特别当$p=\frac{1}{2}$时,$x_{\frac{1}{2}}$称为中位数.

2.定义:设$(X_1,\cdots,X_n)$为总体X的一个样本,$(X_{(1)},\cdots,X_{(n)})$为次序统计量,对任$0<p<1$,称$x_p^*=X_{([np]+1)}$ 为样本p分位数.其中$[a]$表示不超过a的最大整数, 特别当$p=\frac{1}{2}$时,$x_{\frac{1}{2}}^*$称为样本中位数.

$$x_{\frac{1}{2}}^*=\begin{cases} X_{(\frac{n+1}{2})}, & \text{当}n\text{为奇数;} \\ \frac{1}{2}[X_{(\frac{n}{2})}+X_{(\frac{n}{2}+1)}], & \text{当}n\text{为偶数.} \end{cases}$$

第三节　次序统计量及其分布

四.极值分布

1.**定义**:设$(X_1,\cdots,X_n)$是取自分布为$F(x)$的总体X的一个样本,如果存在常数列a_n及$b_n>0$, 使$(X_{(n)}-a_n)/b_n$有非退化的极限分布$G(x)$,则称$G(x)$为**极大值分布**.

注：类似可以定义极小值分布,极大值分布和极小值分布统称极值分布.

I型极大值分布(也称贡贝尔(Gumbel)分布)

$$G_1(x)=\exp(-e^{-x}),-\infty<x<\infty,$$

II型极大值分布(也称弗莱契(Fréchet)分布)

$$G_2(x)=\begin{cases}\exp(-x^{-k}), & x>0;\\ 0, & x\le 0,\end{cases}$$

其中$k>0$是参数.

第三节 次序统计量及其分布

III型极大值分布(也称威布尔(Weibull)分布)

$$G_3(x) = \begin{cases} 1, & x > 0; \\ \exp(-(-x)^k), & x \leq 0, \end{cases}$$

其中$k > 0$是参数.

第四节　统计中常用的分布族

第四节 统计中常用的分布族

$F(x;\theta)$表示X的分布,参数θ可能取值的集合称为**参数空间**, 记作Θ, 称$\{F(x;\theta):\theta\in\Theta\}$为$X$的分布函数族.

(1)正态分布族：$\{N(\mu,\sigma^2):(\mu,\sigma^2)\in\Theta\}$, 其中$\Theta=\{(\mu,\sigma^2):-\infty<\mu<\infty,\sigma^2>0\}$.

(2)二项分布族：$\{b(n,p):0<p<1\}$

(3)Possion分布族：$\{P(\lambda):\lambda>0\}$

(4)均匀分布族:$\{U(a,b):-\infty<a<b<\infty\}$

(5)指数分布族:$\{Exp(\lambda):\lambda>0\}$

第四节 统计中常用的分布族

一.Gamma分布族

1.定义：若随机变量X具有密度函数

$$f(x;\alpha,\lambda)=\frac{\lambda^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x},\ x>0,$$

则称X所服从的分布为Gamma分布,记作$Ga(\alpha,\lambda)$,其中$\alpha>0$是形状参数,$\lambda>0$ 是尺度参数，$\{Ga(\alpha,\lambda):\alpha>0,\lambda>0\}$称为Gamma分布族.

2.性质：
(1)特征函数:$\varphi(t)=Ee^{iXt}=(1-\frac{it}{\lambda})^{-\alpha}$
(2)$E(X)=\frac{\alpha}{\lambda},\ D(X)=\frac{\alpha}{\lambda^2}$

第四节 统计中常用的分布族

(3)当$\alpha=1$时,$Ga(1,\lambda)$就是参数为λ的指数分布.$Ga(\frac{n}{2},\frac{1}{2})$称为$n$个自由度的$\chi^2$分布,记作$\chi^2(n)$分布,$\chi^2(n)$分布的密度函数为$f(x,n)=\frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}x^{\frac{n}{2}-1}e^{-\frac{x}{2}},x>0$

(4)可加性:设$X_1\sim Ga(\alpha_1,\lambda),X_2\sim Ga(\alpha_2,\lambda)$,且$X_1$与$X_2$相互独立,则$X_1+X_2\sim Ga(\alpha_1+\alpha_2,\lambda)$

注: (1) $\chi^2(n)$分布是Gamma分布的特殊情况，也具有可加性.

(2) 自由度是指独立随机变量的个数.

第四节 统计中常用的分布族

例1：设$X \sim Ga(\alpha, \lambda)$,其中$\alpha > 0, \lambda > 0$为常数，$(X_1, \cdots, X_n)$是取自总体的样本,求样本和$\sum_{i=1}^{n} X_i$的分布密度函数。

解：因为Gamma分布具有可加性，且$X_i, i = 1, 2, ..., n$独立同分布，所以$\sum_{i=1}^{n} X_i \sim Ga(n\alpha, \lambda)$

第四节　统计中常用的分布族

例2：设$X \sim Ga(\alpha,\lambda)$，$Y = kX$,证明：$Y \sim Ga(\alpha,\frac{\lambda}{k})$，$k > 0$，并且求$\bar{X}$的分布。

解：因为$X \sim Ga(\alpha_1,\lambda)$，即密度函数为

$$f(x;\alpha,\lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x},\ \ x > 0,$$

又因为$Y = kX$,所以$X = Y/k$,因而有

$$\begin{aligned} f(y;\alpha,\lambda) &= \frac{\lambda^\alpha}{\Gamma(\alpha)} (\frac{y}{k})^{\alpha-1} e^{-\lambda\frac{y}{k}} \frac{1}{k} \\ &= \frac{(\frac{\lambda}{k})^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\frac{\lambda}{k}}, \\ &\sim Ga(\alpha,\frac{\lambda}{k}),\quad y > 0 \end{aligned}$$

$\bar{X} = \frac{1}{n}\sum_{i=1}^n x_i \sim Ga(n\alpha, n\lambda)$

第四节 统计中常用的分布族

例3：设$X \sim N(0,\sigma^2)$,则$Y = X^2 \sim Ga(\frac{1}{2}, \frac{1}{2\sigma^2})$.
解：当$y > 0$时,Y的分布函数为

$$\begin{aligned} F_Y(y) &= P\{Y \le y\} = P\{X^2 \le y\} = P\{-\sqrt{y} \le X \le \sqrt{y}\} \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}), \end{aligned}$$

其分布密度函数为

$$f_Y(y) = [f_X(\sqrt{y}) + f_X(-\sqrt{y})]\frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi}\sigma} y^{-\frac{1}{2}} e^{-\frac{y}{2\sigma^2}},$$

因为$\Gamma(\frac{1}{2}) = \sqrt{\pi}$,因此$Y \sim Ga(\frac{1}{2}, \frac{1}{2\sigma^2})$.

第四节 统计中常用的分布族

定理： 设$(X_1,...,X_n)$是取自正态总体$N(0,\sigma^2)$的一个样本，记$\chi^2=\sum_{i=1}^n X_i^2$，则$\chi^2\sim Ga(\frac{n}{2},\frac{1}{2\sigma^2})$.

例4： 设$X\sim U(0,1)$，则$Y=-\alpha\ln X\sim Exp(\frac{1}{\alpha}),\alpha>0$.

3.χ^2分布另一种定义方式

(1)若$X_i\sim N(0,1),i=1,2,...,n,i.i.d.$,则$Y=X_1^2+X_2^2+\cdots+X_n^2\sim\chi^2(n)$

(2)若$X_i\sim N(\mu,\sigma^2),i=1,2,...,n,i.i.d.$,则$\sum_{i=1}^n(\frac{X_i-\mu}{\sigma})^2\sim\chi^2(n)$

例5： 设$(X_1,...,X_4)$是来自正态总体$N(0,4)$的样本，$T=a(X_1-2X_2)^2+b(3X_3-4X_4)^2$，求常数$a,b$，使得$T\sim\chi^2(2)$.

第四节 统计中常用的分布族

4.χ^2分布的性质

(1)$\chi^2(n)$分布的特征函数$\phi(t)=(1-2it)^{-\frac{n}{2}}$;

(2)设$X\sim\chi^2(n)$, 则$EX=n$, $VarX=2n$;

(3)设$X_i\sim\chi^2(n_i), i=1,2,...,k$, 且相互独立,
则$\sum_{i=1}^{k}X_i\sim\chi^2(\sum_{i=1}^{k}n_i)$.

Cochran分解定理: 设$X_1,...,X_n$是独立同分布的随机变量，$X_i\sim N(0,1), i=1,2,...,n$, $Q_i(i=1,...,k)$ 是$X_1,...,X_n$的二次型，其秩为n_i. 如果$Q_1+\cdots+Q_k=\sum_{i=1}^{n}X_i^2$且$\sum_{i=1}^{k}n_i=n$, 则$Q_i\sim\chi^2(n_i), i=1,2,...,k$,且$Q_1,\cdots,Q_k$相互独立.

第四节 统计中常用的分布族

二.Beta分布族

1.定义:若随机变量X具有密度函数

$$f(x;a,b)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1},\qquad 0<x<1$$

则称X所服从的分布为Beta分布,记作$Be(a,b)$,其中$a>0,\ b>0$是两个参数.$\{Be(a,b):a>0,b>0\}$称为**Beta分布族**.

2.性质:

(1)$E(X)=\frac{a}{a+b},\qquad D(X)=\frac{ab}{(a+b)^2(a+b+1)}$

(2)当$a=b=1$时,$Be(1,1)$分布就是$(0,\ 1)$上的均匀分布$U(0,1)$

第四节 统计中常用的分布族

3.定理：若$X_1 \sim \chi^2(n_1), X_2 \sim \chi^2(n_2)$,且$X_1, X_2$相互独立,则
(1)$F = \frac{X_1/n_1}{X_2/n_2} \sim F(n_1, n_2)$, 其中$F(n_1, n_2)$表示自由度为$n_1$, n_2的F分布.
(2)$\frac{X_1}{X_1+X_2} \sim Be(\frac{n_1}{2}, \frac{n_2}{2})$

4.结论
(1)若$X \sim F(n_1, n_2)$,则$\frac{CX}{1+CX} \sim Be(\frac{n_1}{2}, \frac{n_2}{2})$ 其中$C = n_1/n_2$,反之亦然.
(2)若$X_1 \sim \chi^2(n_1), X_2 \sim \chi^2(n_2)$,且$X_1, X_2$相互独立,
则$Y_1 = X_1 + X_2$与$Y_2 = X_1/X_2$相互独立.
(3)若$X \sim F(n_1, n_2)$, 则$\frac{1}{X} \sim F(n_2, n_1)$.

第四节 统计中常用的分布族

三 t分布

1.定义：设随机变量$X \sim N(0,1)$, $Y \sim \chi^2(n)$,且X, Y相互独立,则称随机变量$T = \dfrac{X}{\sqrt{Y/n}}$所服从的分布为$n$个自由度的$t$分布,记作$T \sim t(n)$.

第四节 统计中常用的分布族

2.性质：$t(n)$分布的密度函数为

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

(1)当$n=1$时,$t(x;1)=\frac{1}{\pi}\frac{1}{1+x^2}$即为Cauchy分布.
当$n>2$时,则$E(T)=0$, $D(T)=\frac{n}{n-2}$
(2)若$X\sim t(n)$,则$X^2\sim F(1,n)$.
(3)若$X\sim t(n)$,则$Y=\frac{n}{n+X^2}\sim Be(\frac{n}{2},\frac{1}{2})$.
(4)$\lim_{n\to\infty}t(x;n)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$
t分布的极限分布是标准正态分布。

第四节 统计中常用的分布族

四 指数型分布族

1.定义：设$\mathcal{F}=\{f(x;\theta):\theta\in\Theta\}$是分布族,若样本$(X_1,\cdots,X_n)$的密度函数(或分布列)$f(x_1,\cdots,x_n;\theta)$可以表示成

$$f(x_1,\cdots,x_n;\theta)=a(\theta)\exp\{\sum_{j=1}^{k}Q_j(\theta)T_j(x_1,\cdots,x_n)\}h(x_1,\cdots,x_n)$$

并且它的支撑$\{x:f(x_1,\cdots,x_n;\theta)>0\}$不依赖于$\theta$,则称此分布族为指数型分布族,简称**指数族**。

第四节 统计中常用的分布族

例1:证明二项分布族$\{b(m,p):0<p<1\}$是指数型分布族. 样本空间$\mathcal{X}=\{(x_1,...,x_n):x_i=0,1,...,m,i=1,2,...,n\}$.

证明：样本$(X_1,\cdots,X_n)$的联合概率分布为

$$\begin{aligned} f(x_1,\cdots,x_n;p) &= [\prod_{i=1}^{n}\binom{m}{x_i}]p^{\sum\limits_{i=1}^{n}x_i}(1-p)^{nm-\sum\limits_{i=1}^{n}x_i} \quad (5) \\ &= (1-p)^{nm}\exp\{\sum_{i=1}^{n}x_i\ln\frac{p}{1-p}\}\prod_{i=1}^{n}\binom{m}{x_i} \quad (6) \end{aligned}$$

取$a(p)=(1-p)^{nm}$, $Q_1(p)=\ln\frac{p}{1-p}$, $T_1(x_1,\cdots,x_n)=\sum\limits_{i=1}^{n}x_i$, $h(x_1,\cdots,x_n)=\prod\limits_{i=1}^{n}\binom{m}{x_i}$.

例2：均匀分布族$\{U(-\theta,\theta):\theta>0\}$不是指数型分布族,这是因为它的支撑$\{x:f(x;\theta)>0\}=(-\theta,\theta)$依赖于未知参数$\theta$

第四节 统计中常用的分布族

例1:证明二项分布族$\{b(m,p):0<p<1\}$是指数型分布族. 样本空间$\mathcal{X}=\{(x_1,...,x_n):x_i=0,1,...,m,i=1,2,...,n\}$.

证明：样本$(X_1,\cdots,X_n)$的联合概率分布为

$$\begin{aligned} f(x_1,\cdots,x_n;p) &= [\prod_{i=1}^{n}\binom{m}{x_i}]p^{\sum\limits_{i=1}^{n}x_i}(1-p)^{nm-\sum\limits_{i=1}^{n}x_i} \quad (5)\\ &= (1-p)^{nm}\exp\{\sum_{i=1}^{n}x_i\ln\frac{p}{1-p}\}\prod_{i=1}^{n}\binom{m}{x_i}, \quad (6)\end{aligned}$$

取$a(p)=(1-p)^{nm}$, $Q_1(p)=\ln\frac{p}{1-p}$, $T_1(x_1,\cdots,x_n)=\sum\limits_{i=1}^{n}x_i$, $h(x_1,\cdots,x_n)=\prod\limits_{i=1}^{n}\binom{m}{x_i}$.

例2：均匀分布族$\{U(-\theta,\theta):\theta>0\}$不是指数型分布族,这是因为它的支撑$\{x:f(x;\theta)>0\}=(-\theta,\theta)$依赖于未知参数$\theta$

第四节 统计中常用的分布族

例3：设X服从正态分布$N(\mu,\sigma^2)$,样本空间为n维欧氏空间R^n.记$\mathbf{x}=(x_1,\cdots,x_n)$,那么样本$(X_1,\cdots,X_n)$的分布密度为

$$\begin{aligned} f(\mathbf{x};\mu,\sigma^2) &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\{-\sum_{i=1}^n \frac{1}{2\sigma^2}(x_i-\mu)^2\} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\{-\frac{n\mu^2}{2\sigma^2}+\frac{n\mu}{\sigma^2}\bar{x}-\sum_{i=1}^n \frac{x_i^2}{2\sigma^2}\} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\{-\frac{n\mu^2}{2\sigma^2}\}\exp\{\frac{n\mu}{\sigma^2}\bar{x}-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2\}, \end{aligned}$$

第四节 统计中常用的分布族

例3：设X服从正态分布$N(\mu,\sigma^2)$,样本空间为n维欧氏空间R^n.记$\mathbf{x}=(x_1,\cdots,x_n)$,那么样本$(X_1,\cdots,X_n)$的分布密度为

$$\begin{aligned}
f(\mathbf{x};\mu,\sigma^2) &= (\frac{1}{\sqrt{2\pi}\sigma})^n \exp\{-\sum_{i=1}^n \frac{1}{2\sigma^2}(x_i-\mu)^2\} \\
&= (\frac{1}{\sqrt{2\pi}\sigma})^n \exp\{-\frac{n\mu^2}{2\sigma^2}+\frac{n\mu}{\sigma^2}\bar{x}-\sum_{i=1}^n \frac{x_i^2}{2\sigma^2}\} \\
&= (\frac{1}{\sqrt{2\pi}\sigma})^n \exp\{-\frac{n\mu^2}{2\sigma^2}\}\exp\{\frac{n\mu}{\sigma^2}\bar{x}-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2\},
\end{aligned}$$

第四节 统计中常用的分布族

只要使

$$a(\mu,\sigma^2)=(\frac{1}{\sqrt{2\pi}\sigma})^n\exp\{-\frac{n\mu^2}{2\sigma^2}\},\ \ Q_1(\mu,\sigma^2)=\frac{n\mu}{\sigma^2},$$

$$Q_2(\mu,\sigma^2)=-\frac{1}{2\sigma^2},$$

$$T_1(\mathbf{x})=\bar{x},\ \ T_2(\mathbf{x})=\sum_{i=1}^{n}x_i^2,\ \ h(\mathbf{x})=1$$

因此正态分布族$\{N(\mu,\sigma^2):-\infty<\mu<\infty,\sigma^2>0\}$ 是指数型分布族.

第五节 正态总体

第五节　正态总体

一 多元正态总体

1.定义：若随机向量X的联合分布密度函数为

$$f(x)=\frac{1}{(2\pi)^{\frac{n}{2}}|B|^{\frac{1}{2}}}\exp\{-\frac{1}{2}(x-a)^{'}B^{-1}(x-a)\}$$

其中B为正定阵,$|B|$为其行列式.B^{-1}是B的逆矩阵,则称随机向量X所服从的分布为**多元正态分布**,简记为$X\sim N_n(a,B)$.

2.特征函数：设随机向量$X\sim N_n(a,B)$,则其特征函数为

$$\varphi(t)=\exp\{ia^{'}t-\frac{1}{2}t^{'}Bt\}$$

其中$t'=(t_1,\cdots,t_n)$

第五节 正态总体

3.期望 方差 协方差阵

设$X=(X_1,\cdots,X_n)'$, $Y=(Y_1,\cdots,Y_m)'$是两个随机向量,
$Z=(Z_{ij})_{r\times s}$是随机矩阵,记

$$\begin{aligned} E(X) &= (E(X_1),\cdots,E(X_n))', \\ E(Z) &= (E(Z_{ij}))_{r\times s}, \end{aligned}$$

$$\begin{aligned} D(X) &= E(X-E(X))(X-E(X))' \\ &= \begin{pmatrix} D(X_1) & Cov(X_1,X_2) & \cdots & Cov(X_1,X_n) \\ Cov(X_2,X_1) & D(X_2) & \cdots & Cov(X_2,X_n) \\ \cdots & \cdots & \cdots & \cdots \\ Cov(X_n,X_1) & Cov(X_n,X_2) & \cdots & D(X_n) \end{pmatrix} \end{aligned}$$

其中$D(X_i)$为X_i的方差,$Cov(X_i,X_j)$为X_i和X_j的协方差.

第五节 正态总体

$$
\begin{aligned}
Cov(X,Y) &= (Cov(Y,X))' = E(X-E(X))(Y-E(Y))' \\
&= \begin{pmatrix} Cov(X_1,Y_1) & \cdots & Cov(X_1,Y_m) \\ \vdots & \ddots & \vdots \\ Cov(X_n,Y_1) & \cdots & Cov(X_n,Y_m) \end{pmatrix}
\end{aligned}
$$

4.$\rho_{ij}=\frac{Cov(X_i,X_j)}{\sqrt{D(X_i)D(X_j)}}$ 称为X_i与X_j之间的线性相关系数,简称为相关系数.

第五节 正态总体

5.性质

(1)$X \sim N_n(a,B)$, 则对X的任一子向量$\tilde{X}' = (X_{k1},...,X_{km})(m \le n)$, 有$\tilde{X} \sim N_m(\tilde{a},\tilde{B})$, 其中$\tilde{a} = (a_{k1},...,a_{km})'$, $\tilde{B}$是B中保留$k1,...,km$行列所得的m阶子矩阵. 特别地, $X_j \sim N(a_j, b_{jj}), j = 1,2,...,n$.

(2)$X \sim N_n(a,B)$, 则$EX = a, VarX = B$.

(3)若$X = (X_1,X_2)' \sim N_n(a,B)$, X_1是n_1维向量, X_2是n_2维向量, $n_1 + n_2 = n$, $X_1 \sim N_{n_1}(a_1,B_{11}), X_2 \sim N_{n_2}(a_2,B_{22})$, 其中$a_1 \in R^{n_1}, a_2 \in R^{n_2}$, $B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$, 则$X_1,X_2$相互独立$\Leftrightarrow Cov(X_1,X_2) = B_{12} = 0$.

第五节 正态总体

(4)$X \sim N_n(a, B), r(A) = m, A \in R^{m\times n}, b = (b_1, \ldots, b_m)' \in R^m$,
则$Y = AX + b \sim N_m(Aa + b, ABA')$.

(5)若$X \sim N_n(a, B)$, 则存在一个正交变换Γ, 使得$Y = \Gamma(X - a)$的各分量是相互独立，均值都为零的正态分布随机变量。特别地，若$X \sim N_n(0, \sigma^2 I_n)$, 则$Y = \Gamma X \sim N_n(0, \sigma^2 I_n)$.

第五节 正态总体

二 正态总体统计量的分布

定理： 设$(X_1,\cdots,X_n)$是取自正态总体$N(\mu,\sigma^2)$的一个样本，$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$，　$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 分别为样本均值和样本方差，则

(1) $\overline{X}\sim N(\mu,\frac{\sigma^2}{n})$;

(2) $\dfrac{\sum\limits_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}\sim\chi^2(n)$;

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

(4) $\overline{X}$ 和 S^2相互独立.

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

第五节　正态总体

推论2：设$(X_1,\cdots,X_{n_1})$是取自正态总体$N(\mu_1,\sigma_1^2)$的一个样本,$(Y_1,\cdots,Y_{n_2})$ 是取自正态总体$N(\mu_2,\sigma_2^2)$的一个样本,且$(X_1,\cdots,X_{n_1})$和$(Y_1,\cdots,Y_{n_2})$相互独立,则

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

(2) 若 $\sigma_1^2=\sigma_2^2=\sigma^2$, 则

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

其中$\overline{X}=\frac{1}{n_1}\sum_{i=1}^{n_1}X_i$　　$\overline{Y}=\frac{1}{n_2}\sum_{i=1}^{n_2}Y_i$,

$S_1^2=\frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i-\overline{X})^2$, $S_2^2=\frac{1}{n_2-1}\sum_{i=1}^{n_2}(Y_i-\overline{Y})^2$,

$S_w^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$, S_w^2称为二个样本的合并方差.

第六节　充分统计量和完备统计量

第六节 充分统计量和完备统计量

一.充分统计量

例1:了解产品的不合格率p,检验员随机抽取了10件产品进行检查,发现第3件和第9件为不合格品,记作$X_3=1, X_9=1$,其余都是合格品,记$X_i=0,\ i=1,2,\cdots,10,$ 且$i\neq 3,\ i\neq 9$.当领导问及检验结果时,检验员作了如下二种回答:

(1) 10件产品中有2件不合格品,即$\sum_{i=1}^{10} X_i=2$;

(2)第9件产品不合格,即$X_9=1$.

第六节 充分统计量和完备统计量

1.定义：设$(X_1,\cdots,X_n)$是从具有分布族$\{F(x;\theta):\theta\in\Theta\}$的总体中抽取的一个样本，$T(X_1,\cdots,X_n)$是一统计量.如果在给定$T(X_1,\cdots,X_n)=t$下,$(X_1,\cdots,X_n)$的条件分布与未知参数$\theta$无关，则称统计量$T(X_1,\cdots,X_n)$是分布族$\{F(x;\theta):\theta\in\Theta\}$ 的充分统计量,或称它是θ的**充分统计量**

例2：证明$T_1=\sum_{i=1}^{n}X_i$是p的充分统计量,而$T_2=X_j$不是充分统计量.

第六节 充分统计量和完备统计量

对于两点分布总体,大小为n的样本的联合分布为

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

其中x_i非0即1. 统计量$T_1=\sum_{i=1}^n X_i$服从二项分布$b(n,p)$,所以在给定$T_1=\sum_{i=1}^n X_i=k$下,样本$(X_1,\cdots,X_n)$的条件分布

$$\begin{aligned}
&\mathrm{P}\{X_1=x_1,\cdots,X_n=x_n|T_1=k;p\}\\
=&\frac{\mathrm{P}\{X_1=x_1,\cdots,X_n=x_n,\sum_{i=1}^n X_i=k;p\}}{P\{T_1=k;p\}}\\
=&\frac{p^k(1-p)^{n-k}}{\binom{n}{k}p^k(1-p)^{n-k}}=\frac{1}{\binom{n}{k}},
\end{aligned}$$

与p无关,即这个条件分布已不包含有关p的任何信息了.

第六节 充分统计量和完备统计量

故$T_1=\sum_{i=1}^{n}X_i$是p的充分统计量,或T_1是两点分布族的充分统计量.

在给定$T_2=X_j=k$条件下,样本$(X_1,\cdots,X_n)$的条件分布

$$P\{X_1=x_1,\cdots,X_n=x_n|T_2=k;p\}=p^{\sum\limits_{i\neq j}x_i}(1-p)^{n-\sum\limits_{i\neq j}x_i-1}$$

与p有关,可见$T_2=X_j$不是充分统计量.

第六节 充分统计量和完备统计量

2.定理：(因子分解定理)设总体X为连续型随机变量,具有分布密度族$\{f(x;\theta):\theta\in\Theta\}$, $(X_1,\cdots,X_n)$取自X的一个样本,则统计量$T(X_1,\cdots,X_n)$为θ的充分统计量的充分必要条件为：样本的联合概率密度函数可分解为

$$
\begin{aligned}
L(x_1,\cdots,x_n;\theta) &= \prod_{i=1}^{n} f(x_i;\theta) \\
&= h(x_1,\cdots,x_n)g(T(x_1,\cdots,x_n),\theta)
\end{aligned}
$$

其中$h(x_1,\cdots,x_n)$是非负函数且与θ无关,$g(T(x_1,\cdots,x_n),\theta)$仅通过$T(x_1,\cdots,x_n)=t$依赖于$(x_1,\cdots,x_n)$.

第六节 充分统计量和完备统计量

例3：设$(X_1, \cdots, X_n)$是取自二点分布$b(1,p)$总体的一个样本,其联合概率分布为

$$\begin{aligned} P\{X = x_1, \cdots, X_n = x_n; p\} &= p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i} \\ &= (1-p)^n(\frac{p}{1-p})^{\sum_{i=1}^n x_i} \end{aligned}$$

取

$$T(x_1, \cdots, x_n) = \sum_{i=1}^n x_i, \ \ h(x_1, \cdots, x_n) = 1,$$

$$g(T(x_1, \cdots, x_n), p) = (1-p)^n(\frac{p}{1-p})^T$$

则

$$P\{X = x_1, \cdots, X_n = x_n; p\} = h(x_1, \cdots, x_n) \cdot g(T(x_1, \cdots, x_n), p),$$

由因子分解定理可知$T(X_1, \cdots, X_n) = \sum_{i=1}^n X_i$是$p$的充分统计量.

第六节 充分统计量和完备统计量

3.若分布有n个参数,则θ是参数向量,若定理条件成立,则称$T(X_1,\cdots,X_n)$ 为关于θ的**联合充分统计量**

例4.设$X\sim N(\mu,\sigma^2)$,$\theta=(\mu,\sigma^2)$是未知参数向量, 证明: $(\overline{X},\sum_{i=1}^n X_i^2)$是$(\mu,\sigma^2)$的联合充分统计量.

第六节 充分统计量和完备统计量

证：$(X_1,\cdots,X_n)$是取自总体的一个样本,其联合密度函数为

$$
\begin{aligned}
L(x_1,\cdots,x_n;\theta) &= \prod_{i=1}^{n} f(x_i;\theta) \\
&= \frac{1}{(\sqrt{2\pi}\sigma)^n}\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\} \\
&= \frac{1}{(\sqrt{2\pi}\sigma)^n}\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2+\frac{n\mu}{\sigma^2}\overline{x}-\frac{n\mu^2}{2\sigma^2}\}.
\end{aligned}
$$

取$h(x_1,\cdots,x_n)=1$,

$$g(T(x_1,\cdots,x_n),\theta)=\frac{1}{(\sqrt{2\pi}\sigma)^n}\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2+\frac{n\mu}{\sigma^2}\overline{x}-\frac{n\mu^2}{2\sigma^2}\}$$

由因子分解定理可知

$T(X_1,\cdots,X_n)=(\overline{X},\sum_{i=1}^{n}X_i^2)$是$\theta=(\mu,\sigma^2)$的联合充分统计量.

第六节 充分统计量和完备统计量

注:$T(X_1,\cdots,X_n)$是θ的联合充分统计量不能推出T_i是θ_i的充分统计量. 因此不能说明$\overline{X},\sum_{i=1}^n X_i^2$分别是$\mu,\sigma^2$的充分统计量.

4.定理: 设$T(X_1,X_2,\cdots,X_n)$是θ的充分统计量, $v=\psi(t)$是单值可逆函数, 则$V=\psi(T)$也是θ的充分统计量.

第六节 充分统计量和完备统计量

二.完备统计量

1.定义: 设$\{F(x;\theta):\theta\in\Theta\}$是一个分布族, 如果由$\mathrm{E}_\theta\left[g(X)\right]=0,\forall\theta\in\Theta$, 总可推出$\mathrm{P}_\theta\{g(X)=0\}=1,\forall\theta\in\Theta$, 则称分布族$\{F(x;\theta):\theta\in\Theta\}$是完备的.

例1. 二项分布族$\{b(n,p):0<p<1\}$是完备的.

例2. 正态分布族$\{N(0,\sigma^2):\sigma^2>0\}$是不完备的.

第六节 充分统计量和完备统计量

2.定义：设$(X_1,\cdots,X_n)$是取自分布族$\{F(x;\theta):\theta\in\Theta\}$的一个样本，若统计量$T(X_1,\cdots,X_n)$的对应分布族$\{F^T(t;\theta):\theta\in\Theta\}$是完备的，则称$T$是完备的.

注：T完备时，原分布族$\{F(x;\theta):\theta\in\Theta\}$可能不完备.

例3. 证明：对分布族$\{N(0,\sigma^2):\sigma^2>0\}$，$T=\sum_{i=1}^n X_i^2$是完备统计量.

例4. 设$(X_1,\cdots,X_n)$是取自均匀分布族$\{U(0,\theta):0<\theta<1\}$的一个样本，证明统计量$T=\max\limits_{1\le i\le n}X_i$是$\theta$的充分完备统计量.

第六节 充分统计量和完备统计量

证: $(X_1, \cdots, X_n)$的联合密度函数为

$$L(x_1, x_2, \cdots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta) = \frac{1}{\theta^n} \mathbf{1}_{\{0 < \max\limits_{1 \le i \le n} x_i < \theta\}}.$$

由因子分解定理, $T = \max\limits_{1 \le i \le n} X_i$是$\theta$的充分统计量.

设T的分布函数为$F^T(t; \theta)$, 对$0 < t < \theta$,

$$F^T(t; \theta) = P\{\max_{1 \le i \le n} X_i \le t; \theta\} = [F(t; \theta)]^n = (\frac{t}{\theta})^n.$$

$\therefore f^T(t; \theta) = \frac{nt^{n-1}}{\theta^n}, 0 < t < \theta.$

第六节 充分统计量和完备统计量

设对$\forall 0<\theta<1$, $g(t)$满足

$$E_\theta[g(T)] = \int_0^\theta g(t) f^T(t;\theta)\mathrm{d}t = \frac{n}{\theta^n}\int_0^\theta g(t)t^{n-1}\mathrm{d}t = 0,$$

即$\int_0^\theta g(t)t^{n-1}\mathrm{d}t = 0, \forall 0<\theta<1$, 两边关于$\theta$求导,

$$g(\theta)\theta^{n-1} = 0, a.s., \forall 0<\theta<1, \quad \therefore g(\theta) = 0, a.s. \forall 0<\theta<1.$$

所以, T的对应分布族完备, $T = \max\limits_{1\le i\le n} X_i$是$\theta$的充分完备统计量.

$\square$

第六节 充分统计量和完备统计量

3.定理: 设$\{f(x;\theta):\theta=(\theta_1,\ldots,\theta_k)\in\Theta\}$是含$k$个参数的指数族,样本$(X_1,\cdots,X_n)$的联合分布密度具有如下形式

$$L(x_1,\cdots,x_n;\theta)=a(\theta)\exp\{\sum_{j=1}^{k}Q_j(\theta)T_j(x_1,\cdots,x_n)\}h(x_1,\cdots,x_n),$$

如果Θ包含一k维矩形,且$Q=(Q_1,\cdots,Q_k)$的值域包含一k维开集,则$T(X_1,\cdots,X_n)=(T_1(X_1,\cdots,X_n),\cdots,T_k(X_1,\cdots,X_n))$是$k$维参数向量$\theta=(\theta_1,\ldots,\theta_k)$ 的充分完备统计量.

第六节 充分统计量和完备统计量

例5. 设$(X_1,\cdots,X_n)$是取自Poisson分布族$\{P(\lambda):\lambda>0\}$的一个样本, 证明: $T(X_1,\cdots,X_n)=\sum_{i=1}^n X_i$是$\lambda$的充分完备统计量.

证明:

$$\begin{aligned} P\{X_1=x_1,\cdots,X_n=x_n;\lambda\} &= \mathrm{e}^{-n\lambda}\frac{\lambda^{\sum_{i=1}^n x_i}}{x_1!\cdots x_n!} \\ &= \mathrm{e}^{-n\lambda}\mathrm{e}^{\sum_{i=1}^n x_i \ln\lambda}\frac{1}{x_1!\cdots x_n!}. \end{aligned}$$

$\therefore \{P(\lambda):\lambda>0\}$是单参数指数族. $\Theta=\{\lambda:\lambda>0\}$包含一维开区间, $Q(\lambda)=\ln\lambda$的值域包含一维开区间, 由定理, $T(X_1,\cdots,X_n)=\sum_{i=1}^n X_i$是$\lambda$的充分完备统计量. $\square$

第六节 充分统计量和完备统计量

例6. 设$(X_1,\cdots,X_n)$是取自正态分布族$\{N(\mu,\sigma^2):-\infty<\mu<\infty,\sigma^2>0\}$的一个样本，证明：$(\sum_{i=1}^n X_i,\sum_{i=1}^n X_i^2)$是$(\mu,\sigma^2)$的充分完备统计量.

证明：$(X_1,\cdots,X_n)$的联合密度函数为

$$\begin{aligned}L(x_1,\cdots,x_n;\mu,\sigma^2) &= (\frac{1}{\sqrt{2\pi}\sigma})^n\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\mu)^2\} \\ &= (\frac{1}{\sqrt{2\pi}\sigma})^n\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2+\frac{\mu}{\sigma^2}\sum_{i=1}^n x_i\} \\ &\quad\cdot\exp\{-\frac{n\mu}{2\sigma^2}\}.\end{aligned}$$

$\{(\mu,\sigma^2):-\infty<\mu<\infty,\sigma^2>0\}$包含2维开集，
$Q(\mu,\sigma^2)=(\frac{\mu}{\sigma^2},-\frac{1}{2\sigma^2})$ 的值域包含2维开区间，所以，
$(\sum_{i=1}^n X_i,\sum_{i=1}^n X_i^2)$是$(\mu,\sigma^2)$的充分完备统计量. □