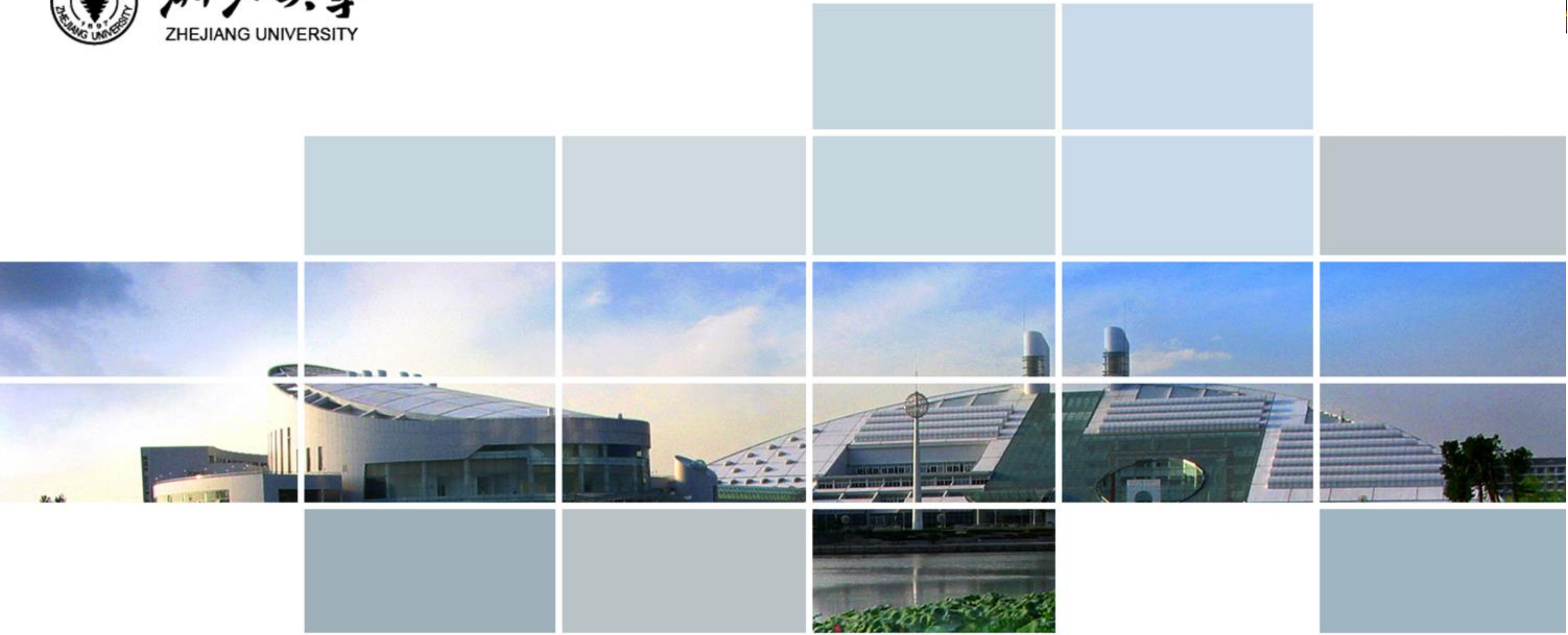
浙江大学
ZHEJIANG UNIVERSITY



第38讲　　总体与样本

数理统计学是一门以数据为基础的科学，可以定义为收集数据，分析数据和由数据得出结论的一组概念、原则和方法的学科.

例如：生产厂家声称他们生产的灯泡平均寿命不低于6000小时，如何验证厂家说法的真伪？由于灯泡寿命试验是破坏性试验，不可能把整批灯泡逐一检测，只能抽取一部分灯泡作为样本进行检验，以样本的信息来推断总体的信息，这是数理统计研究问题的基础。

- **总体**：研究对象的全体；
- **个体**：总体中的成员；
- **总体的容量**：总体中包含的个体数；
- **有限总体**：容量有限的总体；
- **无限总体**：容量无限的总体，通常将容量非常大的总体也按无限总体处理。

例：1）了解某校大学生“没有吃早饭习惯”的比例.

总体是该校大学生全体.这是一个有限总体，每个大学生有许多指标，我们关心的是每个学生是否有吃早饭习惯这一指标.

- 2）了解某城市的空气质量情况，调查该城市的PM2.5值。这是一个无限总体，描述空气质量有许多指标，而我们仅关心PM2.5值.
- 3）研究某种药物在人体中的吸收情况。这是一个有限总体，但数量非常巨大，我们常把它看出无限总体.

为了采用数理统计方法进行分析，首先要收集数据，数据收集方法一般有两种.

（1）通过调查、记录收集数据。如为了调查大学生是否有吃早饭习惯，可以进行问卷调查；要了解PM2.5值，需要在城市设立若干PM2.5监测站点，定时收集数据.

（2）通过实验收集数据。如为了了解药物吸收情况，要征集若干志愿者，把他们分成若干组，观察他们服药后不同时间点药物含量数据.

关于调查数据和实验数据的收集可以根据数据本身的特点有多种不同的方法和设计，有专门的课程讲授，这里不作详细介绍.

- 实际中人们往往只关注总体的某个指标.
- 总体的某个指标X，对于不同的个体来说有不同的取值，这些取值可以构成一个分布，因此X可以看成一个随机变量.
- 有时候就把X称为总体. 假设X的分布函数为F(x)，也称F(x)为总体.

例1:研究某品牌牛奶的质量合格情况，任取一包牛奶，以1表示合格，以0表示不合格，则总体

$X \sim B(1,p)$, p是合格率.

即$P(X=x)=p^x(1-p)^{1-x}, x=0,1.$

- 数理统计主要任务是从总体中抽取一部分个体，根据这部分个体的数据对总体分布给出推断.
- 被抽取的部分个体叫做总体的一个 样本.

- **简单随机样本**：满足以下两个条件的随机样本 $(X_1, X_2, \cdots, X_n)$ 称为容量是n的简单随机样本。

1 **代表性**：每个X_i与X同分布；

2 **独立性**：$X_1, X_2, \cdots, X_n$是相互独立的随机变量。

[说明]：后面提到的样本均指简单随机样本。

[注意]：(1)一个容量为n的样本$X_1, X_2, \ldots, X_n$是指n个独立与总体分布相同的随机变量.

(2)一旦对样本进行观察，得到实际数值$x_1, x_2, \ldots, x_n$称为样本观察值（或样本值）.

(3)两次观察，样本值可能是不同的.

- 如何取得的样本才称是简单随机样本?

对于有限总体，采用放回抽样就能得到简单随机样本.

但当总体容量很大的时候,放回抽样有时候很不方便，因此在实际中当总体容量比较大时,通常将不放回抽样所得到的样本近似当作简单随机样本来处理. 对于无限总体，一般采取不放回抽样.

例2:有四个同学参加了《概率论与数理统计》课程考试,成绩分别为88,75,70,63.现从中抽取容量为2的样本,列出全部的样本值.

答:共有16个样本值,见下表.

(88,88)	(88,75)	(88,70)	(88,63)
(75,88)	(75,75)	(75,70)	(75,63)
(70,88)	(70,75)	(70,70)	(70,63)
(63,88)	(63,75)	(63,70)	(63,63)

例3：为考察某校概率统计的成绩(5分制),从该校有放回地取两名学生，他们的成绩分别用X_1, X_2表示.设该校20%的同学成绩为3分，70%为4分,10%为5分.则总体X的分布律为：

X	3	4	5
p	0.2	0.7	0.1

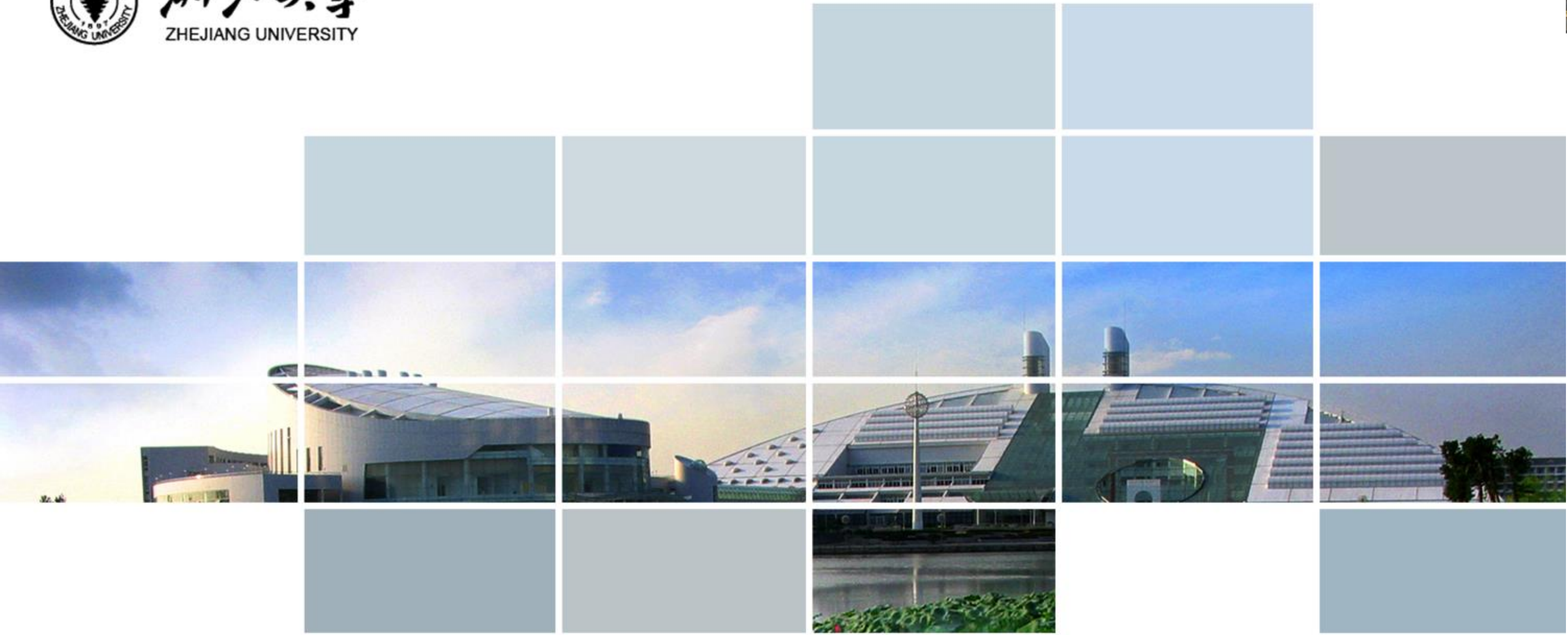
X_1, X_2是容量为2的简单样本，共有9个样本值，且

$$P(X_1 = x_1, X_2 = x_2) = P(X = x_1)P(X = x_2).$$

具体联合分布律见下表.

所有可能的样本观察值（共9个）

第一个观察值	第二个观察值		
	3	4	5
3	0.04	0.14	0.02
4	0.14	0.49	0.07
5	0.02	0.07	0.01



第39讲　　统计量与常用统计量

样本观察值往往是一堆杂乱无章的数据,不经过一定的整理,加工,就很难从样本中提取有用的信息来研究总体的分布及各种特征.

常用的整理加工数据的方法是构造各种统计量.

例 1: 在全校参加概率统计的学生中有放回地取20名学生,他们的成绩分别用$X_1, X_2, \ldots, X_{20}$表示. 那么你觉得全校的平均分该怎么估计呢?

若观察到这20名学生的成绩如下:

65, 68, 56, 98, 93, 78, 89, 86, 93, 68,

87, 45, 79, 93, 67, 88, 76, 90, 66, 88

你觉得全校的平均分大概是多少呢?

通常我们会用这20名同学的平均分来估计全校的平均分. 也就是令

$$\overline{X} = \frac{X_1 + X_2 + \dots + X_{20}}{20},$$

则用$\overline{X}$来估计全校的平均分.

特别地,对于这20个具体值而言,$\overline{X}$的观察值为:

$$\overline{x} = (\mathbf{65} + \ \mathbf{68} + \mathbf{56} + \dots + \ \mathbf{88}) \div 20 = 78.65,$$

因此全校的平均分估计值为78.65分.

- **统计量：**样本的不含任何未知参数的函数。

设$(X_1, X_2, \ldots, X_n)$为样本，若$g(X_1, X_2, \ldots, X_n)$不含任何未知参数，则称$g(X_1, X_2, \ldots, X_n)$为统计量.

一旦有了样本观察值$x_1, x_2, \ldots, x_n$,就可以算出统计量的具体值$g(x_1, x_2, \ldots, x_n)$.

上例中$\overline{X}$就是一个统计量.

常用统计量：

1. 样本均值 $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i,$

2. 样本方差 $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2,$

样本标准差 $S=\sqrt{S^2}$

常用统计量：

3. 样本矩　　k阶矩：$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$

k阶中心矩：$B_k = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^k$

$k = 1, 2, \ldots$

例2：设X为总体，$X_1,\ldots,X_n$是样本，$E(X)=\mu$存在，

则$\overline{X}=\mu$，对吗？

不对

$E(X)=\mu$是一个数，可能已知，可能未知；

$\overline{X}$是随机变量，依赖于样本值

对于不同的样本值，$\overline{X}$的取值可能不一样.

例3 接上一讲例2，总体为88，75，70，63，总体均值为74，计算全部16个样本值的样本均值.

样本编号	样本	样本均值	样本编号	样本	样本均值	样本编号	样本	样本均值
1	(88,88)	88	7	(75,70)	72.5	13	(63,88)	75.5
2	(88,75)	81.5	8	(75,63)	69	14	(63,75)	69
3	(88,70)	79	9	(70,88)	79	15	(63,70)	66.5
4	(88,63)	75.5	10	(70,75)	72.5	16	(63,63)	63
5	(75,88)	81.5	11	(70,70)	70	16个样本均值的平均为74		
6	(75,75)	75	12	(70,63)	66.5			

用样本均值估计总体均值，可能估计过高，也可能估计过低.

所有样本均值的平均值恰好是总体均值.（无偏）

例4：接上一讲例3, 全校学生中任取两个学生考察成绩,对应于X_1, X_2的具体观察值, $\bar{X}$的取值见下表：

所有可能的样本观察值（共9个）			
X_1	X_2		
	3	4	5
3	3	3.5	4
4	3.5	4	4.5
5	4	4.5	5

即观察到的样本均值可能是3，3.5，4，4.5，5.

而总体X的均值$\mu = 3\times 0.2 + 4\times 0.7 + 5\times 0.1 = 3.9.$

当总体数字特征未知时

- 用样本均值$\overline{X}$估计总体均值$\mu = E(X)$
- 用样本方差S^2估计总体方差$\sigma^2 = E(X-\mu)^2$
- 用样本原点矩A_k估计总体原点矩$\mu_k = E(X^k)$
- 用样本中心矩B_k估计总体中心矩$\nu_k = E(X-\mu)^k$

- 总体方差σ^2的估计可以用S^2,也可以用B_2

- $S^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$， $B_2=\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$

- 当$\sigma^2>0$时，$E(S^2)=\sigma^2, E(B_2)\neq\sigma^2$,

所以S^2是σ^2的无偏估计，而B_2是有偏估计

浙江大學
ZHEJIANG UNIVERSITY

第40讲 χ^2分布

- 在数理统计中，除了正态分布外，最重要的三个分布分别为：

χ^2-分布　$t-$分布　$F-$分布

定义: 设随机变量$X_1,\cdots,X_n$相互独立，都服从$N(0,1)$，

则称 $$\chi_n^2=\sum_{i=1}^{n}X_i^2 \qquad (1)$$

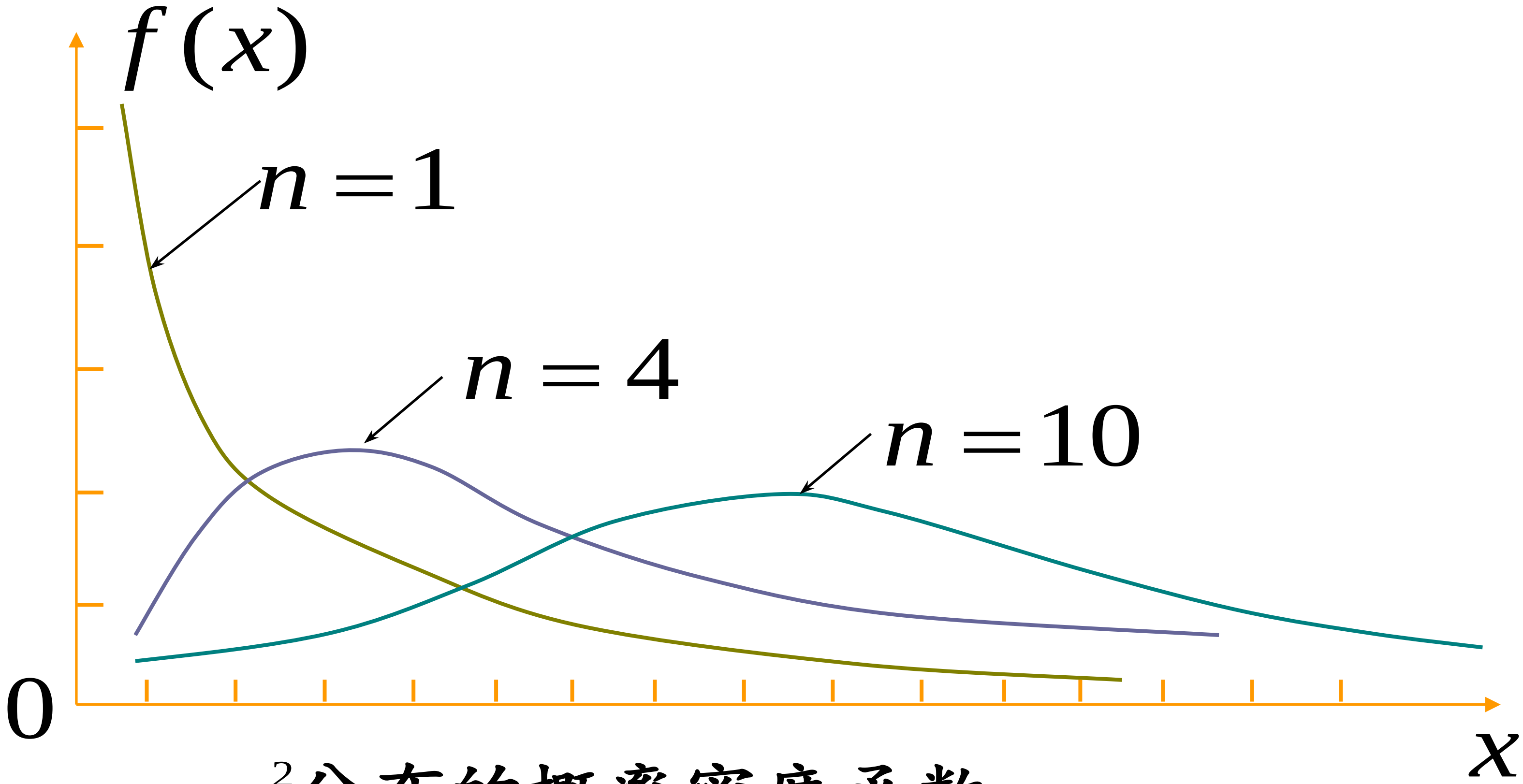
服从自由度为n的χ^2分布，记为$\chi^2 \sim \chi^2(n)$

自由度指(1)式右端包含的独立变量的个数.

$\chi^2(n)$分布的概率密度为：

$$f_n(y)=\begin{cases}\dfrac{1}{2\Gamma(n/2)}\left(\dfrac{y}{2}\right)^{\frac{n}{2}-1}e^{-\frac{y}{2}}, & y>0,\\ 0, & y\le 0.\end{cases}$$

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



χ^2分布的概率密度函数

性质：

1.设$\chi^2 \sim \chi^2(n)$,则$E(\chi^2)=n, D(\chi^2)=2n.$

2. χ^2分布的可加性：

设$Y_1 \sim \chi^2(n_1), Y_2 \sim \chi^2(n_2)$,且$Y_1, Y_2$相互独立，则

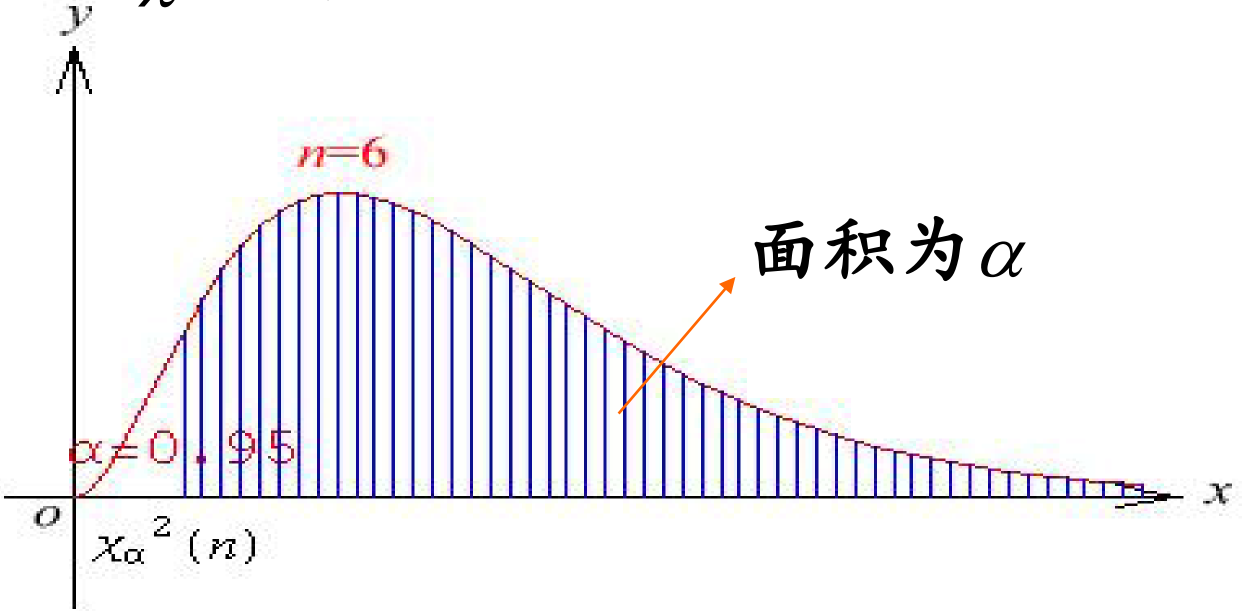
$$Y_1+Y_2 \sim \chi^2(n_1+n_2).$$

设$Y_1,\ldots,Y_m$相互独立，$Y_i \sim \chi^2(n_i)$,则 $\sum_{i=1}^{m} Y_i \sim \chi^2(\sum_{i=1}^{m} n_i).$

给定$\alpha, 0<\alpha<1$,称满足条件$P(\chi^2>\chi_\alpha^2(n))=\alpha$的点

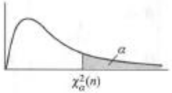
$\chi_\alpha^2(n)$为$\chi^2(n)$分布的上α分位数,

$\chi_\alpha^2(n)$的值可查χ^2分布表.



附表 4　χ^2 分布表

$P\{\chi^2(n) > \chi^2_\alpha(n)\} = \alpha$



n \ α	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.832	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.041	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.203	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.033	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.090	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.558
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.878	14.573	16.151	18.114	36.741	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.772	49.588	52.335
30	13.787	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
31	14.458	15.655	17.539	19.281	21.434	41.422	44.985	48.232	52.191	55.002
32	15.134	16.362	18.291	20.072	22.271	42.585	46.194	49.480	53.486	56.328
33	15.815	17.073	19.047	20.867	23.110	43.745	47.400	50.725	54.775	57.648
34	16.501	17.789	19.806	21.664	23.952	44.903	48.602	51.966	56.061	58.964
35	17.192	18.509	20.569	22.465	24.797	46.059	49.802	53.203	57.342	60.275
36	17.887	19.233	21.336	23.269	25.643	47.212	50.998	54.437	58.619	61.581
37	18.586	19.960	22.106	24.075	26.492	48.363	52.192	55.668	59.893	62.883
38	19.289	20.691	22.878	24.884	27.343	49.513	53.384	56.896	61.162	64.181
39	19.996	21.426	23.654	25.695	28.196	50.660	54.572	58.120	62.428	65.475
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766

当 $n > 40$ 时，$\chi^2_\alpha(n) \approx \frac{1}{2}(z_\alpha + \sqrt{2n-1})^2$.

例1：设总体$X \sim N\left(\mu,\sigma^2\right),\mu,\sigma^2$ 已知.

$\left(X_1,X_2,\cdots,X_n\right)$是取自总体$X$的样本.

求统计量 $\chi^2 = \dfrac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$ 的分布.

解：作变换 $Y_i = \dfrac{X_i - \mu}{\sigma}$ $i = 1, 2, \cdots, n$

则$Y_1, Y_2, \cdots, Y_n$相互独立，且$Y_i \sim N(0,1)$ $i = 1, 2, \cdots, n$

于是 $\chi^2 = \sum_{i=1}^{n} (\dfrac{X_i - \mu}{\sigma})^2 = \sum_{i=1}^{n} Y_i^2 \sim \chi^2(n).$

浙江大学
ZHEJIANG UNIVERSITY

第41讲　　t分布与F分布

$t-$分布

定义：设$X \sim N(0,1), Y \sim \chi^2(n)$，且$X$和$Y$相互独立.

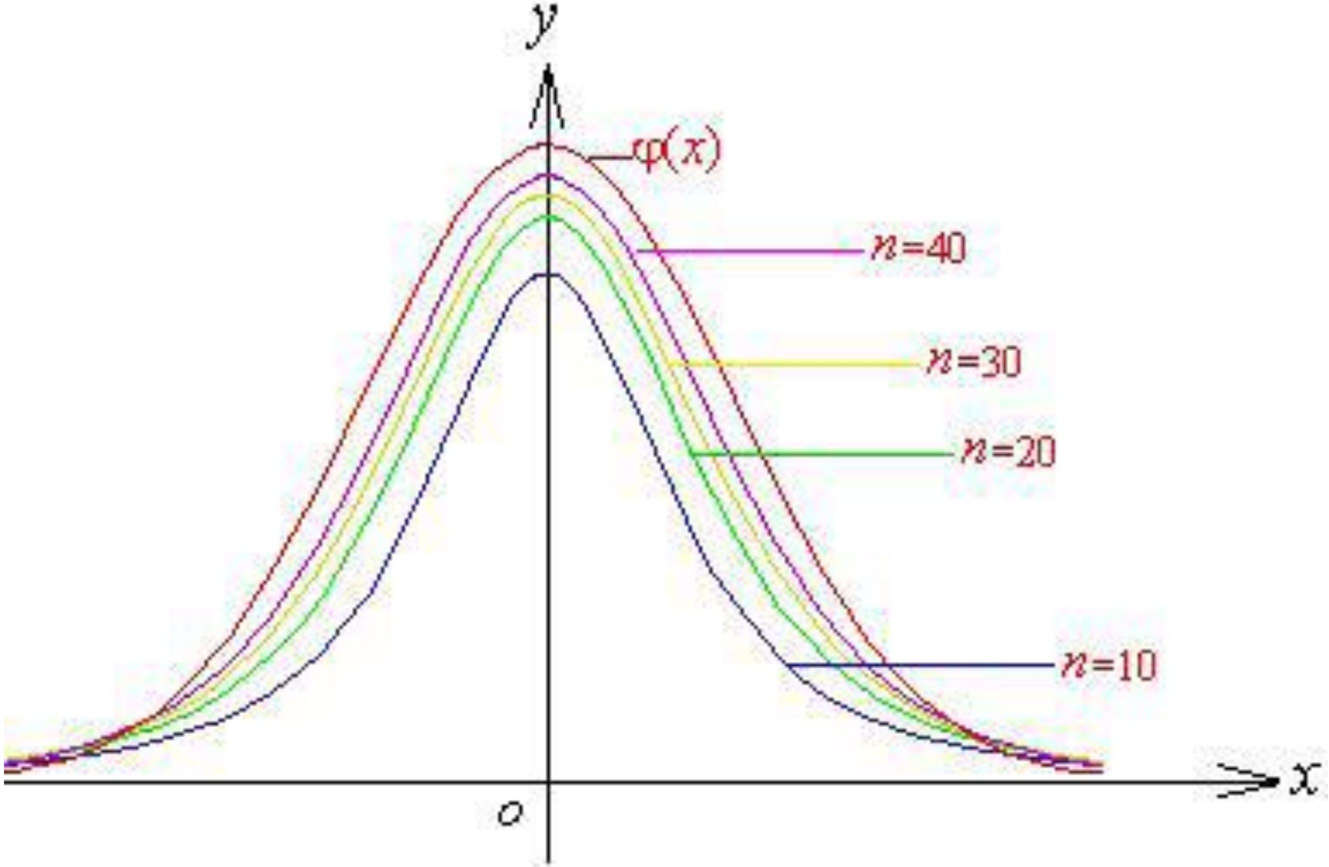
则称随机变量

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

服从自由度为n的t分布. 记为 $T \sim t(n)$.

$t(n)$分布的概率密度为：

$$f(t,n)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}},-\infty<t<+\infty$$

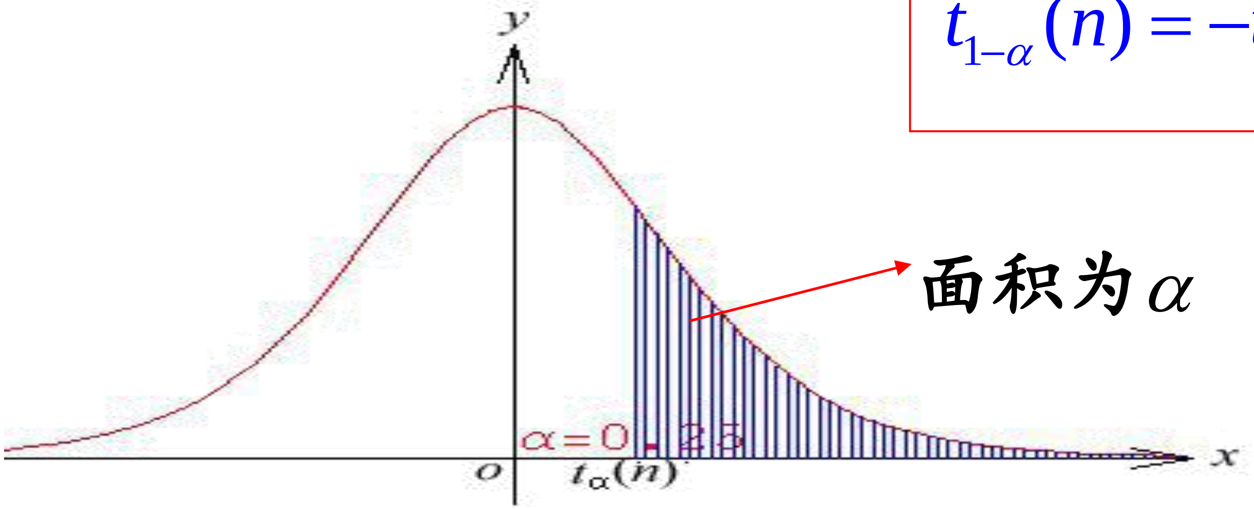


$t(n)$分布概率密度函数

给定α, $0<\alpha<1$,称满足条件$\int_{t_\alpha(n)}^{\infty} f(t,n)dt=\alpha$的点

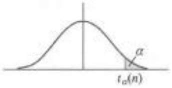
$t_\alpha(n)$为$t(n)$分布的上α**分位数**. $t_\alpha(n)$可查t分布表.

$$\boxed{t_{1-\alpha}(n)=-t_\alpha(n)}$$



面积为α

$P\{t(n) > t_\alpha(n)\} = \alpha$



		α						
		0.20	0.15	0.10	0.05	0.025	0.01	0.005
n	1	1.376	1.963	3.077 7	6.313 8	12.706 2	31.820 7	63.657 4
	2	1.061	1.386	1.885 6	2.920 0	4.302 7	6.964 6	9.924 8
	3	0.978	1.250	1.637 7	2.353 4	3.182 4	4.540 7	5.840 9
	4	0.941	1.190	1.533 2	2.131 8	2.776 4	3.746 9	4.604 1
	5	0.920	1.156	1.475 9	2.015 0	2.570 6	3.364 9	4.032 2
	6	0.906	1.134	1.439 8	1.943 2	2.446 9	3.142 7	3.707 4
	7	0.896	1.119	1.414 9	1.894 6	2.364 6	2.998 0	3.499 5
	8	0.889	1.108	1.396 8	1.859 5	2.306 0	2.896 5	3.355 4
	9	0.883	1.100	1.383 0	1.833 1	2.262 2	2.821 4	3.249 8
	10	0.879	1.093	1.372 2	1.812 5	2.228 1	2.763 8	3.169 3
	11	0.876	1.088	1.363 4	1.795 9	2.201 0	2.718 1	3.105 8
	12	0.873	1.083	1.356 2	1.782 3	2.178 8	2.681 0	3.054 5
	13	0.870	1.079	1.350 2	1.770 9	2.160 4	2.650 3	3.012 3
	14	0.868	1.076	1.345 0	1.761 3	2.144 8	2.624 5	2.976 8
	15	0.866	1.074	1.340 6	1.753 1	2.131 5	2.602 5	2.946 7
	16	0.865	1.071	1.336 8	1.745 9	2.119 9	2.583 5	2.920 8
	17	0.863	1.069	1.333 4	1.739 6	2.109 8	2.566 9	2.898 2
	18	0.862	1.067	1.330 4	1.734 1	2.100 9	2.552 4	2.878 4
	19	0.861	1.066	1.327 7	1.729 1	2.093 0	2.539 5	2.860 9
	20	0.860	1.064	1.325 3	1.724 7	2.086 0	2.528 0	2.845 3
	21	0.859	1.063	1.323 2	1.720 7	2.079 6	2.517 7	2.831 4
	22	0.858	1.061	1.321 2	1.717 1	2.073 9	2.508 3	2.818 8
	23	0.858	1.060	1.319 5	1.713 9	2.068 7	2.499 9	2.807 3
	24	0.857	1.059	1.317 8	1.710 9	2.063 9	2.492 2	2.796 9
	25	0.856	1.058	1.316 3	1.708 1	2.059 5	2.485 1	2.787 4
	26	0.856	1.058	1.315 0	1.705 6	2.055 5	2.478 6	2.778 7
	27	0.855	1.057	1.313 7	1.703 3	2.051 8	2.472 7	2.770 7
	28	0.855	1.056	1.312 5	1.701 1	2.048 4	2.467 1	2.763 3
	29	0.854	1.055	1.311 4	1.699 1	2.045 2	2.462 0	2.756 4
	30	0.854	1.055	1.310 4	1.697 3	2.042 3	2.457 3	2.750 0
	31	0.853 5	1.054 1	1.309 5	1.695 5	2.039 5	2.452 8	2.744 0
	32	0.853 1	1.053 6	1.308 6	1.693 9	2.036 9	2.448 7	2.738 5
	33	0.852 7	1.053 1	1.307 7	1.692 4	2.034 5	2.444 8	2.733 3
	34	0.852 4	1.052 6	1.307 0	1.690 9	2.032 2	2.441 1	2.728 4
	35	0.852 1	1.052 1	1.306 2	1.689 6	2.030 1	2.437 7	2.723 8
	36	0.851 8	1.051 6	1.305 5	1.688 3	2.028 1	2.434 5	2.719 5
	37	0.851 5	1.051 2	1.304 9	1.687 1	2.026 2	2.431 4	2.715 4
	38	0.851 2	1.050 8	1.304 2	1.686 0	2.024 4	2.428 6	2.711 6
	39	0.851 0	1.050 4	1.303 6	1.684 9	2.022 7	2.425 8	2.707 9
	40	0.850 7	1.050 1	1.303 1	1.683 9	2.021 1	2.423 3	2.704 5
	41	0.850 5	1.049 8	1.302 5	1.682 9	2.019 5	2.420 8	2.701 2
	42	0.850 3	1.049 4	1.302 0	1.682 0	2.018 1	2.418 5	2.698 1
	43	0.850 1	1.049 1	1.301 6	1.681 1	2.016 7	2.416 3	2.695 1
	44	0.849 9	1.048 8	1.301 1	1.680 2	2.015 4	2.414 1	2.692 3
	45	0.849 7	1.048 5	1.300 6	1.679 4	2.014 1	2.412 1	2.689 6

F分布

定义: 设$X \sim \chi^2(n_1), Y \sim \chi^2(n_2)$, 且$X, Y$独立, 则称随机

变量$F = \dfrac{X/n_1}{Y/n_2}$服从自由度为(n_1, n_2)的F分布,

记为　　$F \sim F(n_1, n_2)$,

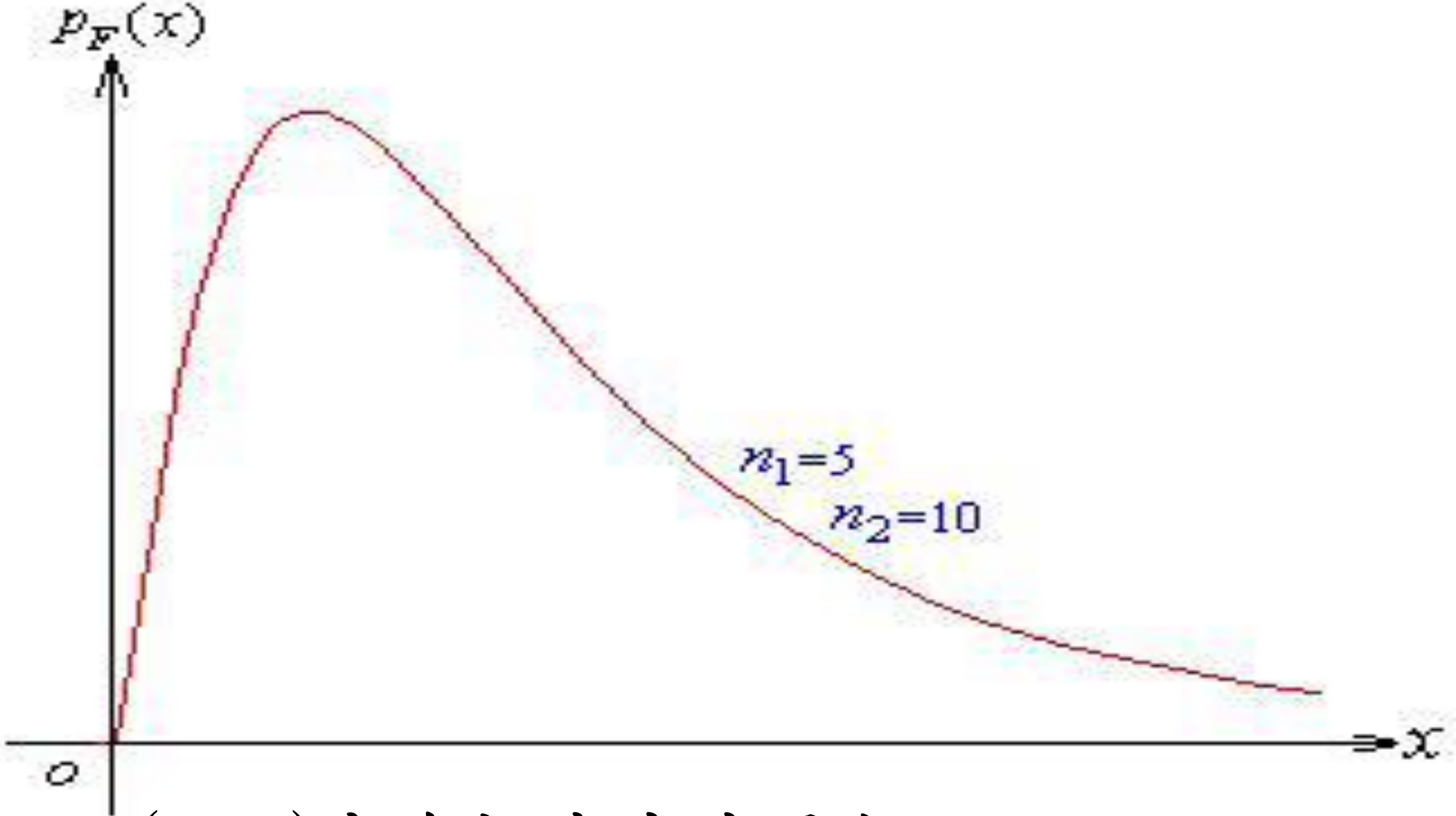
其中n_1称为第一自由度, n_2称为第二自由度.

性质: 若 $F \sim F(n_1, n_2)$,则 $\dfrac{1}{F} \sim F(n_2, n_1)$.

$F(n_1,n_2)$分布的概率密度为：

$$f(x;n_1,n_2)=\begin{cases}\dfrac{1}{B\left(\frac{n_1}{2},\frac{n_2}{2}\right)}n_1^{\frac{n_1}{2}}n_2^{\frac{n_2}{2}}x^{\frac{n_1}{2}-1}\left(n_2+n_1x\right)^{-\frac{n_1+n_2}{2}}, & x>0,\\ 0, & x\le 0.\end{cases}$$

其中，$B(a,b)=\int_0^1 x^{\alpha-1}(1-x)^{b-1}dx=\dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$



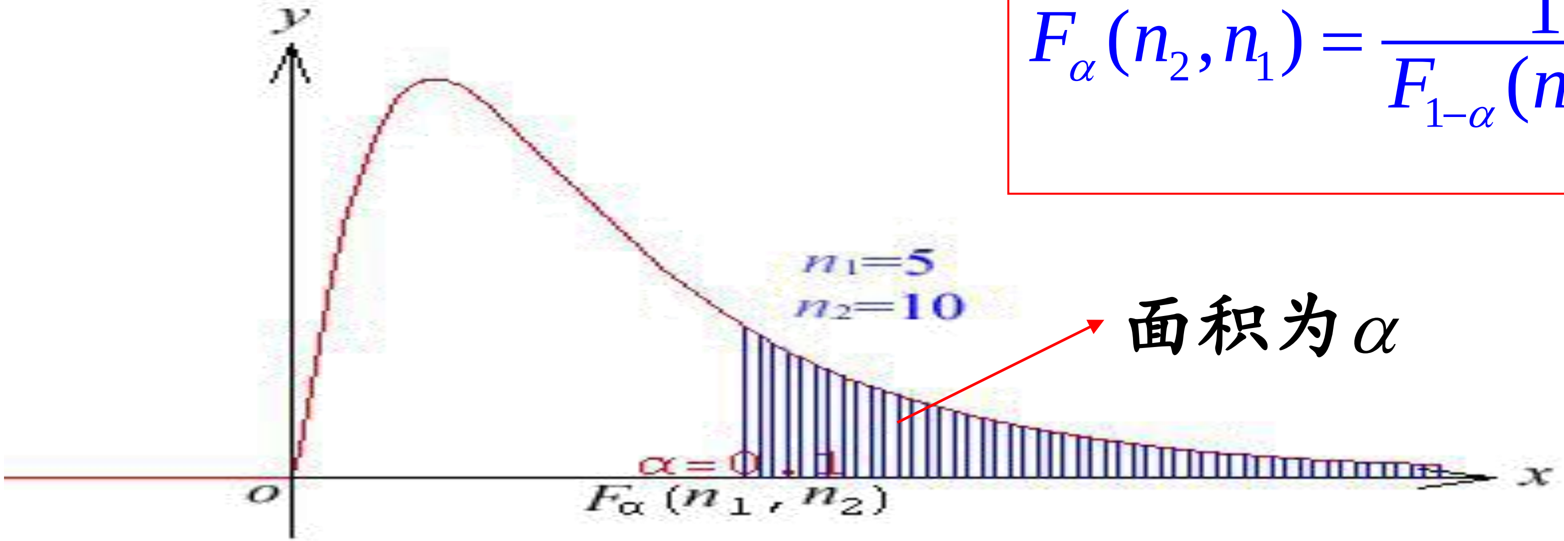
$F(5,10)$分布概率密度函数

给定 $0<\alpha<1$,称满足条件 $\int_{F_\alpha(n_1,n_2)}^{\infty} f(x;n_1,n_2)dx=\alpha$

的点 $F_\alpha(n_1,n_2)$ 为 $F(n_1,n_2)$ 分布的上 α 分位数.

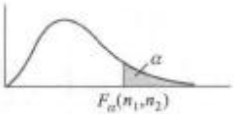
$F_\alpha(n_1,n_2)$ 可查 F 分布表得到.

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



附表 5　F 分布表

$P\{F(n_1,n_2) > F_\alpha(n_1,n_2)\} = \alpha$



($\alpha = 0.10$)

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00

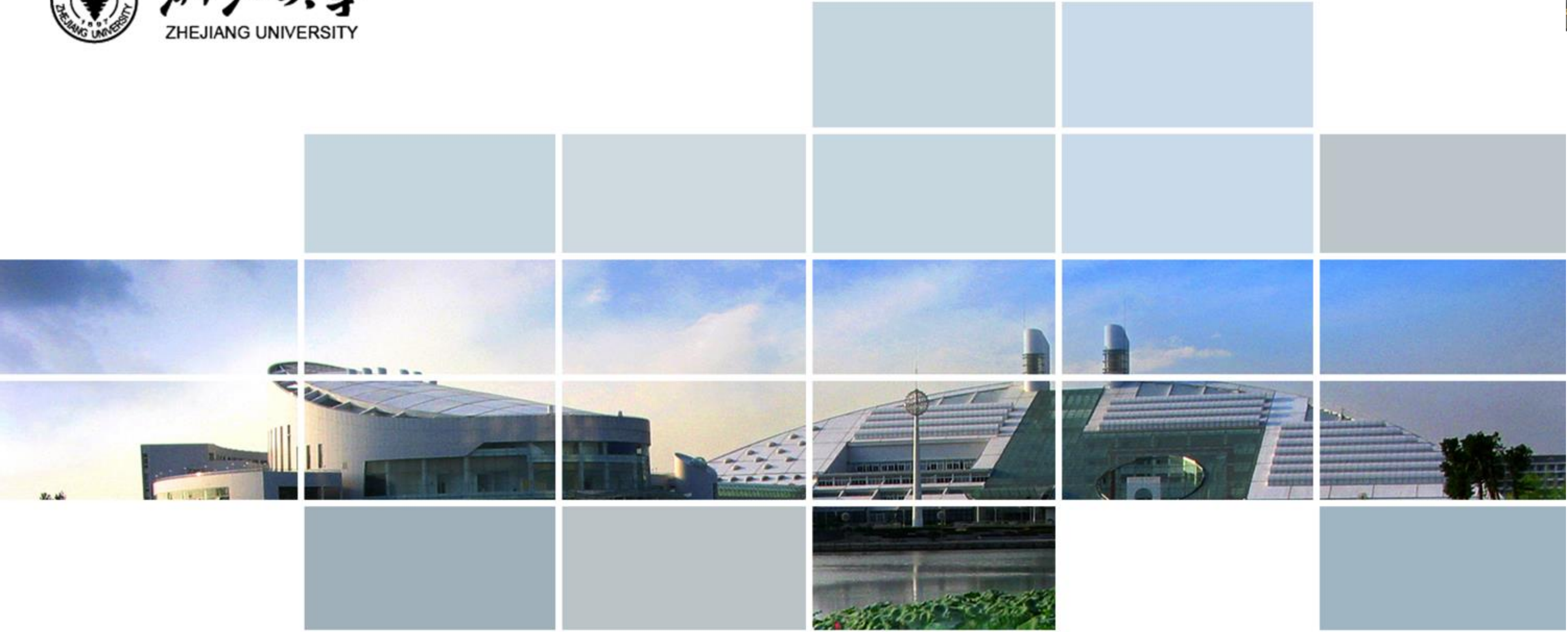
例: X,Y,Z相互独立, 均服从$N(0,1)$, 则

(1) $X^2+Y^2+Z^2 \sim \chi^2(3)$;

(2) $\dfrac{X}{\sqrt{(Y^2+Z^2)/2}} \sim t(2)$;

(3) $\dfrac{2X^2}{Y^2+Z^2} \sim F(1,2)$.

若$t \sim t(n)$,则 $t^2 \sim F(1,n)$.



第42讲 单个正态总体的抽样分布

- 统计量的分布称为抽样分布.

定理一：设总体$X \sim N(\mu, \sigma^2), X_1, X_2, \cdots, X_n$是样本，

样本均值$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$,样本方差$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

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

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

(1) 证明: $E(\bar{X}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu,$

$$D(\bar{X}) = D(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) = \frac{\sigma^2}{n},$$

$X_1, X_2, \ldots X_n$独立且都服从正态分布,

而且$\bar{X}$是$X_1, X_2, \ldots X_n$的线性组合

$\Rightarrow \bar{X}$ 服从正态分布, 即$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}).$

(2) 证略.

例1： 设总体$X \sim N(\mu, \sigma^2), X_1, X_2, \cdots, X_n$是样本，

(1) $\dfrac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2}$服从什么分布？

(2) $\dfrac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2}$服从什么分布？

答：(1) $\chi^2(n-1)$，　(2) $\chi^2(n)$.

由定理一(1)知，$\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$.

当σ未知时,可用S来近似σ,此时有

定理二: 设总体$X \sim N(\mu,\sigma^2), X_1, X_2, \cdots, X_n$是样本,

样本均值$\bar{X}=\dfrac{1}{n}\sum_{i=1}^{n}X_i$,样本方差$S^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$.

则 $\dfrac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$.

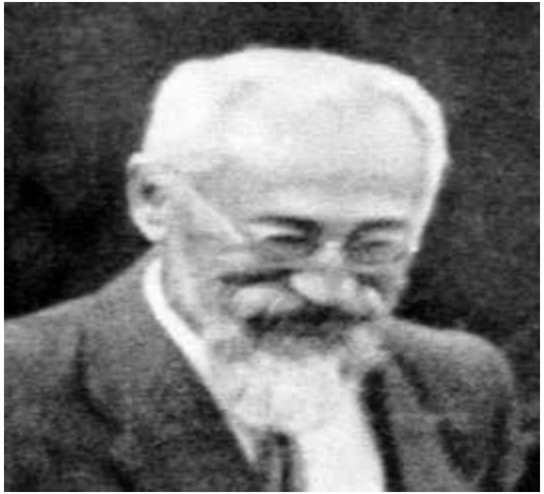
证明: $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1),\ \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

$\bar{X}$与S^2相互独立,

$$\therefore \frac{\bar{X}-\mu}{S/\sqrt{n}} = \frac{\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2(n-1)}}} \sim t(n-1).$$

William Gosset(1876-1937)

- 1908年提出t-分布



例2：设总体X的均值μ,方差σ^2存在.

$\left(X_1,\cdots,X_n\right)$是取自总体$X$的样本,

$\bar{X},S^2$为样本均值和样本方差.

(1) 求$E(S^2)$;

(2)若$X\sim N\left(\mu,\sigma^2\right)$, 求$D(S^2)$.

解(1): $\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n}(X_i^2 - 2\bar{X}X_i + \bar{X}^2)$

$$= \sum_{i=1}^{n} X_i^2 - 2\bar{X}\sum_{i=1}^{n} X_i + n\bar{X}^2$$

$$= \sum_{i=1}^{n} X_i^2 - 2\bar{X}n\bar{X} + n\bar{X}^2$$

$$= \sum_{i=1}^{n} X_i^2 - n\bar{X}^2$$

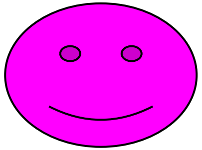
$$E(S^2) = E[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2]$$

$$= E[\frac{1}{n-1}(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2)]$$

$$= \frac{1}{n-1}[\sum_{i=1}^{n} E(X_i^2) - nE(\bar{X}^2)]$$

$$= \frac{1}{n-1}[\sum_{i=1}^{n}(\sigma^2 + \mu^2) - n(\frac{\sigma^2}{n} + \mu^2)] = \sigma^2$$

无偏！



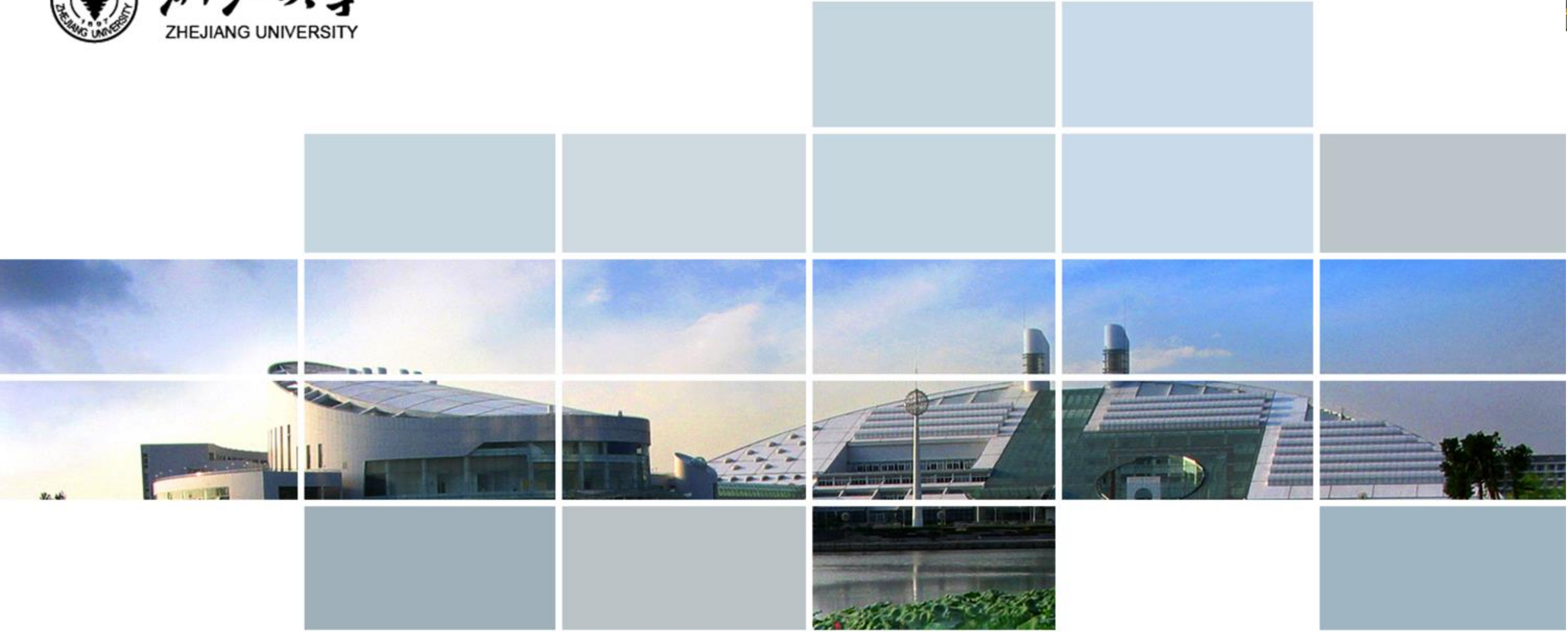
(2)因为总体$X \sim N\left(\mu,\sigma^2\right)$, 所以$\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$,

$$\Rightarrow D\left[\frac{(n-1)S^2}{\sigma^2}\right] = 2(n-1)$$

$$\Rightarrow D(S^2) = \frac{2\sigma^4}{n-1}.$$

随样本量n增大，$D(S^2)$减小.

浙江大学
ZHEJIANG UNIVERSITY



第43讲 两个正态总体的抽样分布

本讲中，设样本$\left(X_1,\cdots,X_{n_1}\right)$和$\left(Y_1,\cdots,Y_{n_2}\right)$分别来自总体$N\left(\mu_1,\sigma_1{}^2\right)$和$N\left(\mu_2,\sigma_2{}^2\right)$,并且它们相互独立.

样本均值分别为$\bar{X},\bar{Y}$;样本方差分别为S_1^2,S_2^2.

则可以得到下面三个抽样分布.

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

证明: $$\chi_1^2=\frac{(n_1-1)S_1^2}{\sigma_1^2}\sim\chi^2(n_1-1)$$

$$\chi_2^2=\frac{(n_2-1)S_2^2}{\sigma_2^2}\sim\chi^2(n_2-1)$$ S_1^2,S_2^2 独立,

$$\therefore F=\frac{S_1^2}{S_2^2}\Big/\frac{\sigma_1^2}{\sigma_2^2}=\frac{\chi_1^2/(n_1-1)}{\chi_2^2/(n_2-1)}\sim F(n_1-1,n_2-1)$$

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

证明：$\bar{X} \sim N(\mu_1, \frac{\sigma_1^2}{n_1}),\ \bar{Y} \sim N(\mu_2, \frac{\sigma_2^2}{n_2}),$

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

$$\Rightarrow [(\bar{X}-\bar{Y})-(\mu_1-\mu_2)] \Big/ \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \sim N(0,1).$$

（3）当$\sigma_1^2 = \sigma_2^2 = \sigma^2$时，

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

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

证明：当$\sigma_1^2=\sigma_2^2=\sigma^2$时，

由(2) $$\frac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_1-\mu_2\right)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim N(0,1)$$

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

$$\therefore\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{\sigma^2}\sim\chi^2(n_1+n_2-2),$$

由t分布定义，$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

$$=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\Bigg/\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{\sigma^2(n_1+n_2-2)}}$$

$$\sim t(n_1+n_2-2)$$

思考：若σ^2未知，为什么用S_w^2来估计σ^2，

而不用S_1^2或S_2^2来估计σ^2呢？

$E(S_1^2)=\sigma^2$ 无偏 $\qquad E(S_2^2)=\sigma^2$ 无偏

$D(S_1^2)=\dfrac{2\sigma^4}{n_1-1}$ $\qquad D(S_2^2)=\dfrac{2\sigma^4}{n_2-1}$ —第42讲例2

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

从直观来看，S_w^2比S_1^2或S_2^2包含更多σ^2的信息

$E(S_w^2)=\sigma^2$ 无偏

$D(S_w^2)=\dfrac{2\sigma^4}{n_1+n_2-2}$

比$D(S_1^2),D(S_2^2)$更小.

例：设总体$X\sim N\left(\mu,\sigma^2\right)$. $\left(X_1,X_2,\cdots,X_5\right)$是取自总体$X$的样本. 设$a,b$是都不为0的数.

若$a(X_1-X_2)^2+b(2X_3-X_4-X_5)^2\sim\chi^2(k)$,

则a,b,k各为多少？

解：$X_1 - X_2 \sim N(0, 2\sigma^2)$, $\dfrac{X_1 - X_2}{\sqrt{2}\sigma} \sim N(0,1)$

$2X_3 - X_4 - X_5 \sim N(0, 6\sigma^2)$, $\dfrac{2X_3 - X_4 - X_5}{\sqrt{6}\sigma} \sim N(0,1)$

$\dfrac{X_1 - X_2}{\sqrt{2}\sigma}$与$\dfrac{2X_3 - X_4 - X_5}{\sqrt{6}\sigma}$相互独立，

故$\dfrac{(X_1 - X_2)^2}{2\sigma^2} + \dfrac{(2X_3 - X_4 - X_5)^2}{6\sigma^2} \sim \chi^2(2)$

$\therefore a = \dfrac{1}{2\sigma^2}, b = \dfrac{1}{6\sigma^2}, k = 2.$

例：设总体$X \sim N\left(\mu,\sigma^2\right),\left(X_1,\cdots,X_4\right)$

与$\left(Y_1,\cdots,Y_9\right)$是取自总体$X$的两个独立样本，

$\bar{X},S_1^2$和$\bar{Y},S_2^2$分别为样本均值和样本方差；

求(1)若$a\dfrac{\bar{X}-\bar{Y}}{S_1} \sim t(k)$，则$a,k$各为多少？

(2)$\sum_{i=1}^{4}(X_i-\mu)^2 \Big/ 4S_2^2$服从什么分布？

解:(1) $\because \bar{X} \sim N(\mu, \frac{\sigma^2}{4}), \bar{Y} \sim N(\mu, \frac{\sigma^2}{9})$,且$\bar{X}$与$\bar{Y}$相互独立,

$$\therefore \bar{X} - \bar{Y} \sim N(0, \frac{13\sigma^2}{36}), \quad \frac{\pm 6(\bar{X} - \bar{Y})}{\sqrt{13}\sigma} \sim N(0,1)$$

又$\frac{3S_1^2}{\sigma^2} \sim \chi^2(3)$,且$\bar{X} - \bar{Y}$与$S_1^2$相互独立,

$$\therefore \frac{\pm 6}{\sqrt{13}} \frac{\bar{X} - \bar{Y}}{\sigma} \Big/ \sqrt{\frac{3S_1^2}{3\sigma^2}} = \frac{\pm 6\sqrt{13}}{13} \frac{\bar{X} - \bar{Y}}{S_1} \sim t(3)$$

$$\Rightarrow a = \pm \frac{6\sqrt{13}}{13}, k = 3.$$

$$(2)\frac{1}{\sigma^2}\sum_{i=1}^{4}(X_i-\mu)^2 \sim \chi^2(4), \quad \frac{8S_2^2}{\sigma^2} \sim \chi^2(8),$$

由F分布定义知，

$$\frac{1}{4\sigma^2}\sum_{i=1}^{4}(X_i-\mu)^2 \Bigg/ \frac{8S_2^2}{8\sigma^2} = \sum_{i=1}^{4}(X_i-\mu)^2 \Bigg/ 4S_2^2 \sim F(4,8).$$