3 抽样分布

3.1 引言：概率论如何认识随机现象的数量规律

3.2 概率分布界定

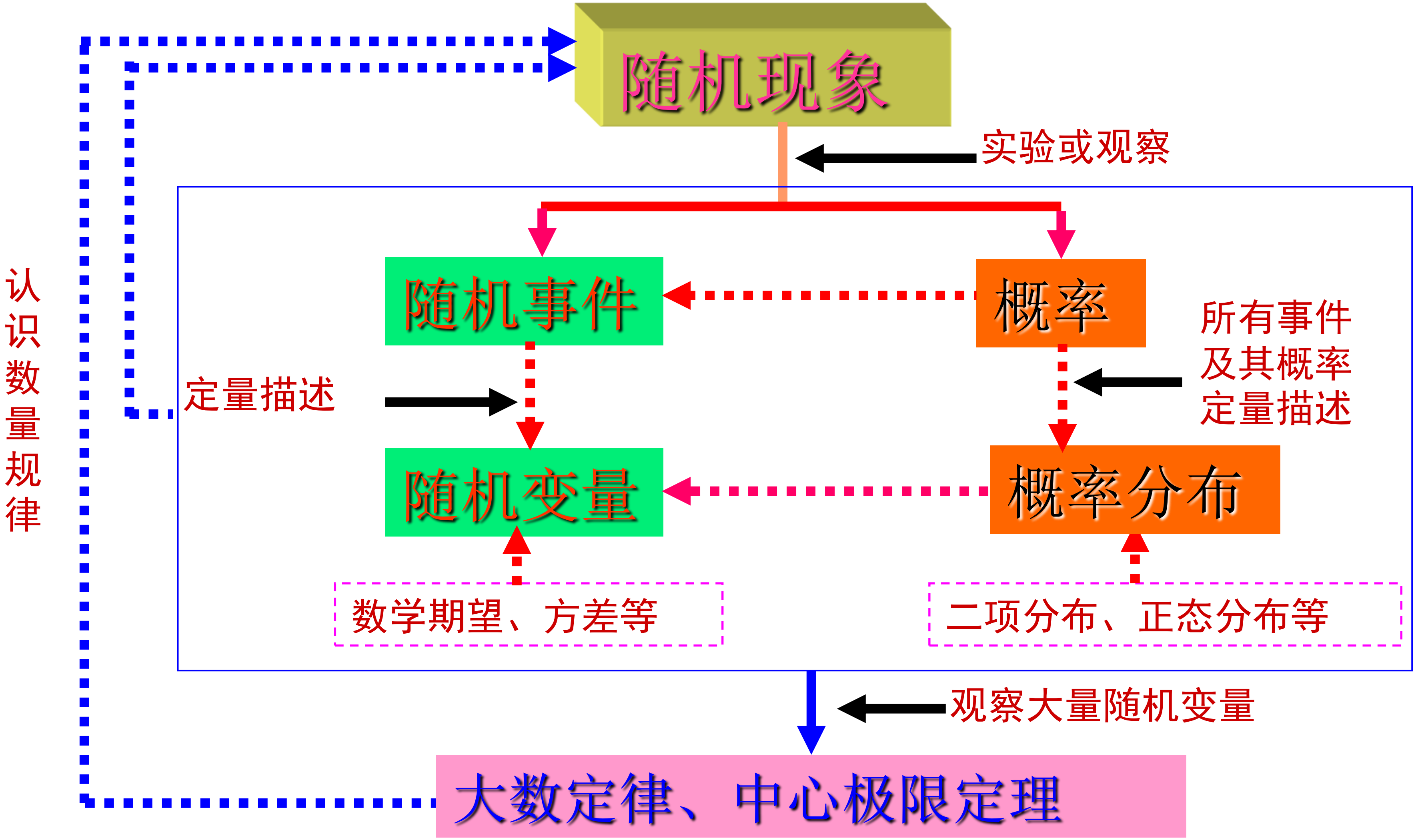
3.3 正态分布

3.4 T分布

3.5 F分布

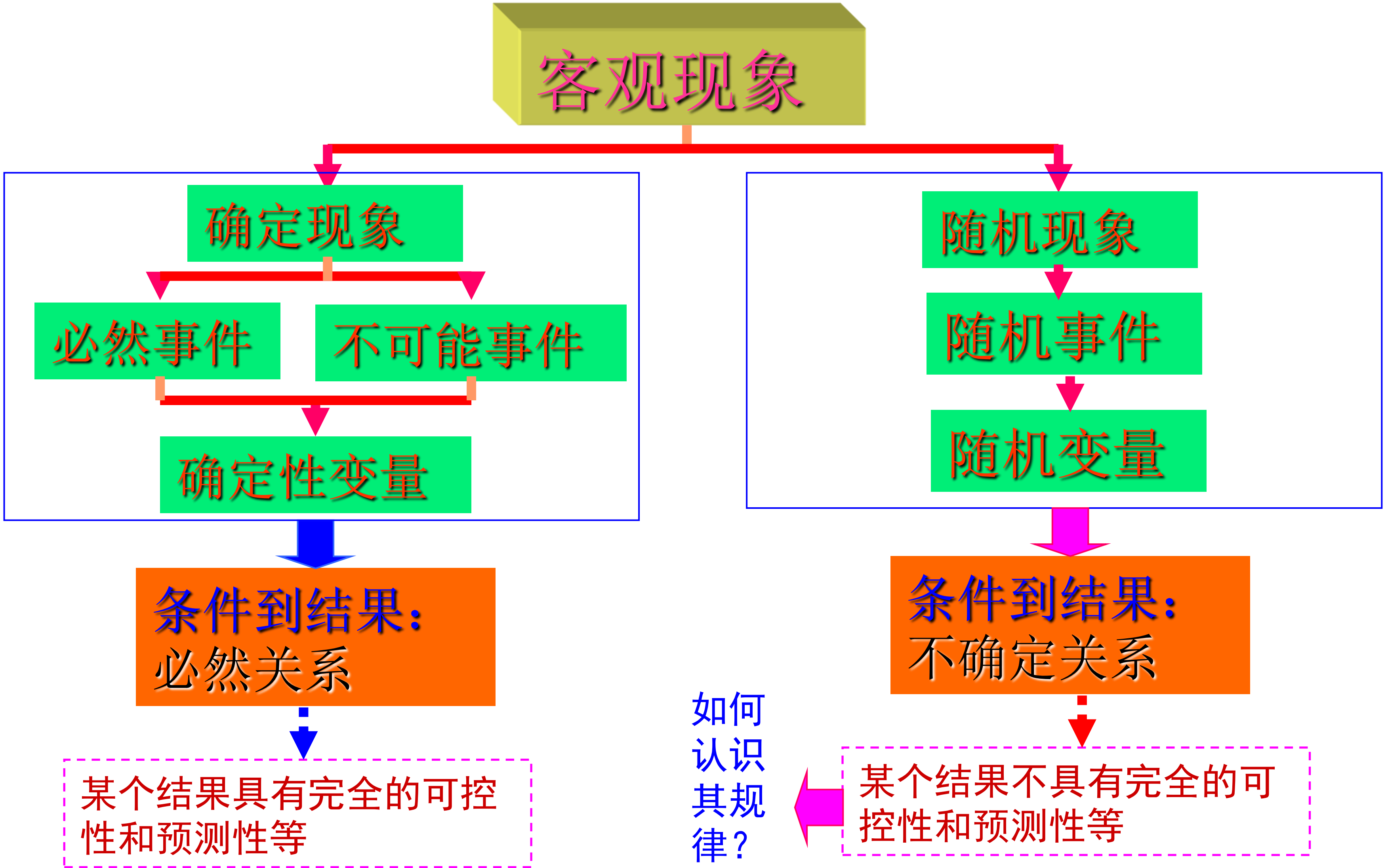
3.6 抽样与抽样分布

3.1 引言：概率论如何认识随机现象的数量规律



概率论知识结构体系

3.1 引言：概率论如何认识随机现象的数量规律（续）



3.1 引言：概率论如何认识随机现象的数量规律（续）

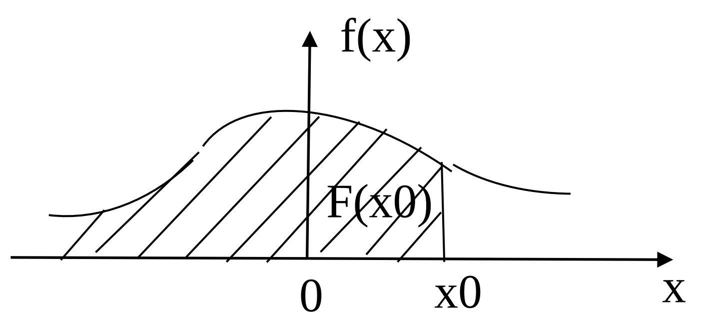
在相同条件下
大量试验或观察

统计规律

事件	事件1	事件2	…	事件n
概率	概率1	概率2	…	概率n

认识复合事件

X	X1	X2	…	Xn
p	p1	p2	…	pn

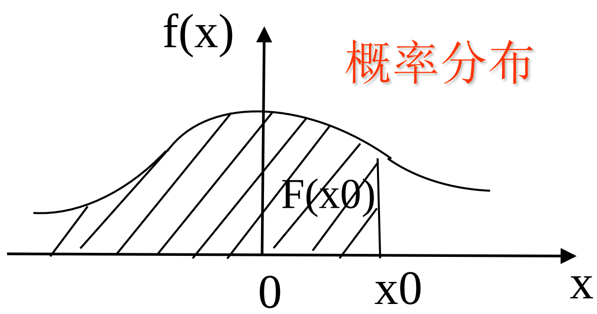


概率分布：随机变量所有可能取值的分布规律。

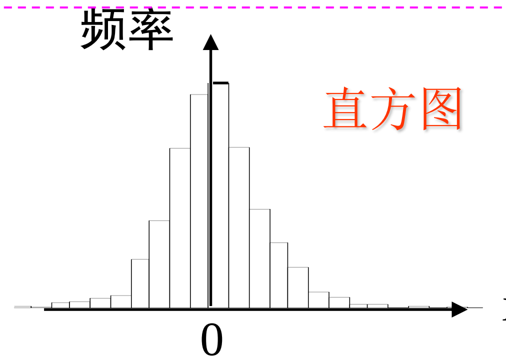
期望和方差：随机变量所有可能取值的平均水平和差异程度。

3.1 引言：概率论如何认识随机现象的数量规律（续）

总体N → **样本n**



$n \to N$



统计规律

总体均值：μ
总体方差：σ^2

$E(\bar{x}) = \mu$
$E(s^2) = \sigma^2$
$\bar{x} \sim N(\mu, \sigma^2/n)$

$n \to N$

样本均值：$\bar{x}$
样本方差：s^2

3.1 引言：概率论如何认识随机现象的数量规律（续）

大数定律的基本思想

- **当试验或观察的次数逐渐增大时，**
  - 某随机事件出现的频率m/n逐渐趋近于该随机事件发生的概率p，即

$$\lim_{n\to\infty}\frac{m}{n}=p。$$

  - 大量随机事件发生结果 $x_1, x_2, \cdots, x_n$的均值 $\bar{x}$ 逐渐趋近于这些结果的期望值 μ，即

$$\lim_{n\to\infty}\bar{x}=\mu,$$

其中，$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$，$\mu=E(X)$。

3.1 引言：概率论如何认识随机现象的数量规律（续）

- **保险的运作机制：风险集中**
  - **保险的必要性：**风险管理者通过风险集中将其风险转移给集体，但为节省交易成本，就需保险公司作为风险集中的纽带（否则风险单位之间必须两两签订协议）。风险管理者购买保险，其实就是参与大量风险单位的风险集中活动，保险公司就是这些大量风险单位的集体，这样风险管理者就可将其风险向保险公司转移；
  - **保险的可能性：**保险公司所承担的风险是大量风险单位组成的集体的风险，而**只要风险单位足够多**，集体的期望损失将趋于确定（大数定律），这样保险公司只需承担较小的风险，收取的保费只需用于确定的损失赔偿及其他费用的开支。

3.1 引言：概率论如何认识随机现象的数量规律（续）

中心极限定理的基本思想及应用

- **当试验或观察的次数逐渐增大时，**
  - 大量随机事件结果 $x_1, x_2, \cdots, x_n$ 的均值逐渐趋近于一个正态分布，即 $\bar{x} \sim N(\mu, \sigma^2/n)$

其中， $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$， $\mu = E(X)$， $\sigma^2 = D(X)$。

- **为样本统计量的抽样分布形式提供了理论基础**
  - 根据中心极限定理，在样本容量足够大时，从任意概率分布总体中抽取的随机样本的均值将近似服从均值为 μ、方差为 σ^2/n 的正态分布。
- **为大量的统计分析和应用提供了理论基础和关键结论**

3.2 概率分布界定

□ 随机变量的概率分布

➢ 离散型随机变量的概率分布

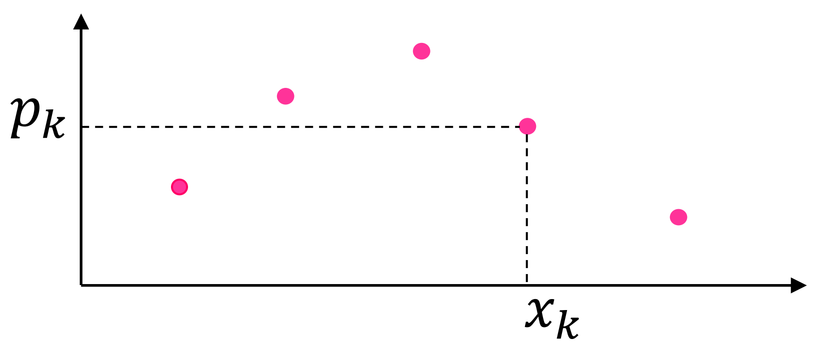
● 定义：以函数、表格、图形形式描述随机变量的可能取值及其相应概率的一种表达方式。

函数形式： $P\{X=x_k\}=p_k, k=1,2,\cdots,n$

表格形式：

X	x_1	x_2	…	x_k	…	x_n
p	p_1	p_2	…	p_k	…	p_n

图形形式：



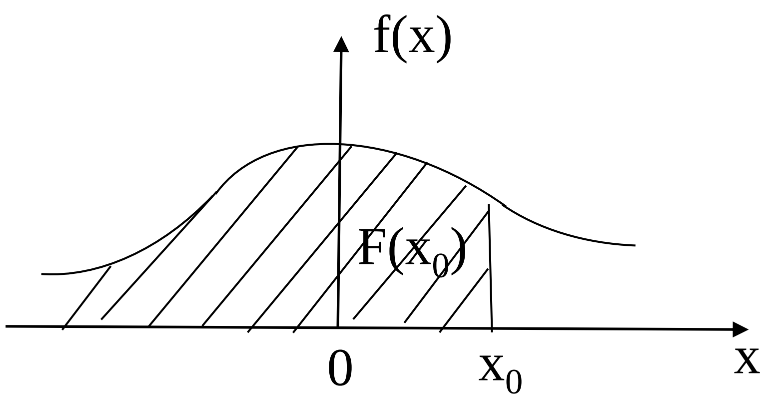
3.2 概率分布界定（续）

➢ 连续型随机变量的概率密度函数（PDF）

- 定义：对于任意实数x，存在一个非负函数f(x)，使得随机变量X的分布函数F(X) 满足：

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

则f(x)为随机变量X的概率密度函数。



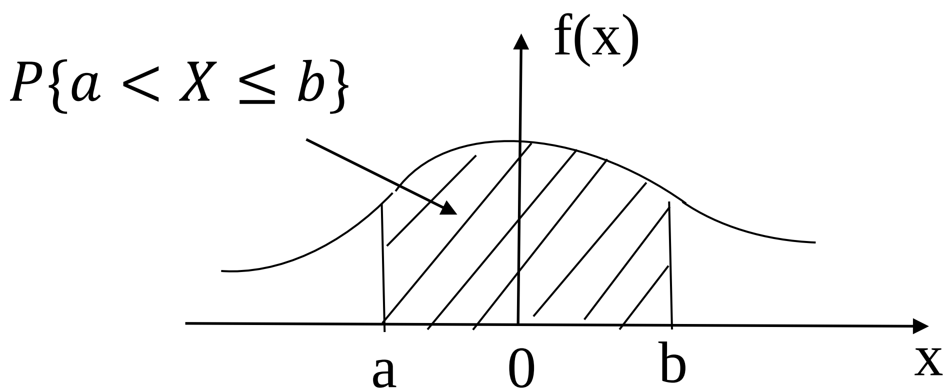
3.2 概率分布界定（续）

- 具有如下性质：

（1） $f(x) \geq 0$；

（2） $P\{a < X \leq b\} = \int_a^b f(x)dx$ ，特别地，$\int_{-\infty}^{+\infty} f(x)dx = 1$；

（3） $F'(x) = f(x)$。



3.2 概率分布界定（续）

- **连续随机变量的常用分布：**

（1）均匀分布：若随机变量X的概率密度函数为：

$$f(x)=\begin{cases}\dfrac{1}{b-a}, a\le x\le b\\ 0, 其它\end{cases};$$

（2）指数分布：若随机变量X的概率密度函数为：

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

（2）正态分布：若随机变量X的概率密度函数为：

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}，\ \sigma>0。$$

3.2 概率分布界定（续）

随机变量的累积分布函数（CDF）

累积分布函数的概念

- 若X为一个随机变量，对任意实数x，则随机变量的累积分布函数为：

$$\boldsymbol{F(x) = P\{X \leq x\}}$$

累积分布函数的性质

- 对任意实数x，$0 \leq F(x) \leq 1$；
- 若a<b，则 $F(\mathrm{a}) \leq F(\mathrm{b})$；
- $F(-\infty) = \lim\limits_{x \to -\infty} F(\mathrm{x}) = 0$，$F(+\infty) = \lim\limits_{x \to +\infty} F(\mathrm{x}) = 1$；
- $\lim\limits_{x_n \to x+} F(\mathrm{x_n}) = F(x)$。

3.2 概率分布界定（续）

- **离散型随机变量的累积分布函数**
  - 函数形式：

$$F(x)=\sum_{k:x_k\le x} p_k$$

- **连续型随机变量的累积分布函数**
  - 函数形式：

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

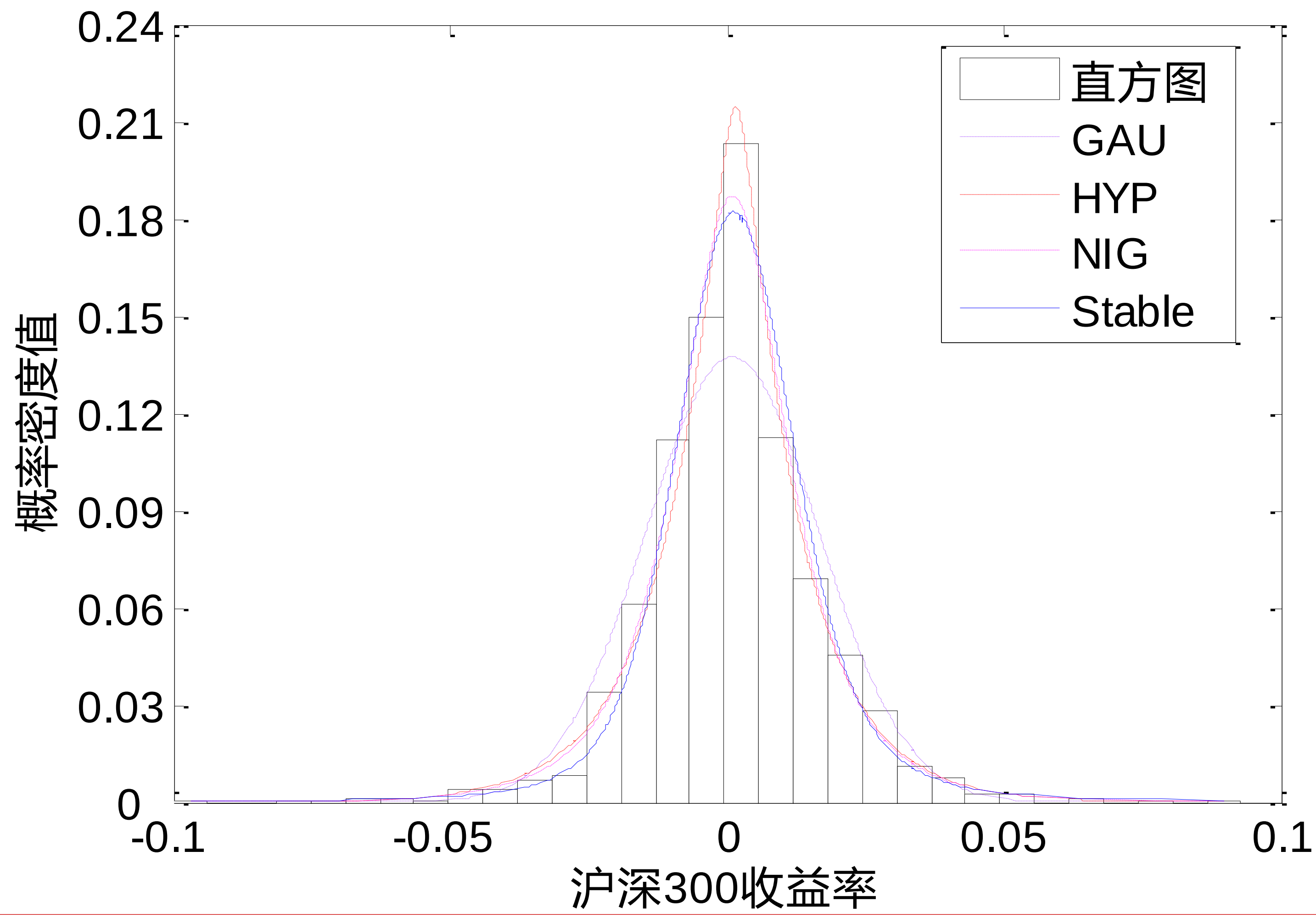
离散型	连续型
1. 分布列: $p_n = P(X=x_n)$	1. 密度函数 $X \sim f(x)$
2. $F(x) = \sum_{x_i \le x} P(X = x_i)$	2. $F(x) = \int_{-\infty}^{x} f(t)dt$

3. $F(a+0) = F(a)$; $P(a<X\le b) = F(b)-F(a)$.

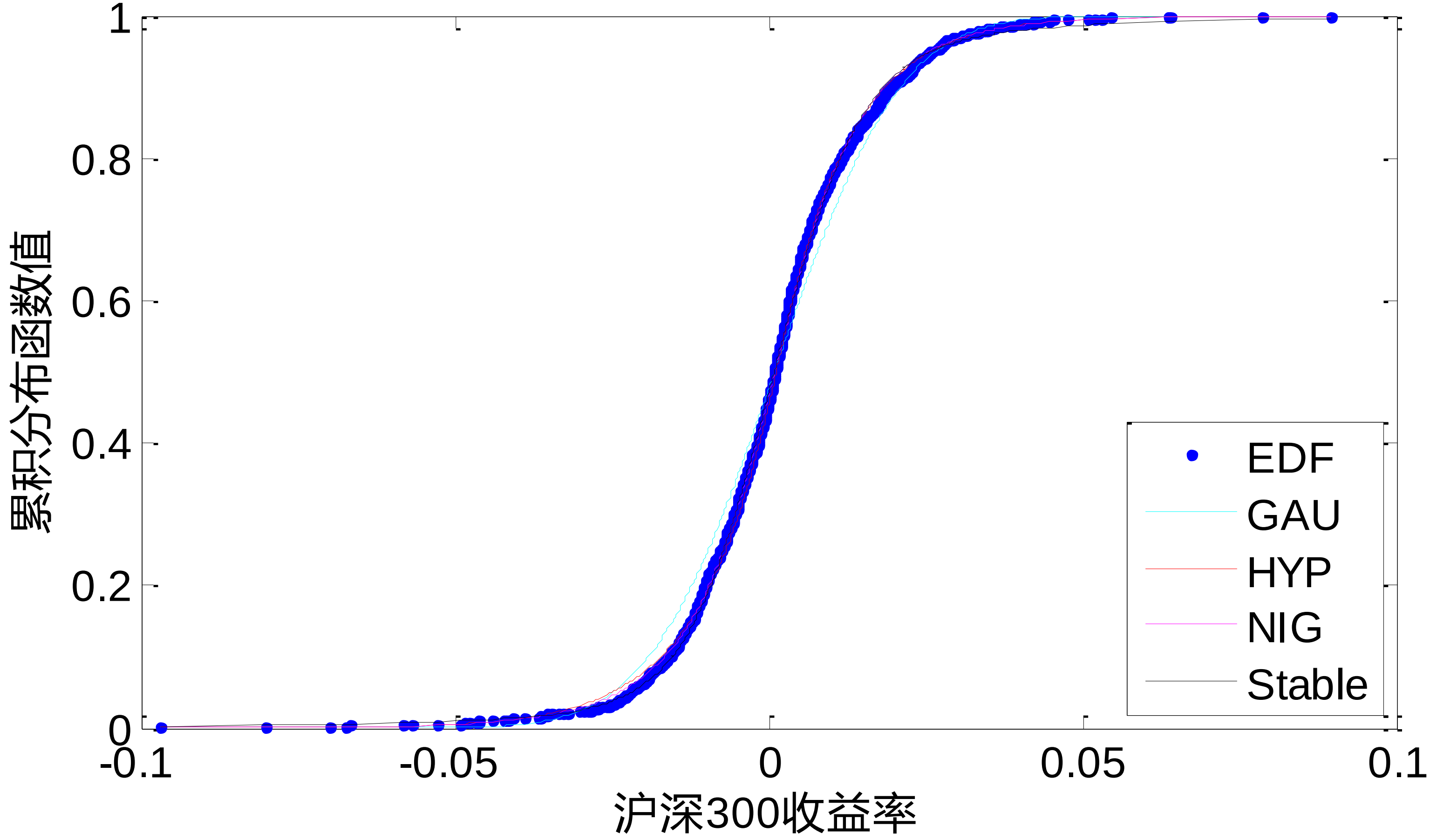
离散型	连续型
4. 点点计较	4. $P\{X=a\} = 0$
5. $F(x)$为阶梯函数。	5. $F(x)$为连续函数。
$F(a-0) \neq F(a)$.	$F(a-0) = F(a)$.

3.2 概率分布界定（续）

■ 应用实例：资产收益的概率分布



3.2 概率分布界定（续）



从图可看出，Stable和NIG分布对沪深300收益率序列拟合较好，比HYP和GAU均更好，GAU最差。

3.2 概率分布界定（续）

金融资产收益分布特征的一些共识

- **非正态性。**
- **资产收益分布的典型统计特征：**

  （1）非对称性；（2）尖峰；（3）厚尾
- **适用的分布类型：稳定分布、t分布**
- **资产收益分布实证研究的未来方向：**
  - 由股票市场向外汇市场、债券市场、基金市场等进行拓展，研究资产收益分布函数的具体形式及其主要特征；
  - 研究不同时间标度（周、月、季、年）的资产收益分布函数之间的关系以及资产收益序列各具体特征（如偏度、峰度等）之间的关系，并对此进行建模和解释；
  - 资产收益分布的应用性研究，包括对基于某种分布形式的资本市场理论以及风险管理研究；
  - 资产收益序列的各种行为规律（如周内效应、月份效应、小公司效应等）及其建模

3.3 正态分布

■ 普通正态分布

➢ 普通正态分布的定义

- 若连续型随机变量X的概率密度函数和累积分布函数分别为：

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

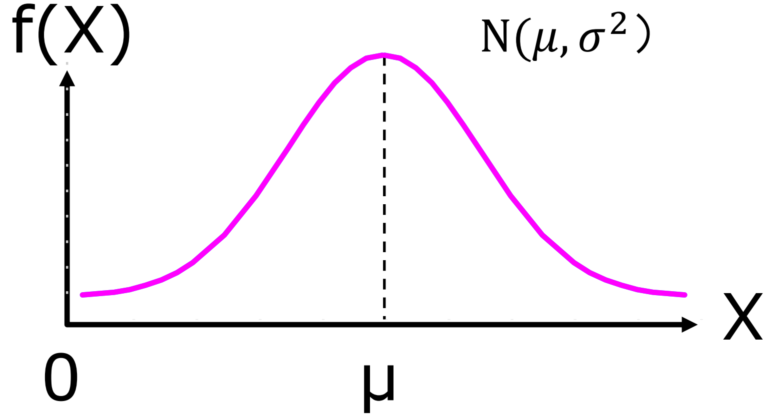
$$\mathrm{F}(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\mathrm{x}}e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt,$$

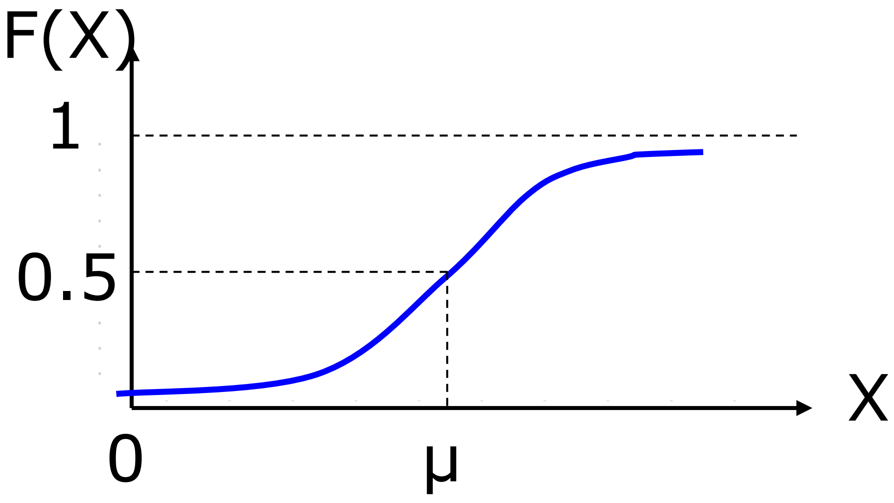
其中， μ、$\sigma>0$ 为常数，$-\infty<x<\infty$，则称随机变量X服从正态分布，记为

$$\mathrm{x}\sim\mathrm{N}(\mu,\sigma^2)。$$

3.3 正态分布（续）

- 正态分布 $N(\mu,\sigma^2)$ 的概率分布函数曲线和累积分布函数曲线分别为：





3.3 正态分布（续）

➢ 普通正态分布的性质

- 若 $X \sim N(\mu, \sigma^2)$，则随机变量X的性质有：

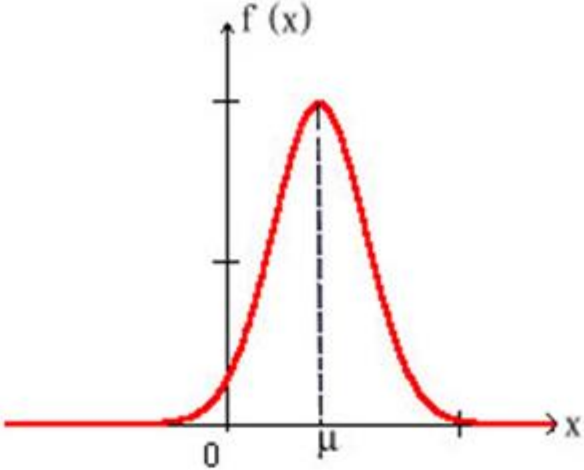
（1）X的均值为：$E(X) = \mu$

（2）X的方差为：$\sigma_X^2 = E[(X-\mu)^2] = \sigma^2$

- 概率密度函数值非负，即为 $f(x) \geq 0$；
- 概率密度曲线以x轴为渐近线，即为

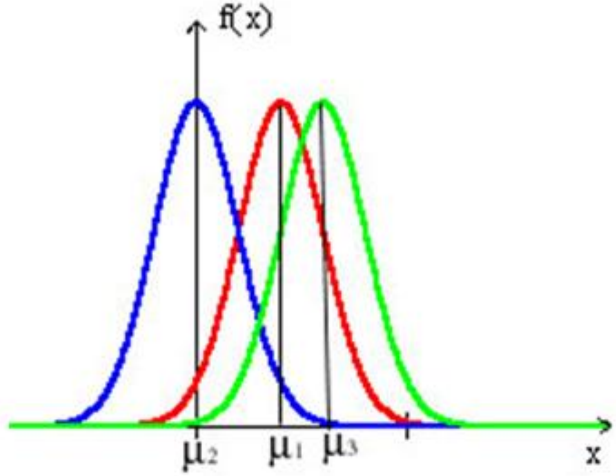
$$\text{当} x \to \infty \text{时，} f(x) \to 0$$

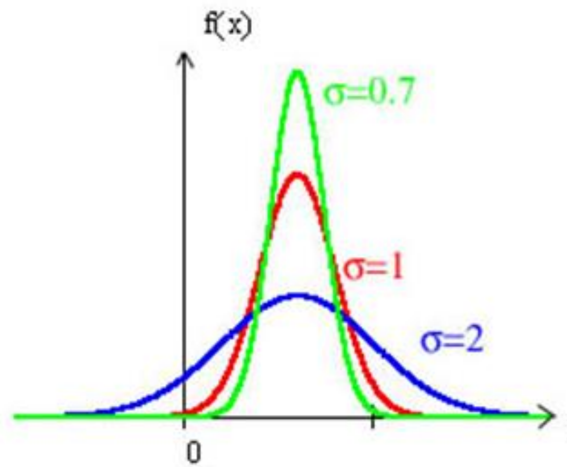
- 概率密度函数曲线是关于均值对称的钟形曲线；

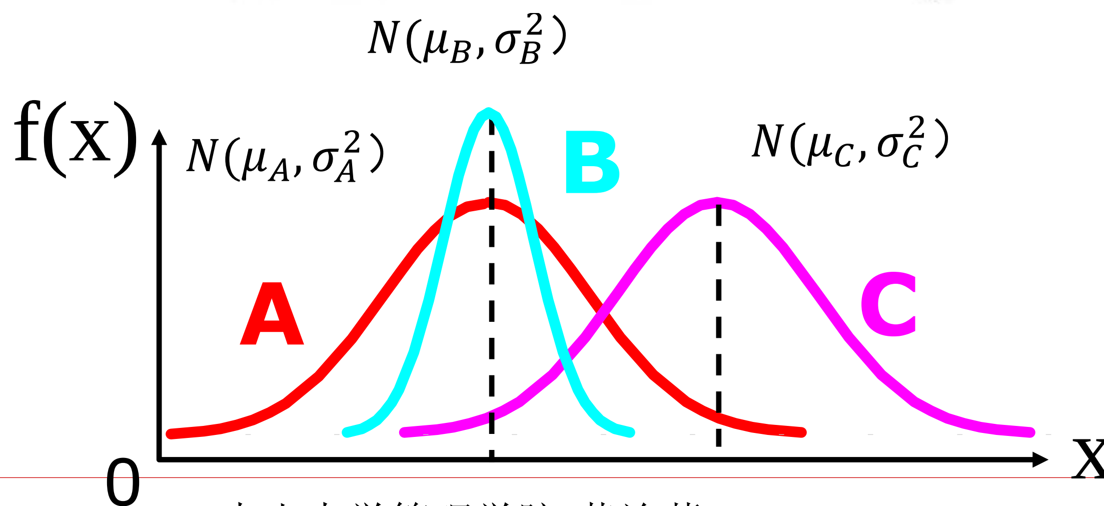


3.3 正态分布（续）

- 均值控制曲线的对称（中心）位置；方差控制曲线的陡峭程度。



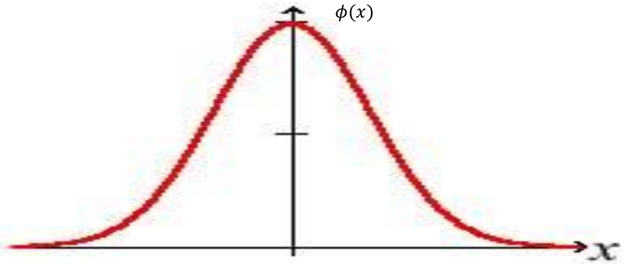




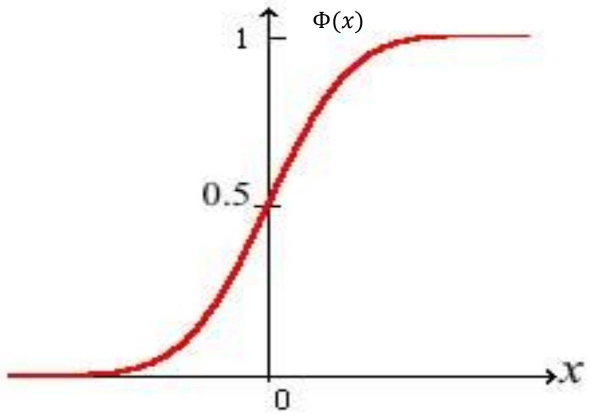
标准正态分布

- 若将普通正态分布中的均值取为0，方差取为1，则普通正态分布转化为标准正态分布，记为 Z~N(0,1)，其概率密度函数和累积分布函数分别为：

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



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



3.3 正态分布（续）

- 标准正态分布 N(0,1) 具有如下性质：

（1） $\phi(x)$ 具有各阶导数；

（2）$\phi(-x)=\phi(x)$ ；

（3）$\phi(x)$ 在 $(-\infty,0]$ 递增，而在 $(0,+\infty)$ 递减；

$$f'(x)=\frac{\mu-x}{\sqrt{2\pi\sigma^3}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\quad -\infty<x<+\infty$$

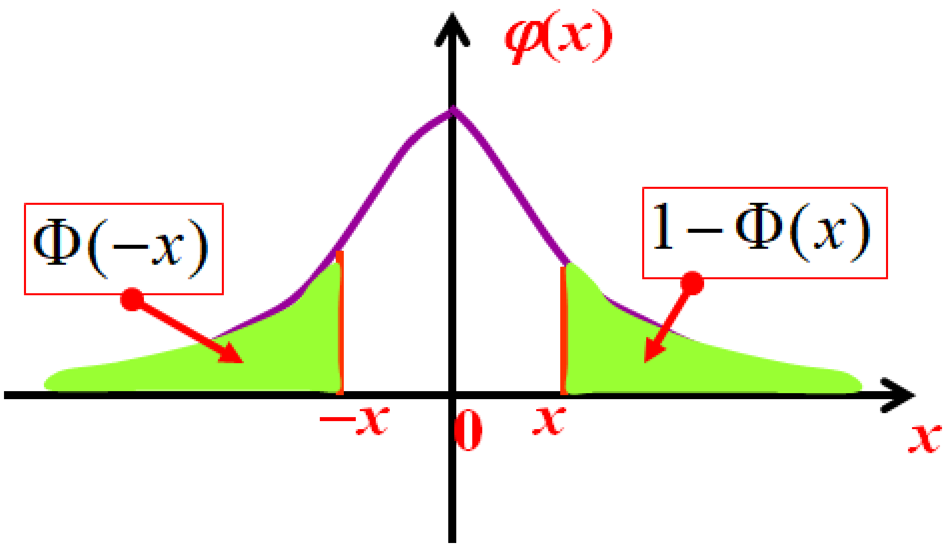
（4）$\phi(x)$ 在-1和1处有两个拐点；

$$f''(x)=\frac{(x-\mu)^2-\sigma^2}{\sqrt{2\pi\sigma^3}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\quad -\infty<x<+\infty$$

3.3 正态分布（续）

（5）以x轴为渐近线：$\lim_{x\to\infty}\phi(x)=0$；

（6）$\Phi(-x)=1-\Phi(x)$；



(1) $\Phi(0)=\frac{1}{2}$,

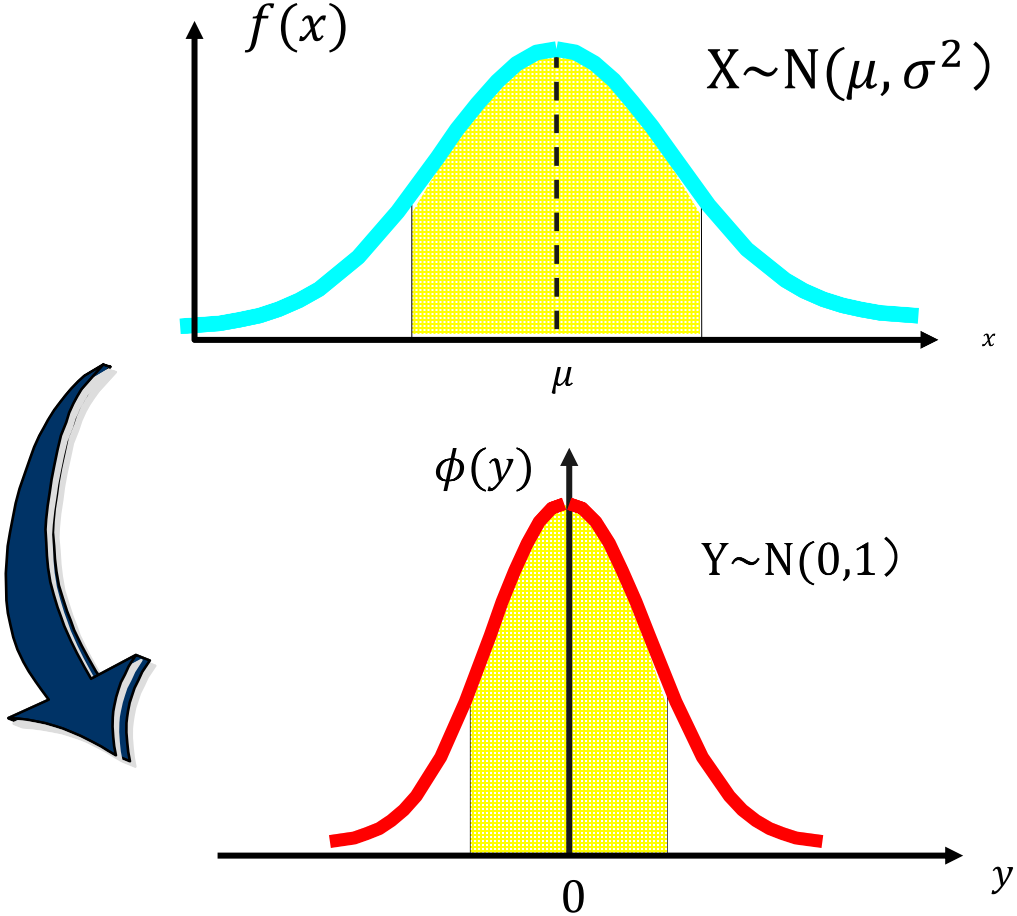
(2) $\Phi(-x)=1-\Phi(x)$

3.3 正态分布（续）

（7）标准正态分布 Y~N(0,1)与普通正态分布 X~N(μ,σ^2）的对应关系：

标准化：

$$Y = \frac{X-\mu}{\sigma}$$

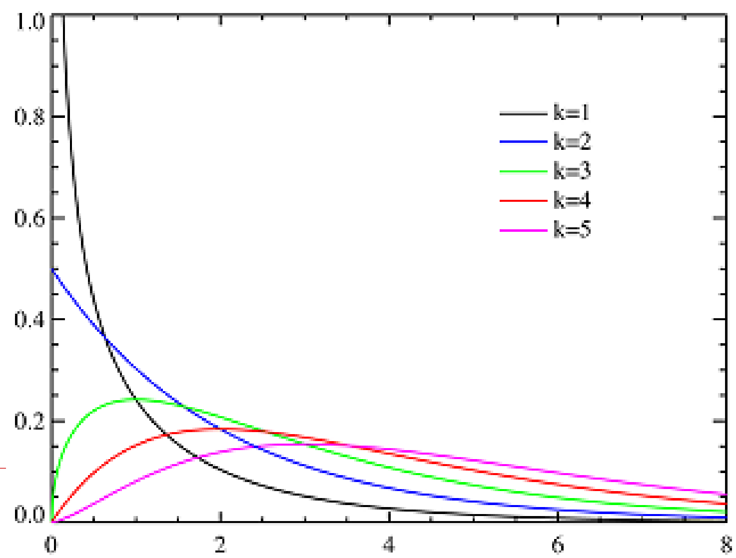


3.3 正态分布（续）

（7）若n个相互独立的随机变量X₁、X₂、…、Xn均服从标准正态分布（也称独立同分布于标准正态分布），则这n个服从标准正态分布的随机变量的平方和X构成一新的随机变量，其分布称为卡方分布（chi-square distribution），记为$V \sim \chi^2(n_2)$：

$$f(x|\nu) = \frac{x^{(\nu-2)/2} e^{-x/2}}{2^{\nu/2}\Gamma(\nu/2)}$$

其中，$x \geq 0$，$\nu > 0$为自由度参数，$\Gamma(\cdot)$为伽玛函数。



伽玛函数$\Gamma(x)$定义为：

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt, \quad x > 0$$

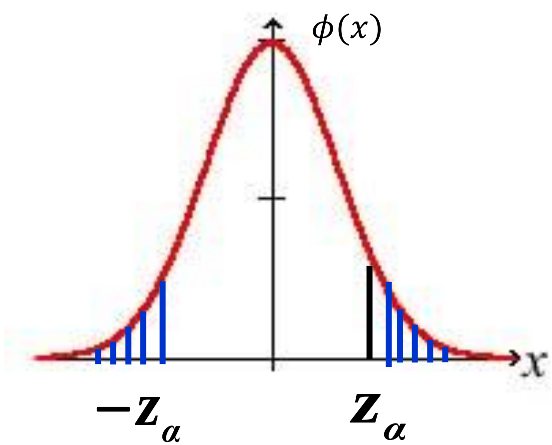
3.3 正态分布（续）

□ 标准正态分布的上α分位点

- 若X~N(0,1)，则满足条件

$$P\{X > z_\alpha\} = \alpha, 0 < \alpha < 1 \Rightarrow P\{X < -z_\alpha\} = \alpha$$

的点 z_α称为标准正态分布的上α分位点。

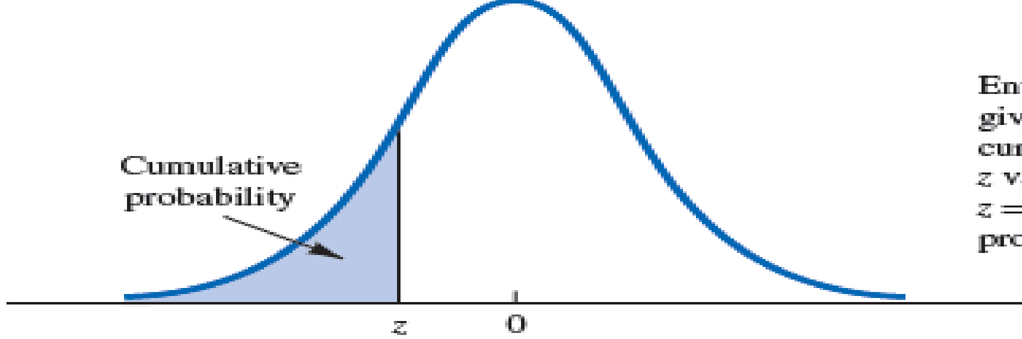


左图可以看出，$P\{X > z_{1-\alpha}\} = 1 - \alpha$

即 $P\{X < z_{1-\alpha}\} = \alpha$，因此，$z_{1-\alpha} = -z_\alpha$

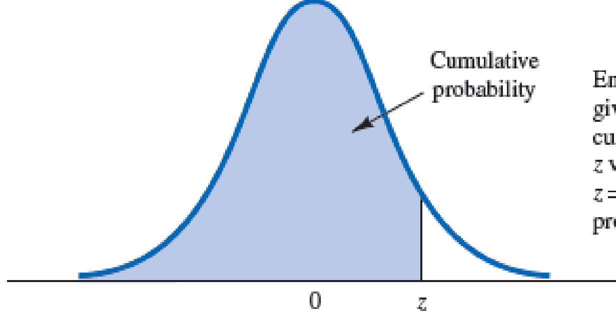
计算标准正态分布的右侧**2.5%**分位点的代码为：

- Excel：NORM.INV(1-0.025,0,1)
- MATLAB：norminv(1-0.025, 0, 1)
- Python：st.norm.ppf(1-0.025, 0, 1)



Entries in the table give the area under the curve to the left of the z value. For example, for $z = -.85$, the cumulative probability is .1977.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641



Entries in the table give the area under the curve to the left of the z value. For example, for $z = 1.25$, the cumulative probability is .8944.

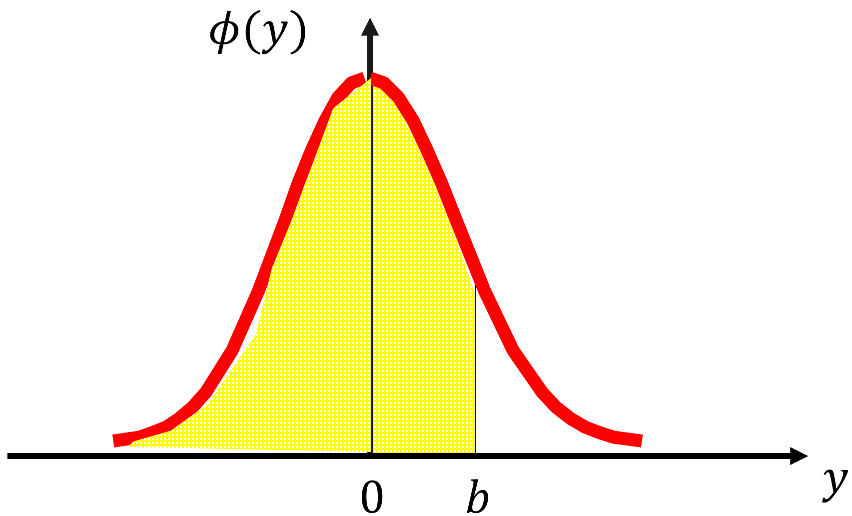
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

3.3 正态分布（续）

□ 标准正态分布概率的计算

➢ 问题1

- 若 $Y \sim N(0,1)$，则 $P\{Y \le b\} = ?$

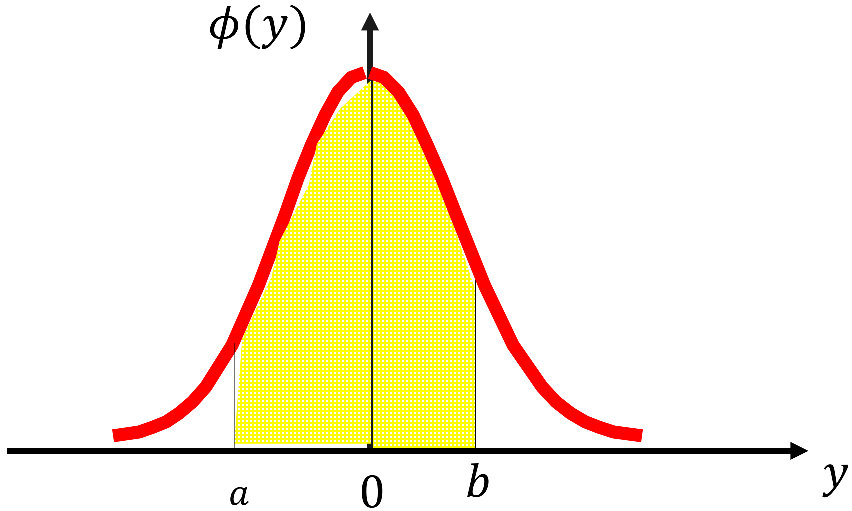


解决方法：直接查询标准正态分布表，或利用统计分析软件求得。

3.3 正态分布（续）

➢ 问题2

- 若 $Y \sim N(0,1)$，则 $P\{a < Y \le b\} = ?$

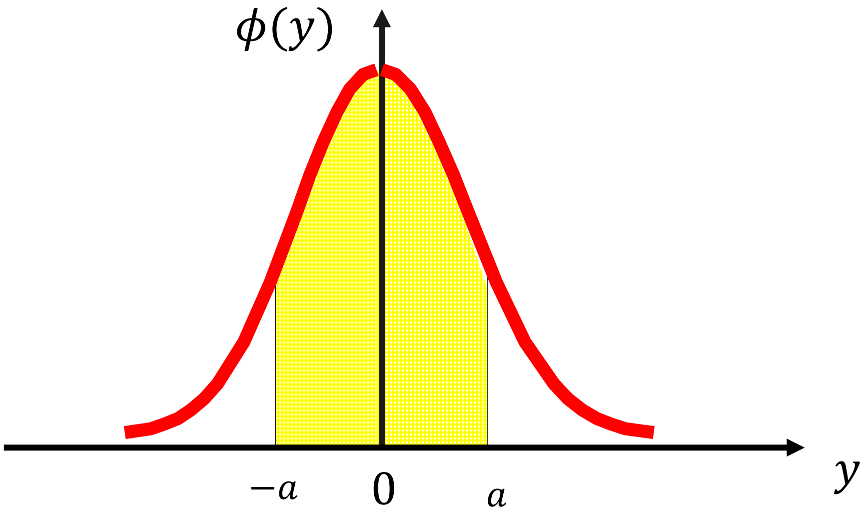


解决方法：先变换 $P\{a < Y \le b\} = P\{Y \le b\} - P\{Y < a\}$，然后查询标准正态分布表，或利用统计分析软件求得。

3.3 正态分布（续）

➢ 问题3

- 若 $Y \sim N(0,1)$，则 $P\{|Y| \le a\} = ?$



解决方法：先变换 $P\{|Y| \le a\} = 2(0.5 - P\{Y \le -a\})$，然后查询标准正态分布表，或利用统计分析软件求得。

3.3 正态分布（续）

- **常用概率值（3σ原则）**：若 Y~N(0,1)，则

$$P\{|Y| \le 1.000\} =$$

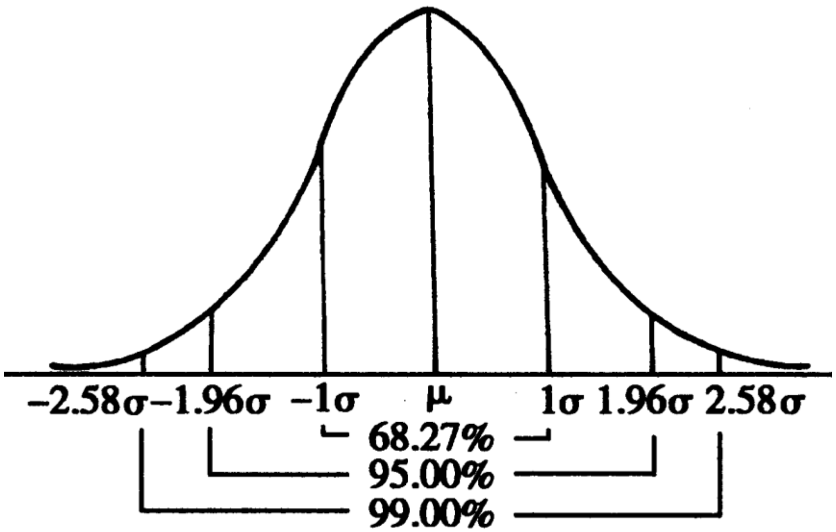
$$P\{|Y| \le 1.645\} =$$

$$P\{|Y| \le 1.960\} =$$

$$P\{|Y| \le 2.000\} =$$

$$P\{|Y| \le 2.576\} =$$

$$P\{|Y| \le 3.000\} =$$



3.3 正态分布（续）

普通正态分布概率的计算

- 问题
  - 若 $X \sim N(\mu,\sigma^2)$，则 $P\{a < X \le b\} = ?$
  - 若 $X \sim N(\mu,\sigma^2)$ 且 $P\{\mu < X \le a\} = p$，则 $a = ?$

- 解决方法
  - **方法1**：直接利用概率密度函数的积分即可求得；
  - **方法2**：首先，正态标准化，然后，查询标准正态分布函数表即可求得。
  - **方法3**：利用统计分析软件求得。

3.3 正态分布（续）

计算练习：

- **例3.3.1：**
  - 设 $X \sim N(0, 1)$，求 $P(X>-1.96)$，$P(|X|<1.96)$。
- **例3.3.2：**
  - 设 $X \sim N(0, 1)$，$P(X\le b) = 0.9515$，$P(X\le a) = 0.04947$，求 a、b。
- **例3.3.3：**
  - 设随机变量$X \sim N(0,1)$，试求：(1). $P\{1\le X<2\}$；(2). $P\{-1<X<2\}$.
- **例3.3.4：**
  - 设随机变量$X \sim N(2,9)$，试求：(1). $P\{1\le X<5\}$；(2). $P\{|X-2|>6\}$；(3). $P\{X>0\}$

3.3 正态分布（续）

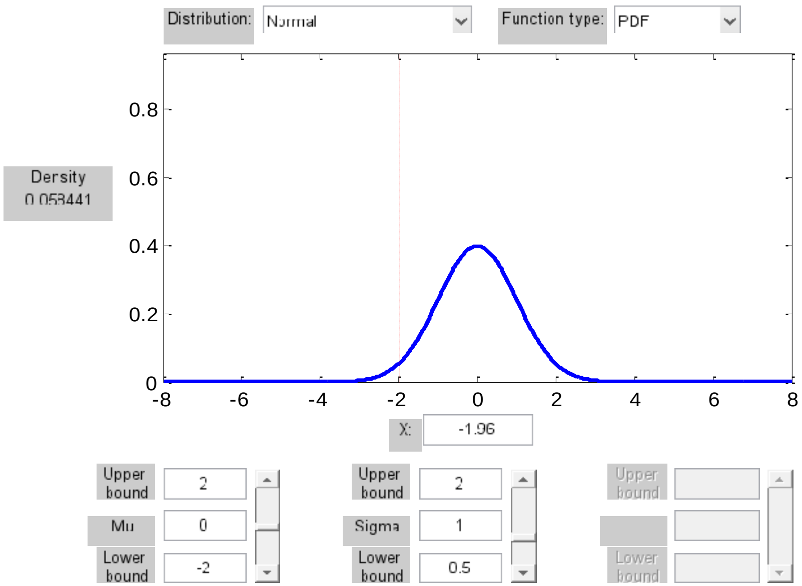
- **例3.3.5：**
  - 设 $X \sim N(10, 4)$，求$P(10<X<13)$，$P(|X-10|<2)$。
- **例3.3.6：**
  - 设 $X \sim N(\mu, \sigma^2)$， $P(X \leq -5) = 0.045$，$P(X \leq 3) = 0.618$, 求 μ 及 σ。
- **例3.3.7：**
  - 公司人事主管部门对2 500名经理的年薪进行了调查，发现所有经理的平均年薪及其标准差分别为51 800元和730.30元，则某位经理的平均年薪在51 300至52 300元之间的概率是多少？

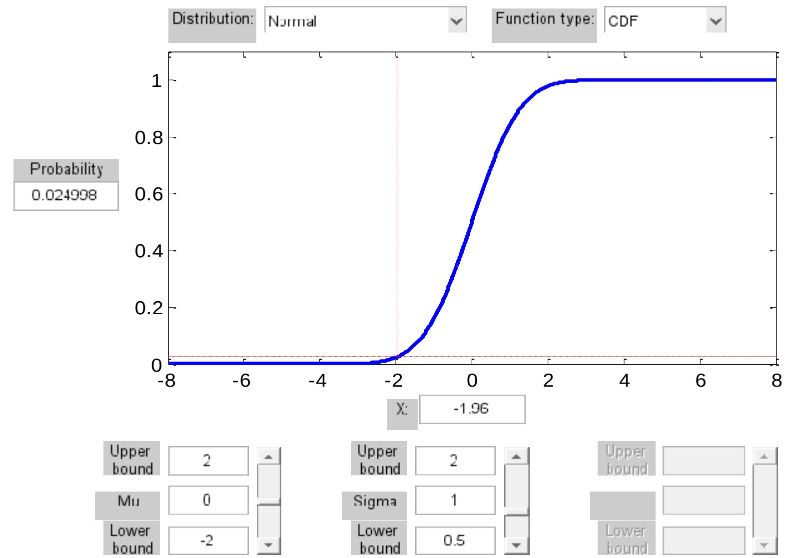
3.3 正态分布（续）

- 例3.3.8：
  - 某类资产的平均价值等于500 000元，标准差为15 000元，则属于该类的某项资产价值在485 000至530 000元之间的概率为多少？
- 例3.3.9：项目设备生产是否采用外包决策
  - 若某项目建设正需一批设备，而且项目组经过分析发现，若自行生产，则机器设备年租金为100 000元，若采用外包生产，则单位外包成本为700元。但是，单位原材料成本（LC）、单位工资成本（MC）以及自行生产的年产量（P）等3个变量是不确定的，通过经验发现，它们分别服从如下的正态分布：$LC \sim N(200,100)$，$MC \sim N(400,25)$，$P \sim N(900,10000)$。

那么，项目经理要求你利用蒙特卡罗模拟方法提供定量决策依据。

MATLAB概率分布函数动态演示界面：disttool





f = normpdf(X,mu,sigma); %计算均值为 mu、标准差为 sigma 的正态分布概率密度函数在 x 处的取值。
F = normcdf(X,mu,sigma); %计算均值为 mu、标准差为 sigma 的正态分布累积分布函数在 x 处的取值。
u = norminv(P, mu,sigma); %计算均值为 mu、标准差为 sigma 的正态分布累积分布逆函数在 P 处的取值。

f = normpdf(-2:2,1,10); F = normcdf(-2:2,1,10); u = norminv(1-0.05, 0, 1); %分别计算均值为 1、标准差为 10 的正态分布概率分布函数值、累积分布函数值、标准正态分布的上侧 0.05 分位数。

例3.3.1： 设 $X \sim N(0, 1)$，求

$$P(X>-1.96),\quad P(|X|<1.96)$$

解: $P(X>-1.96) = 1-\Phi(-1.96) = 1-(1-\Phi(1.96))$

$= \Phi(1.96) = 0.975$ **(查表得)**

$$P(|X|<1.96) = 2\Phi(1.96)-1 = 2\times 0.975-1 = 0.95$$

Excel代码:

```
p1=1-NORMSDIST(-1.96)
p2=NORMSDIST(1.96)-NORMSDIST(-1.96)
```

MATLAB代码:

```
p1=1-cdf('Normal',-1.96,0,1);  %或者p1=1-normcdf(-1.96,0,1);
p2=cdf('Normal',1.96,0,1)-cdf('Normal',-1.96,0,1); %或者
 p2=normcdf(1.96,0,1)-normcdf(-1.96,0,1);
```

Python代码:

```
import scipy.stats as st
p1=1-st.norm.cdf(x=-1.96, loc=0,scale=1) #st.norm.sf(x=-1.96, loc=0,scale=1)
p2=st.norm.cdf(x=1.96, loc=0, scale=1)-st.norm.cdf(x=-1.96, loc=0, scale=1)
```

例3.3.2： 设 $X \sim N(0, 1)$, $P(X \le b) = 0.9515$,

$P(X \le a) = 0.04947$, 求 a, b.

解: $\Phi(b) = 0.9515 > 1/2$,

所以 $b > 0$, 反查表得:

$\Phi(1.66) = 0.9515$,

故 $b = 1.66$

$\Phi(a) = 0.0495 < 1/2$,

所以 $a < 0$, $\Phi(-a) = 0.9505$, 反查表得：$\Phi(1.65) = 0.9505$,

故 $a = -1.65$

Excel代码:

```
a=NORM.INV(0.9515,0,1)
b=NORM.INV(0.04947,0,1)
```

MATLAB代码:

```
a = norminv(0.9515, 0, 1);
b = norminv(0.04947, 0, 1);
```

Python代码:

```
import scipy.stats as st
a = st.norm.ppf(q=0.9515, loc=0,scale=1)
b = st.norm.ppf(q=0.04947, loc=0,scale=1)
```

例3.3.3: 设随机变量$X \sim N(0,1)$，试求：

(1). $P\{1 \le X < 2\}$；(2). $P\{-1 < X < 2\}$.

解：(1). $\boldsymbol{P\{1 \le X < 2\} = \Phi(2) - \Phi(1) = 0.97725 - 0.84134 = 0.1359}$

(2). $\boldsymbol{P\{-1 \le X < 2\} = \Phi(2) - \Phi(-1) = \Phi(2) - [1 - \Phi(1)]}$

$\boldsymbol{= 0.97725 - 1 + 0.84134 = 0.81859}$

Excel代码:

```
p1=NORMSDIST(2)-NORMSDIST(1)
p2= =NORMSDIST(2)-NORMSDIST(-1)
```

MATLAB代码:

```
p1=normcdf(2,0,1)- normcdf(1,0,1)
p2=normcdf(2,0,1)- normcdf(-1,0,1)
```

Python代码:

```
import scipy.stats as st
p1=st.norm.cdf(x=2, loc=0,scale=1)- st.norm.cdf(x=1, loc=0,scale=1)
p2= st.norm.cdf(x=2, loc=0,scale=1)- st.norm.cdf(x=-1,
loc=0,scale=1)
```

例3.3.4： 设随机变量$X \sim N(2,\quad 9)$，试求：
(1). $P\{1 \le X < 5\}$；(2). $P\{|X-2| > 6\}$；(3). $P\{X > 0\}$.

解：(1). $P\{1 \le X < 5\} = F(5) - F(1) = \Phi(\frac{5-2}{3}) - \Phi(\frac{1-2}{3})$

$$= \Phi(1) - \Phi\left(-\frac{1}{3}\right) = \Phi(1) + \Phi\left(\frac{1}{3}\right) - 1$$

$$= 0.84134 + 0.62930 - 1 = 0.47064$$

(2). $P\{|X-2| > 6\} = 1 - P\{|X-2| \le 6\} = 1 - P\{-6 \le X - 2 \le 6\}$

$$= 1 - P\{-4 \le X \le 8\} = 1 - [\Phi(\frac{8-2}{3}) - \Phi(\frac{-4-2}{3})]$$

$$= 1 - [\Phi(2) - \Phi(-2)] = 2 \times [1 - \Phi(2)]$$

$$= 2 \times (1 - 0.97725) = 0.0455$$

(3). $P\{X > 0\} = 1 - P\{X \le 0\} = 1 - \Phi(\frac{0-2}{3}) = 1 - \Phi\left(-\frac{2}{3}\right)$

$$= \Phi\left(\frac{2}{3}\right) = 0.7486$$

Excel代码：

```
p1= NORMDIST(5,2,3,TRUE)-NORMDIST(1,2,3,TRUE)
p2= NORMDIST(-6,0,3,TRUE)-1+NORMDIST(6,0,3,TRUE)
p3= 1-NORMDIST(0,2,3,TRUE)
```

MATLAB代码：

```
p1=cdf('Normal',5,2,3)-cdf('Normal',1,2,3) %或者p1=normcdf(5,2,3)-normcdf(1,2,3)
p2=1-cdf('Normal',8,2,3)+cdf('Normal',-4,2,3) %或者p2=1-normcdf(8,2,3)+normcdf(-4,2,3);
p3=1-cdf('Normal',0,2,3) %或者p2=1-normcdf(0,2,3)
```

Python代码：

```
import scipy.stats as st
p1=st.norm.cdf(x=5, loc=2,scale=3)-st.norm.cdf(x=1, loc=2,scale=3)
p2=st.norm.cdf(x=-6, loc=0,scale=3)+1-st.norm.cdf(x=6, loc=0,scale=3)
p3=1-st.norm.cdf(x=0, loc=0,scale=3)
```

例3.3.5： 设 $X \sim N(10, 4)$，求 $P(10<X<13)$，$P(|X-10|<2)$.

解: $P(10<X<13) = \Phi(1.5)-\Phi(0) = 0.9332 - 0.5 = 0.4332$

$P(|X-10|<2) = P(8<X<12) = 2\Phi(1)-1 = 0.6826$

Excel代码:

```
p1= NORMDIST(13,10,2,TRUE)-NORMDIST(10,10,2,TRUE)
p2= NORMDIST(2,0,2,TRUE)- NORMDIST(-2,0,2,TRUE)
```

MATLAB代码:

```
p1=cdf('Normal',13,10,2)-cdf('Normal',10,10,2) %或者
 p1=normcdf(13,10,2)-normcdf(10,10,2)
p2=1-cdf('Normal',2,0,2)+cdf('Normal',-2,0,2) %或者p2=1-
 normcdf(2,0,2)+normcdf(-2,0,2);
```

Python代码:

```
import scipy.stats as st
p1=st.norm.cdf(x=13, loc=10,scale=2)-st.norm.cdf(x=10, loc=10,scale=2)
p2=st.norm.cdf(x=2, loc=0,scale=2) -st.norm.cdf(x=-2, loc=0,scale=2)
```

例3.3.6： 设 $X \sim N(\mu, \sigma^2)$, $P(X \le -5) = 0.045$, $P(X \le 3) = 0.618$, 求 μ 及 σ.

解:
$$\begin{cases} \dfrac{5+\mu}{\sigma} = 1.69 \\ \dfrac{3-\mu}{\sigma} = 0.3 \end{cases} \Longrightarrow \begin{cases} \mu = 1.76 \\ \sigma = 4 \end{cases}$$

Excel代码:

```
a=NORM.INV(0.045,0,1)
b=NORM.INV(0.618,0,1)
```

MATLAB代码:

```
a = norminv(0.045, 0, 1)
b = norminv(0.618, 0, 1)
```

Python代码:

```
import scipy.stats as st
a = st.norm.ppf(q=0.045, loc=0,scale=1)
b = st.norm.ppf(q=0.618, loc=0,scale=1)
```

例3.3.7：

$$P\{51300 < x \le 52300\} =$$

$$P\{\frac{51300-51800}{730.30} < \frac{x-\mu}{\sigma_x} \le \frac{52300-51800}{730.30}\} = P\{-0.6847 < z \le 0.6847\}$$

$$= 2\times(P\{z \le 0.6847\} - 0.5) = 2\times(0.7533-0.5) = 0.5066$$

Excel代码：

p= NORMDIST(52300, 51800, 730.3,TRUE)-NORMDIST(51300, 51800, 730.3,TRUE)

或CDF.NORMAL(52300,51800,730,30)-CDF.NORMAL(51300,51800,730,30)

MATLAB代码：

```
x_down=51300;  x_up=52300;  x_bar=51800;  standard_deviation=730.3;
p=cdf('Normal',x_up,x_bar,standard_deviation)-cdf('Normal', x_down, x_bar, standard_deviation);
```

Python代码：

```
import scipy.stats as st
st.norm.cdf(x=52300, loc=51800,scale=730.3)-st.norm.cdf(x=51300, loc=51800,scale=730.3)
```

例3.3.8：

$$P\{485000 < x \le 530000\} =$$

$$P\{\frac{485000-500000}{15000} < \frac{x-\mu}{\sigma_x} \le \frac{530000-500000}{15000}\} = P\{-1 < z \le 2\}$$

$$= P\{z \le 2\} - (1 - P\{z \le 1\}) = 0.97725 - (1 - 0.8413) = 0.8186$$

Excel代码：

p= NORMDIST(530000, 500000, 15000,TRUE)-NORMDIST(485000, 500000, 15000, TRUE)

或CDF.NORMAL(530000, 500000, 15000)- CDF.NORMAL(485000, 500000, 15000)

MATLAB代码：

```
x_down= 485000;  x_up= 530000;  x_bar= 500000;  standard_deviation= 15000;
p=cdf('Normal',x_up,x_bar,standard_deviation)-cdf('Normal’, x_down, x_bar, standard_deviation);
```

Python代码：

```
import scipy.stats as st
st.norm.cdf(x= 530000, loc= 500000, scale= 15000)-st.norm.cdf(x= 485000, loc= 500000, scale= 15000)
```

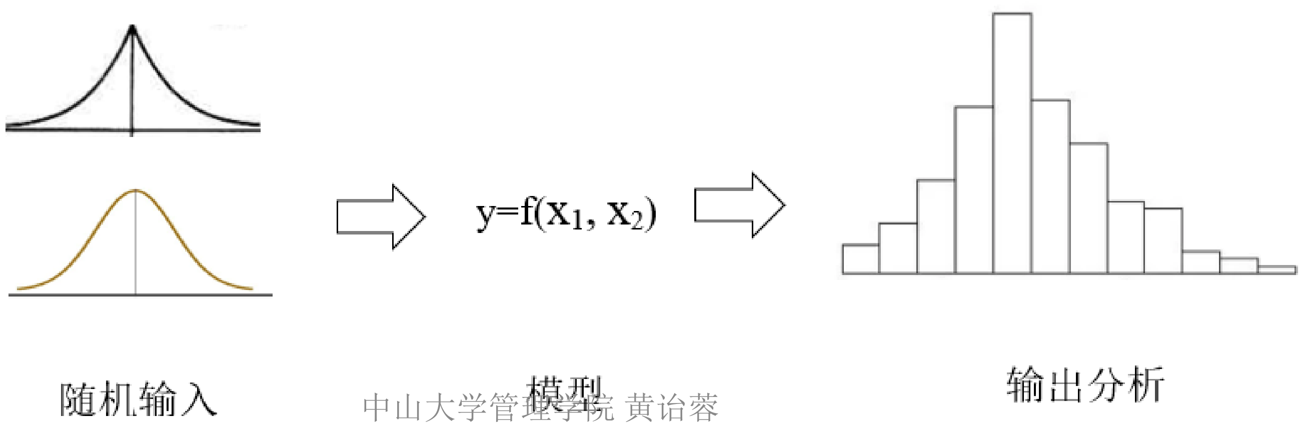
蒙特卡罗模拟法（MC，Monte Carlo Simulation）

基本原理

- 在假定变量服从某概率分布函数条件下，利用计算机和抽样技术为该变量模拟一系列随机数，然后计算出该变量小于等于某给定值的概率，或直接计算目标变量的结果。

基本步骤

- 构建各变量的概率分布；
- 生成大量随机数，一般在1000个以上；
- 计算目标变量的大量取值；
- 输出分析（比如，计算最小值、最大值、均值、标准差、概率等）。



蒙特卡罗模拟法（续）

- （3）利用各风险变量产生的所有随机数计算状态变量的大量取值，并进行统计分析（如计算最小值、最大值、均值和标准差等）；
- （4）根据分析结果，绘制概率分布和累积概率分布曲线，比如，基于正态分布的累积概率曲线；
- （5）依据累积概率曲线评估其风险状况；
- （6）也可利用模拟数据和实际数据对概率模型进行修正，并在实践中不断完善。

常用于模拟的概率分布模型

- 均匀分布、正态分布、对数正态分布、指数分布、极值分布、三角分布、β分布、泊松分布等。

例3.3.9：

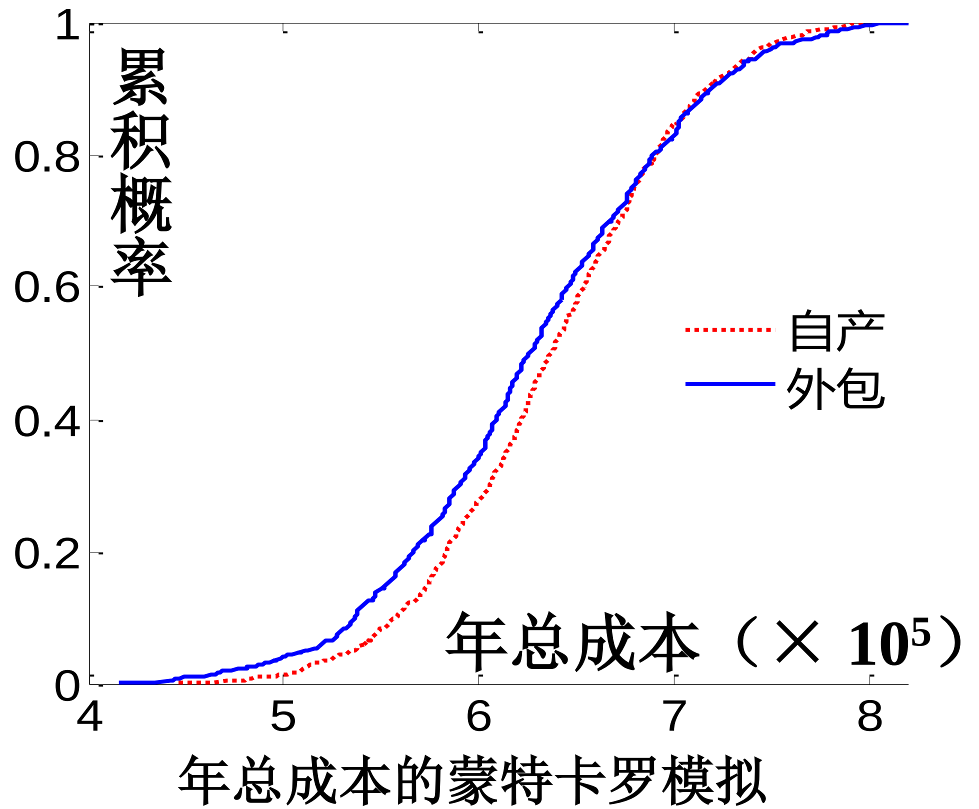
第1步：首先，模拟年产量P、单位原材料成本MC和单位工资成本LC的1000个随机数，然后，根据成本公式计算年可变成本和总成本，如下表A、B、C、D、E例所示（仅显示前5个和最后5个）。

自行生产年总成本的模拟结果：

	A	B	C	D	E	F
1	年产量 P	单位原材料成本 LC	单位工资成本 MC	年可变成本 VC=P*(LC+MC)	自行生产：总成本 TC=VC+100000	小于等于E列显示数值的概率 P
2	1127.1227	214.4094	403.851	696854.7682	796854.7682	1.000
3	1132.4832	206.9711	405.186	693257.8113	793257.8113	0.999
4	1159.4934	198.9097	398.326	692491.0958	792491.0958	0.998
5	1112.5198	222.4671	399.058	691458.4307	791458.4307	0.997
6	1160.1664	199.2945	395.191	689701.8688	789701.8688	0.996
997	636.29019	184.294	402.971	373670.916	473670.916	0.005
998	630.97874	192.2753	395.94	371151.1511	471151.1511	0.004
999	619.71556	195.6067	392.29	364328.4268	464328.4268	0.003
1000	604.95992	198.4714	401.942	363225.7972	463225.7972	0.002
1001	593.78405	184.5891	394.994	344147.1487	444147.1487	0.001

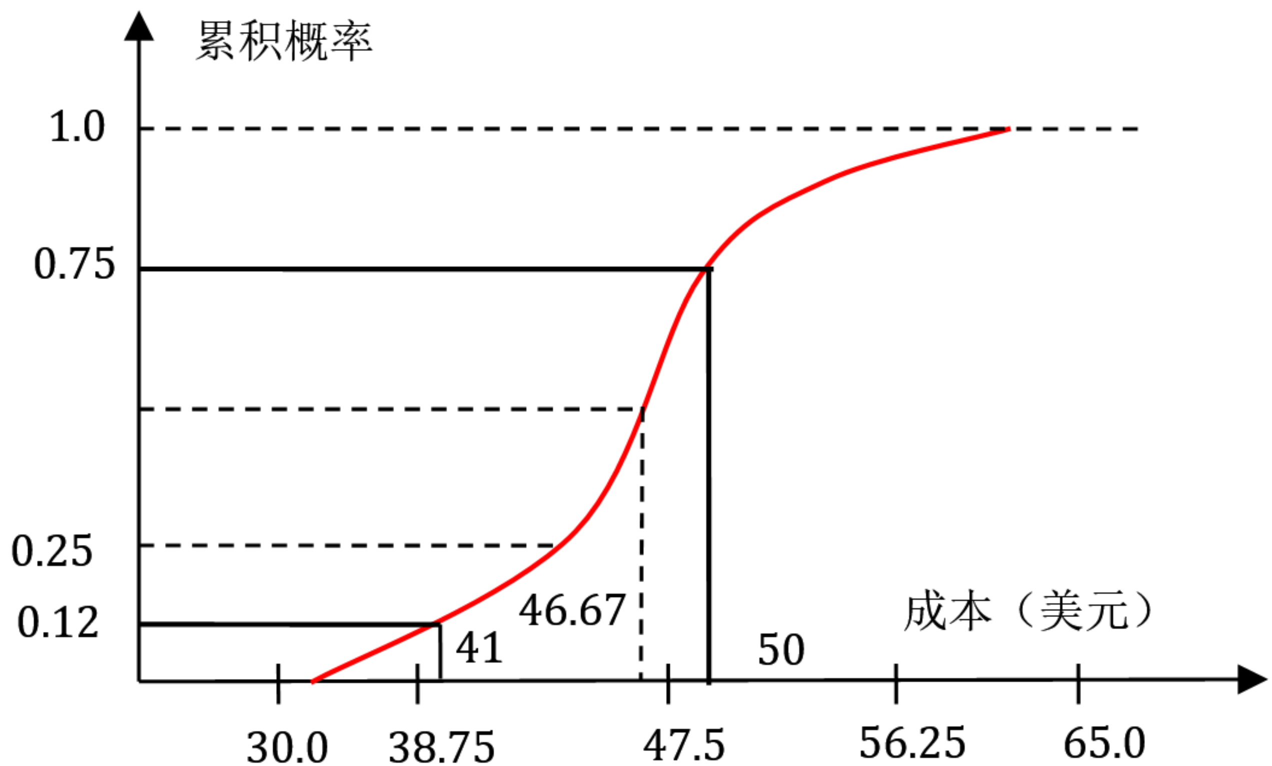
外包年总成本的模拟结果：

	A	B	C
1	年产量 P	外包：总成本 TC=P*700	小于等于E列显示数值的概率 P
2	1180.28	826199.11	1.0000
3	1160.17	812116.49	0.9990
4	1159.49	811645.37	0.9980
5	1149.91	804935.12	0.9970
6	1145.47	801830.83	0.9960
997	633.46	443424.44	0.0050
998	630.98	441685.12	0.0040
999	619.72	433800.89	0.0030
1000	604.96	423471.94	0.0020
1001	593.78	415648.83	0.0010

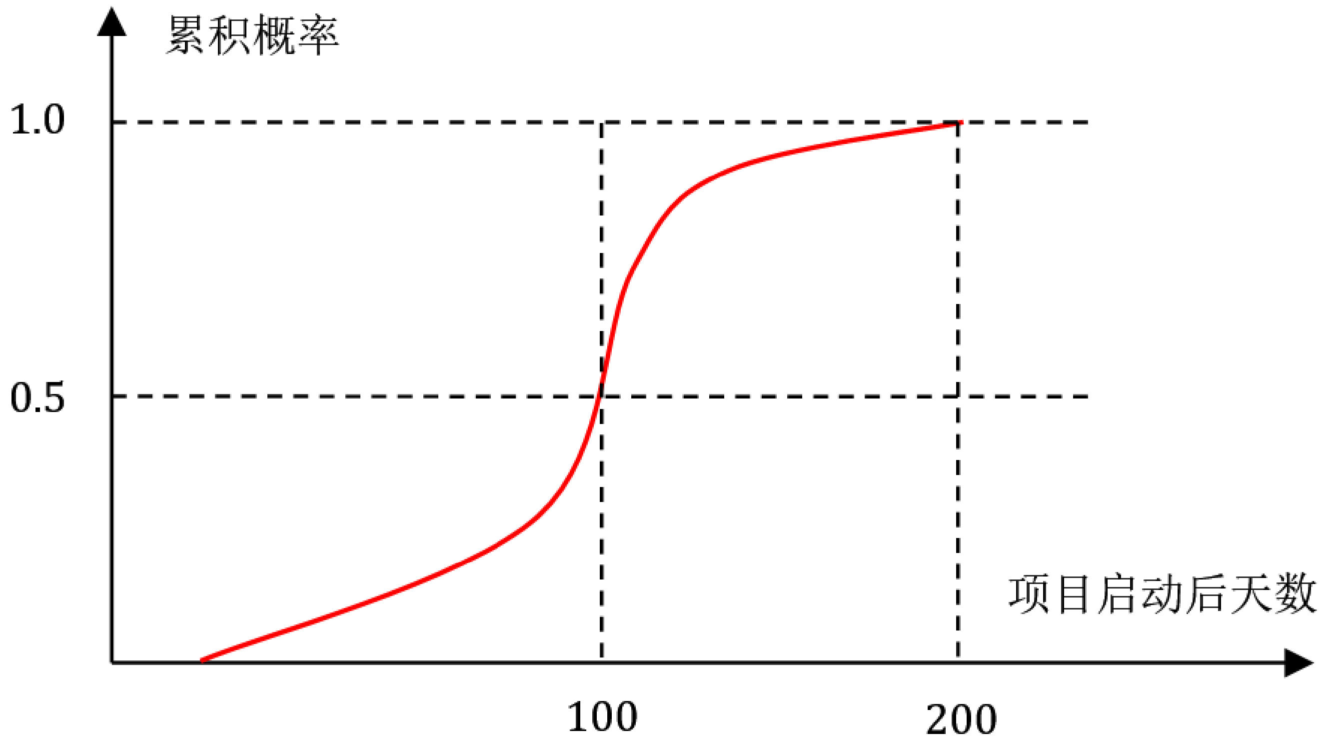


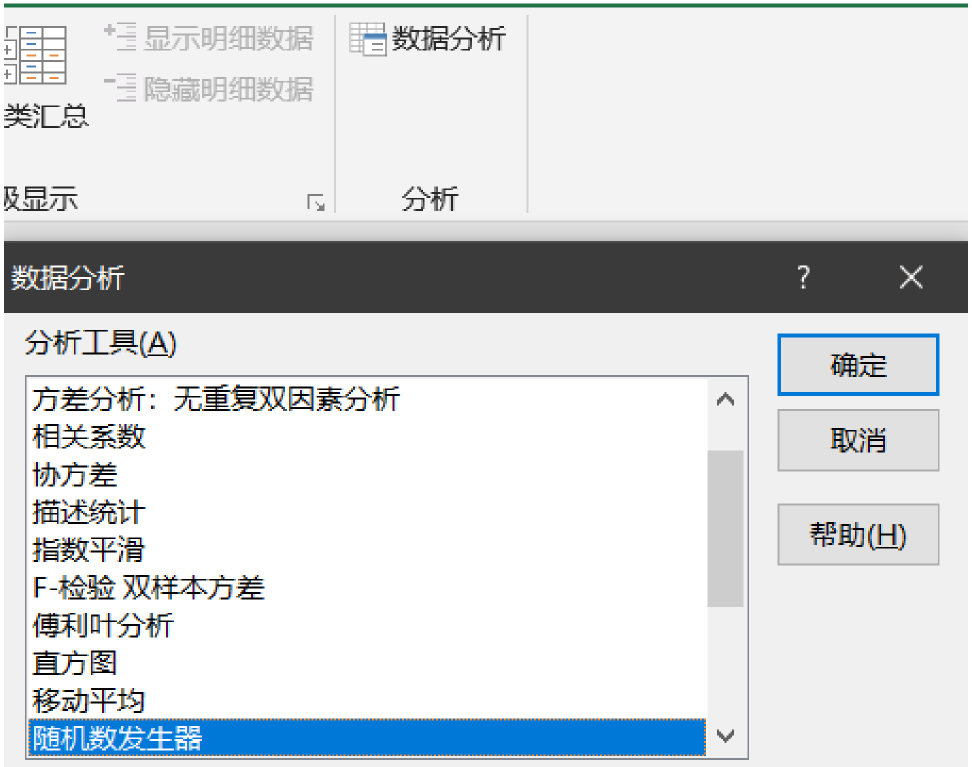
第2步：计算所有年总成本值以下对应的累积概率曲线，如上图所示。从分析结果可看出，应选择外包方式成本更低。

➢ 项目成本蒙特卡罗模拟



➢ 项目进度蒙特卡罗模拟




显示明细数据
隐藏明细数据
类汇总
级显示
数据分析
分析
数据分析
分析工具(A)
方差分析：无重复双因素分析
相关系数
协方差
描述统计
指数平滑
F-检验 双样本方差
傅利叶分析
直方图
移动平均
随机数发生器
确定
取消
帮助(H)


随机数发生器
变量个数(V):
1
随机数个数(B):
1000
分布(D):
正态
参数
平均值(E) =
900
标准偏差(S) =
100
随机数基数(R):
输出选项
输出区域(O):
G1
新工作表组(P):
新工作薄(W)
确定
取消
帮助(H)

MATLAB代码：

```
%% 模拟LC、MC和P的正态分布随机数（1000行1列）
LC = normrnd(200,10, 1000, 1);
MC = normrnd(400,5, 1000, 1);
P = normrnd(900,100, 1000, 1);
%% 计算自行生产的总成本和外包生产的总成本
TC1=P.*(LC+MC)+100000; %自行生产总成本
TC2=P*700; %外包总成本
%% 计算累积概率值
sort1=sort(TC1); sort2=sort(TC2); Prob=0.001:0.001:1;
%% 绘制总成本与累积概率关系曲线
figure('Color','W');
plot(sort1, Prob, sort2, Prob, 'r-.', 'Linewidth',2);
xlabel('总成本'); ylabel('累积概率');
legend('自行生产','外包生产','Location', 'SouthEast');
```

Python代码：

```
#--- 模拟LC、MC和P的正态分布随机数（1000行1列）
import numpy.random as nprS
LC = npr.normal(loc=200,scale=10,size=1000)
MC = npr.normal(loc=400,scale=5,size=1000)
P = npr.normal(loc=900,scale=100,size=1000)
#---计算自行生产的总成本和外包生产的总成本
TC1=P*(LC+MC)+100000  #自行生产总成本
TC2=P*700;  #外包总成本
#--- 计算累积概率值
import numpy as npS
sort1=np.sort(TC1); sort2=np.sort(TC2); Prob=np.linspace(0.001,1,1000);
#--- 绘制总成本与累积概率关系曲线
import matplotlib; matplotlib.__version__
#matplotlib.use('Qt5Agg') #matplotlib.use('?') #matplotlib.get_backend() #如果显示不出
  图片，则运行该语句
import matplotlib.pyplot as plt   #导入pyplot模块
plt.rcParams['font.sans-serif']='Simhei' #以黑体字体正常显示中文
plt.rcParams['axes.unicode_minus']=False #用于正常显示坐标轴，
fig, ax = plt.subplots()  ##显性创建figure对象(创建图 figure 和 坐标系 axes)，然后在
  坐标系中添加图形和相关元素,也可独立创建：#fig=plt.figure()创建画布;
  ax=plt.axes()创建坐标轴
ax.plot(sort1,Prob, label='自行生产' ) ; ax.plot(sort2,Prob, color='r', label='外包') ;
  ax.legend() ; plt.show() ; plt.xlabel('总成本'); plt.ylabel('累积概率')
```

3.4 T分布

T 分布界定

T分布定义

设$X \sim N(0,1)$，$Y \sim \chi^2(n)$，且X与Y相互独立，则称变量：

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

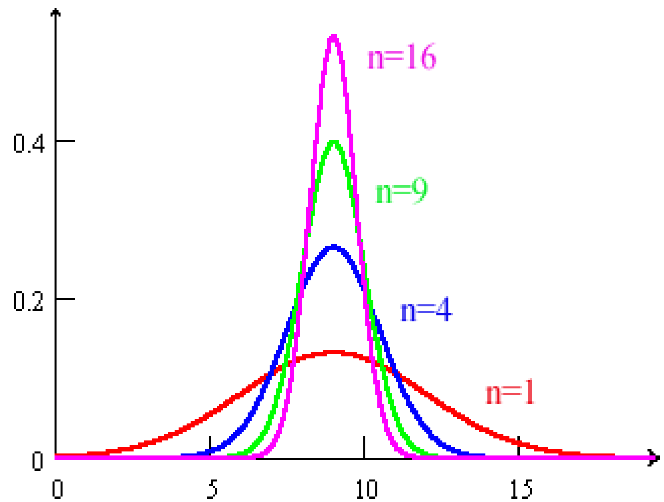
所服从的分布为自由度为 n的 t 分布，又称学生氏分布，记为t~t(n)。

3.4 T分布（续）

- T分布的概率密度函数

$$h(t)=\frac{\Gamma[(n+1)/2]}{\Gamma(n/2)\sqrt{n\pi}}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}\quad -\infty<t<\infty$$

- T分布的概率密度曲线

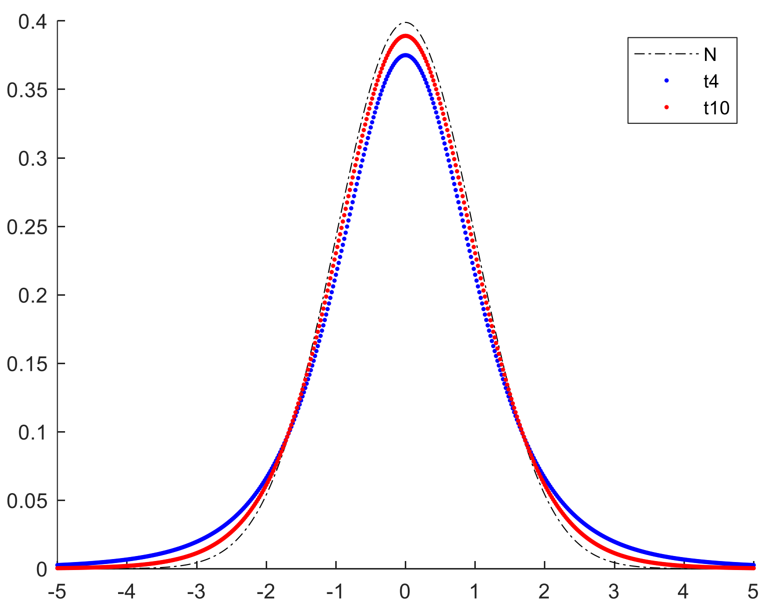


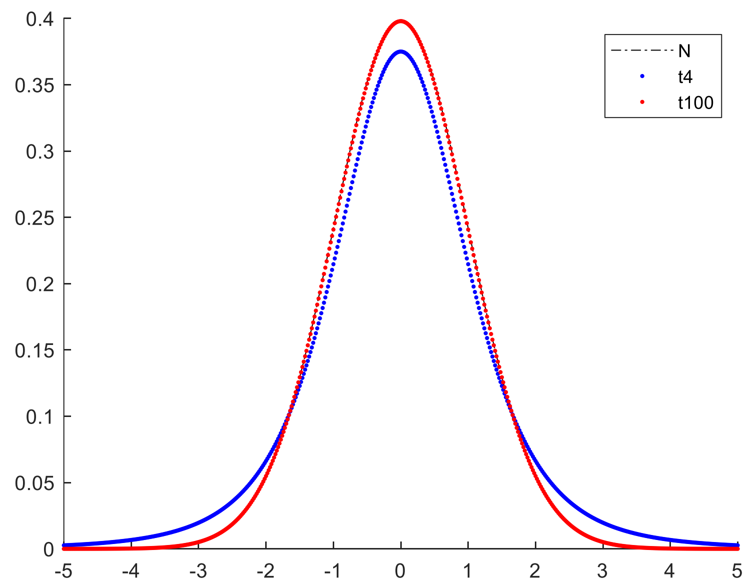
3.4 T分布（续）

➢ T分布的性质

- 具有自由度为n的t分布$t \sim t\,(n)$，其数学期望与方差为：$E(t)=0, D(t)=n/(n-2)\;(n>2)$；
- T分布的密度函数关于t=0对称；
- 当n充分大时，T分布趋近于正态分布，即为

，或者，$\lim\limits_{n\to\infty} h(t)=\dfrac{1}{\sqrt{2\pi}}e^{-t^2/2}$　　$t \overset{近似}{\sim} N(0,1)$





3.4 T分布（续）

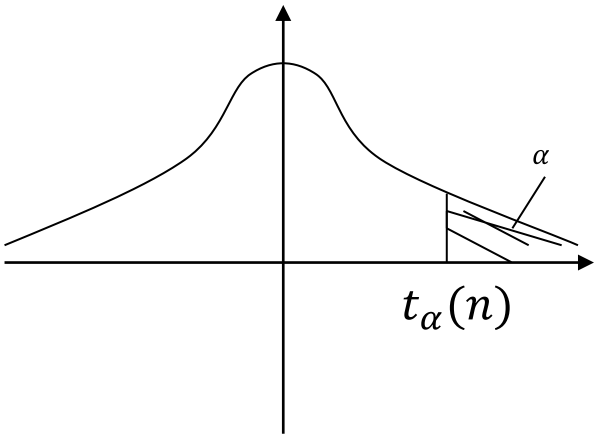
T分布的分位点

- T分布的上α分位点定义：
  - 对于给定α，$0<\alpha<1$，称满足条件：

$$p\{t > t_\alpha(n)\} = \int_{t_\alpha(n)}^{\infty} h(t)dt = \alpha$$

的点$t_\alpha(n)$。

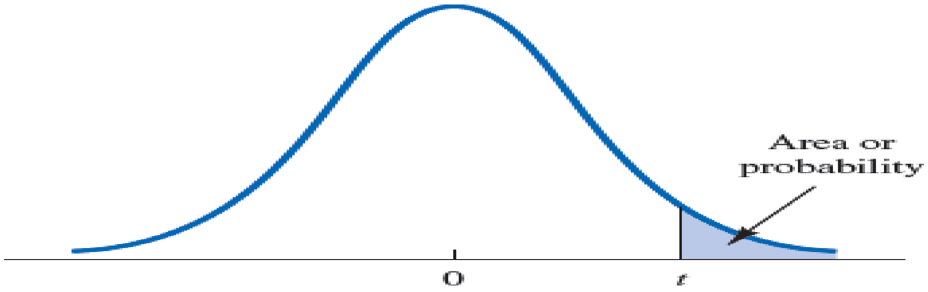
- T分布的上α分位点图示



3.4 T分布（续）

- T分布的上α分位点性质
  - $t_{1-\alpha}(n) = -t_\alpha(n)$
- 计算自由度为15的T分布上侧0.025分位点
  - 查表：$t_{0.025}(15) = 2.131$
  - 近似：当n>45时，对于常用的α的值，可用正态近似，即为$t_\alpha(n) \approx z_\alpha$
  - Excel代码： T.INV(1-0.025,15)
  - MATLAB代码：tinv(1-0.025, 15)
  - Python代码：

```
import scipy.stats as st
st.t.ppf(1-0.025, 15)
```



Entries in the table give t values for an area or probability in the upper tail of the t distribution. For example, with 10 degrees of freedom and a .05 area in the upper tail, $t_{.05} = 1.812$.

Degrees of Freedom	Area in Upper Tail					
	.20	.10	.05	.025	.01	.005
1	1.376	3.078	6.314	12.706	31.821	63.656
2	1.061	1.886	2.920	4.303	6.965	9.925
3	.978	1.638	2.353	3.182	4.541	5.841
4	.941	1.533	2.132	2.776	3.747	4.604
5	.920	1.476	2.015	2.571	3.365	4.032
6	.906	1.440	1.943	2.447	3.143	3.707
7	.896	1.415	1.895	2.365	2.998	3.499
8	.889	1.397	1.860	2.306	2.896	3.355
9	.883	1.383	1.833	2.262	2.821	3.250
10	.879	1.372	1.812	2.228	2.764	3.169
11	.876	1.363	1.796	2.201	2.718	3.106
12	.873	1.356	1.782	2.179	2.681	3.055
13	.870	1.350	1.771	2.160	2.650	3.012
14	.868	1.345	1.761	2.145	2.624	2.977
15	.866	1.341	1.753	2.131	2.602	2.947
16	.865	1.337	1.746	2.120	2.583	2.921
17	.863	1.333	1.740	2.110	2.567	2.898
18	.862	1.330	1.734	2.101	2.552	2.878
19	.861	1.328	1.729	2.093	2.539	2.861
20	.860	1.325	1.725	2.086	2.528	2.845
21	.859	1.323	1.721	2.080	2.518	2.831
22	.858	1.321	1.717	2.074	2.508	2.819
23	.858	1.319	1.714	2.069	2.500	2.807
24	.857	1.318	1.711	2.064	2.492	2.797
25	.856	1.316	1.708	2.060	2.485	2.787
26	.856	1.315	1.706	2.056	2.479	2.779
27	.855	1.314	1.703	2.052	2.473	2.771
28	.855	1.313	1.701	2.048	2.467	2.763
29	.854	1.311	1.699	2.045	2.462	2.756
30	.854	1.310	1.697	2.042	2.457	2.750
31	.853	1.309	1.696	2.040	2.453	2.744
32	.853	1.309	1.694	2.037	2.449	2.738
33	.853	1.308	1.692	2.035	2.445	2.733
34	.852	1.307	1.691	2.032	2.441	2.728

3.5 F分布

F分布界定

F分布的定义

- 设$U \sim \chi^2(n_1), V \sim \chi^2(n_2)$，U与V相互独立，则称随机变量

$$\boldsymbol{F = \frac{U/n_1}{V/n_2}}$$

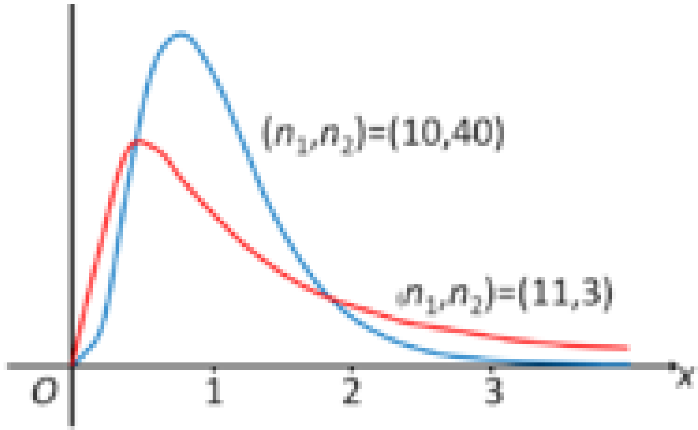
服从自由度为n_1及 n_2的F分布，n_1称为第一（分子）自由度，n_2称为第二（分母）自由度，记作$F \sim F(n_1,n_2)$，易知，$\frac{1}{F} = \frac{V/n_2}{U/n_1} \sim F(n_2,n_1)$。

3.5 F分布（续）

- F分布的概率密度函数
  - 若$F \sim F(n_1,n_2)$， F的概率密度函数为

$$\varphi(y)=\begin{cases}\dfrac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{2})\,\Gamma(\frac{n_2}{2})}\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}}(y)^{\frac{n_1}{2}-1}\left(1+\frac{n_1}{n_2}y\right)^{-\frac{n_1+n_2}{2}} & y>0\\ 0 & y\le 0\end{cases}$$

- F分布的概率密度曲线



3.5 F分布（续）

➢ F分布的性质

- F分布的数学期望为：$E(F) = \frac{n_2}{n_2-2}$，n_2>2；
- 它的数学期望并不依赖于第一自由度n_1。

3.5 F分布（续）

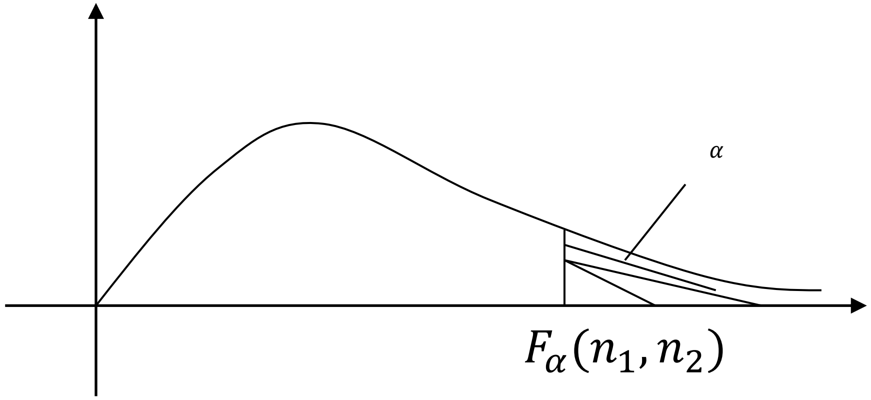
F分布的分位点

- F(n_1, n_2)分布的上α分位点定义
  - 对于给定的α，0<α<1，称满足条件

$$p\{F > F_\alpha(n_1, n_2)\} = \int_{F_\alpha(n_1,n_2)}^{\infty} \phi(y)dy = \alpha$$

的点$F_\alpha(n_1, n_2)$。

- F(n_1, n_2)分布的上α分位点图示



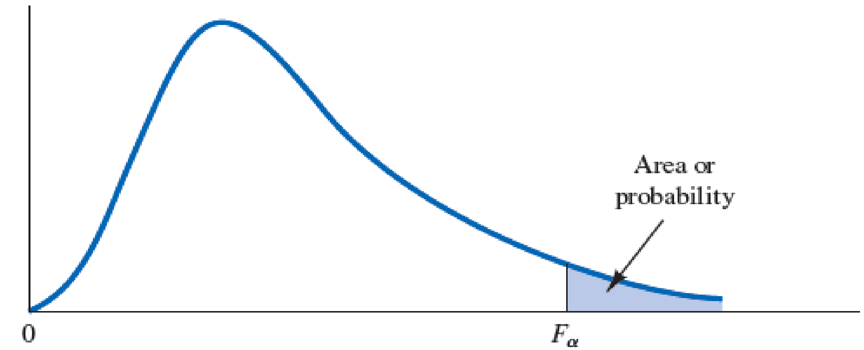
3.5 F分布（续）

- F(n_1, n_2)分布的上α分位点性质

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

- 计算分子和分母自由度分别为4和8的F分布上侧0.025分位点
  - 查表：$F_{0.025}(4,8) = 5.053 = \frac{1}{F_{0.975}(8,4)} = \frac{1}{0.1979}$
  - Excel代码：F.INV(1-0.025,4,8)
  - MATLAB代码：finv(1-0.025, 4, 8)
  - Python代码：

```
import scipy.stats as st
st.f.ppf(1-0.025,4,8)
```



Entries in the table give F_α values, where α is the area or probability in the upper tail of the F distribution. For example, with 4 numerator degrees of freedom, 8 denominator degrees of freedom, and a .05 area in the upper tail, $F_{.05} = 3.84$.

Denominator Degrees of Freedom	Area in Upper Tail	Numerator Degrees of Freedom																	
		1	2	3	4	5	6	7	8	9	10	15	20	25	30	40	60	100	1000
1	.10	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	61.22	61.74	62.05	62.26	62.53	62.79	63.01	63.30
	.05	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	245.95	248.02	249.26	250.10	251.14	252.20	253.04	254.19
	.025	647.79	799.48	864.15	899.60	921.83	937.11	948.20	956.64	963.28	968.63	984.87	993.08	998.09	1001.40	1005.60	1009.79	1013.16	1017.76
	.01	4052.18	4999.34	5403.53	5624.26	5763.96	5858.95	5928.33	5980.95	6022.40	6055.93	6156.97	6208.66	6239.86	6260.35	6286.43	6312.97	6333.92	6362.80
2	.10	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
	.05	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.43	19.45	19.46	19.46	19.47	19.48	19.49	19.49
	.025	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50
	.01	98.50	99.00	99.16	99.25	99.30	99.33	99.36	99.38	99.39	99.40	99.43	99.45	99.46	99.47	99.48	99.48	99.49	99.50
3	.10	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.20	5.18	5.17	5.17	5.16	5.15	5.14	5.13
	.05	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.70	8.66	8.63	8.62	8.59	8.57	8.55	8.53
	.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.25	14.17	14.12	14.08	14.04	13.99	13.96	13.91
	.01	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.34	27.23	26.87	26.69	26.58	26.50	26.41	26.32	26.24	26.14
4	.10	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
	.05	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
	.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.66	8.56	8.50	8.46	8.41	8.36	8.32	8.26
	.01	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.20	14.02	13.91	13.84	13.75	13.65	13.58	13.47
5	.10	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.324	3.21	3.19	3.17	3.16	3.14	3.13	3.11
	.05	6.61.	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.62	4.56	4.52	4.50	4.46	4.43	4.41	4.37
	.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.43	6.33	6.27	6.23	6.18	6.12	6.08	6.02
	.01	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.72	9.55	9.45	9.38	9.29	9.20	9.13	9.03

Denominator Degrees of Freedom	Area in Upper Tail	Numerator Degrees of Freedom																	
		1	2	3	4	5	6	7	8	9	10	15	20	25	30	40	60	100	1000
6	.10	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.87	2.84	2.81	2.80	2.78	2.76	2.75	2.72
	.05	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.94	3.87	3.83	3.81	3.77	3.74	3.71	3.67
	.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.27	5.17	5.11	5.07	5.01	4.96	4.92	4.86
	.01	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.56	7.40	7.30	7.23	7.14	7.06	6.99	6.89
7	.10	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.63	2.59	2.57	2.56	2.54	2.51	2.50	2.47
	.05	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.51	3.44	3.40	3.38	3.34	3.30	3.27	3.23
	.025	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.57	4.47	4.40	4.36	4.31	4.25	4.21	4.15
	.01	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.31	6.16	6.06	5.99	5.91	5.82	5.75	5.66
8	.10	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.30
	.05	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.22	3.15	3.11	3.08	3.04	3.01	2.97	2.93
	.025	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.10	4.00	3.94	3.89	3.84	3.78	3.74	3.68
	.01	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.52	5.36	5.26	5.20	5.12	5.03	4.96	4.87
9	.10	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.34	2.30	2.27	2.25	2.23	2.21	2.19	2.16
	.05	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.01	2.94	2.89	2.86	2.83	2.79	2.76	2.71
	.025	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.77	3.67	3.60	3.56	3.51	3.45	3.40	3.34
	.01	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	4.96	4.81	4.71	4.65	4.57	4.48	4.41	4.32
10	.10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.24	2.20	2.17	2.16	2.13	2.11	2.09	2.06
	.05	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.85	2.77	2.73	2.70	2.66	2.62	2.59	2.54
	.025	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.52	3.42	3.35	3.31	3.26	3.20	3.15	3.09
	.01	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.56	4.41	4.31	4.25	4.17	4.08	4.01	3.92
11	.10	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.17	2.12	2.10	2.08	2.05	2.03	2.01	1.98
	.05	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.72	2.65	2.60	2.57	2.53	2.49	2.46	2.41
	.025	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.33	3.23	3.16	3.12	3.06	3.00	2.96	2.89
	.01	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.25	4.10	4.01	3.94	3.86	3.78	3.71	3.61
12	.10	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.10	2.06	2.03	2.01	1.99	1.96	1.94	1.91
	.05	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.62	2.54	2.50	2.47	2.43	2.38	2.35	2.30
	.025	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.18	3.07	3.01	2.96	2.91	2.85	2.80	2.73
	.01	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.01	3.86	3.76	3.70	3.62	3.54	3.47	3.37
13	.10	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
	.05	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.53	2.46	2.41	2.38	2.34	2.30	2.26	2.21
	.025	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.05	2.95	2.88	2.84	2.78	2.72	2.67	2.60
	.01	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.82	3.66	3.57	3.51	3.43	3.34	3.27	3.18
14	.10	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.01	1.96	1.93	1.99	1.89	1.86	1.83	1.80
	.05	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.46	2.39	2.34	2.31	2.27	2.22	2.19	2.14
	.025	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	2.95	2.84	2.78	2.73	2.67	2.61	2.56	2.50
	.01	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.66	3.51	3.41	3.35	3.27	3.18	3.11	3.02
15	.10	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	1.97	1.92	1.89	1.87	1.85	1.82	1.79	1.76
	.05	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.40	2.33	2.28	2.25	2.20	2.16	2.12	2.07
	.025	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.86	2.76	2.69	2.64	2.59	2.52	2.47	2.40
	.01	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.52	3.37	3.28	3.21	3.13	3.05	2.98	2.88

3.6 抽样与抽样分布

从样本到总体的推断

推断统计的概念

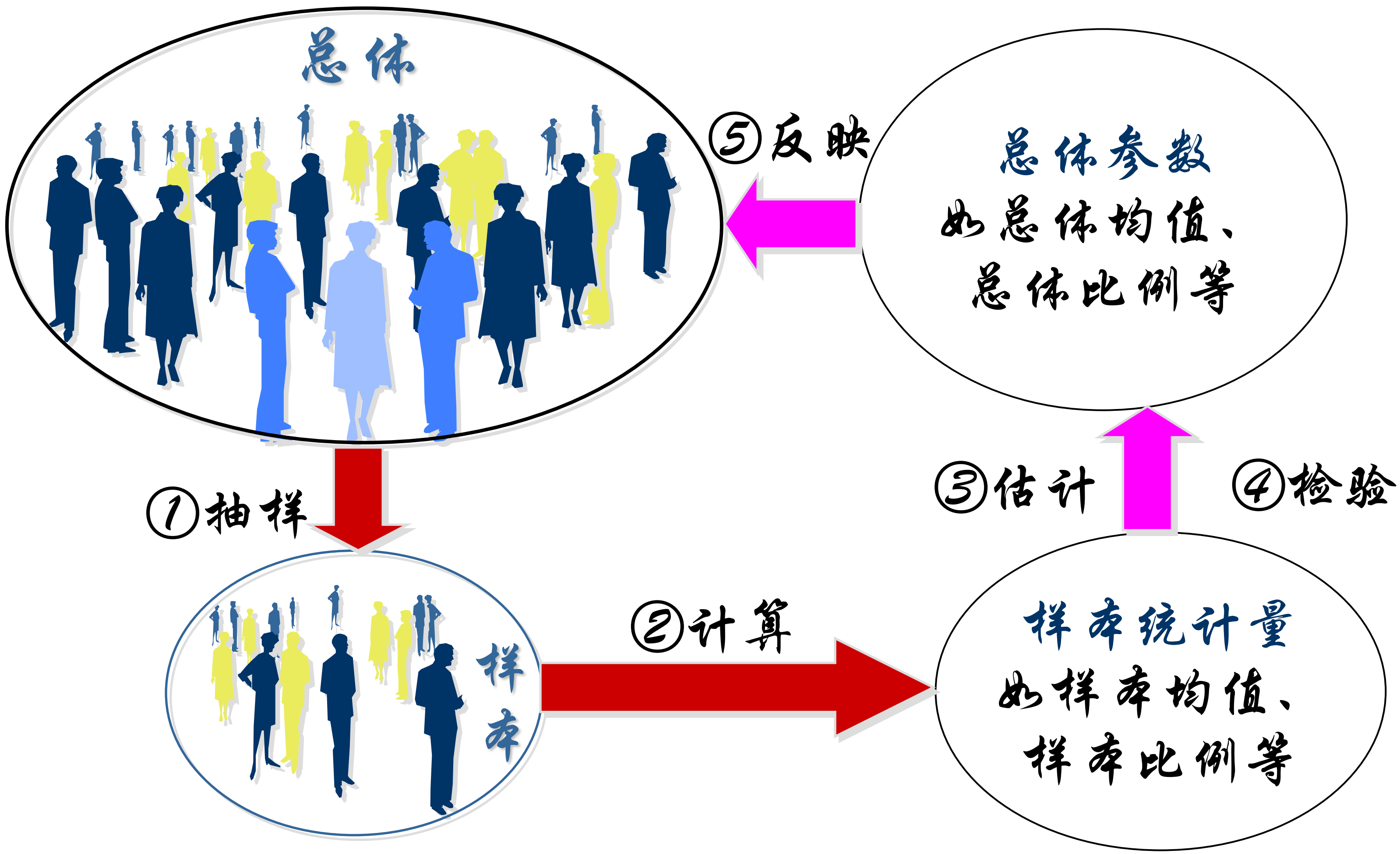
- 按照一定程序和规则从总体中抽取部分单位进行观察，根据被抽取单位的数量特征，运用一定方法对总体信息做出具有一定可靠程度的估计和判断。
- 谚语：“你不必吃完整头牛，才知肉是老的！”

总体 ------ 总体规律（理论分布）？

样本 ------ 样本统计量值

3.6 抽样与抽样分布（续）

➢ 推断统计分析过程



3.6 抽样与抽样分布（续）

- 需要推断统计的情况
  - 推断统计的目的就是利用样本对总体进行认识，当出现下列情况时，需要抽样调查：
  - （1）当研究对象是无限总体时；
  - （2）当进行破坏性或消耗性的检查时；
  - （3）当进行成本太高的调查时；
  - （4）当需对全面调查进行验证、补充和修正时；（5）当用于生产质量控制时；
  - （6）当对某些总体的假设进行检验，判断真伪，为决策服务时。

3.6 抽样与抽样分布（续）

□ 抽样方法

➢ 简单随机抽样（Simple Random Sampling)

- **定义**：将所有总体单位编制成抽样框，然后采用抽签或随机数表从中抽取样本单位。对于有限总体，必须保证每个总体单位被抽中的概率相同，对于无限总体，必须保证每个总体单位的抽取是相互独立。
- **3种方法**：（1）人工抽签；（2）随机数；（3）利用软件直接进行抽样。
- **适合情况**：比较适合于规模不大、总体内部差异小的情况。
- **地位**：简单随机抽样是其他抽样调查组织形式的基础。

3.6 抽样与抽样分布（续）

- 随机数表（Random Numbers）：由一连串的0,1,2,3, 4,5,6,7,8,9这些数字组成，且满足下列两个条件：（1）表中任一位置的数字，其为0-9中任何一个数字的概率相同；（2）不同位置的数字之间是独立的，就是说，知道表某一部分是些什么数字，不会提供你任何关于其他部分是些什么数字的信息。
- 如何从随机数表中进行简单随机抽样：

第1步，根据数据编写代码，要求：（1）所有代码位数相同，保证所有个体都有相同的中选概率；（2）代码尽可能短（一般地，1-10用1位，11-100用2位，101-1000用3位,…）；

第2步，任意选择随机数起点，沿行或沿列读取随机数，选取符合条件的代码，重复代码只选一个。

3.6 抽样与抽样分布（续）

- 利用Excel进行简单随机抽样：

（1）产生随机数：工具/数据分析/随机数发生器/各项设置/确定。

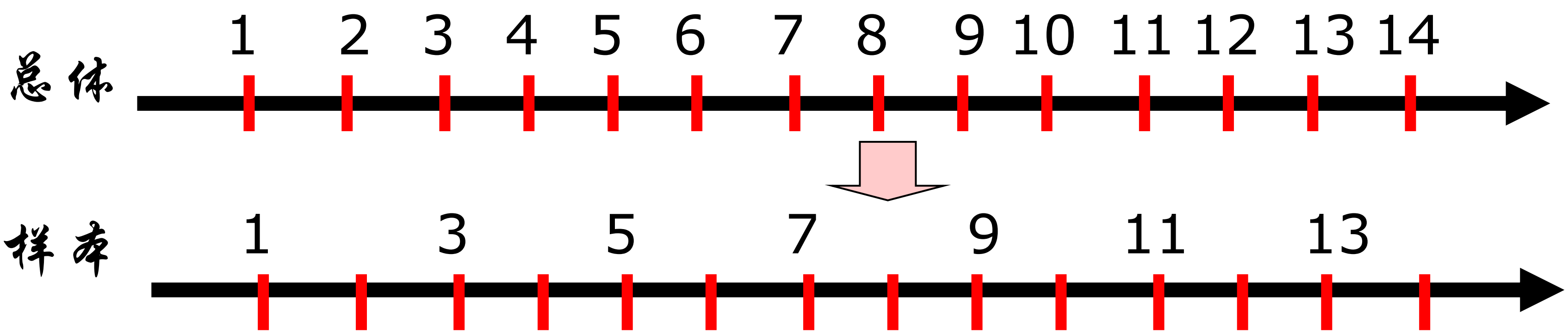
（2）抽样：工具/数据分析/抽样/各项设置/确定。

（3）例：CJW顾客满意得分的简单随机抽样（抽取30样本单位）。

3.6 抽样与抽样分布（续）

➢ 等距抽样（Systematic Sampling)

- 定义：将总体单位按照某一有关或无关标志进行排序，然后在规定的范围内随机选择某一起点并按照一套规则（如按一定的间隔）进行抽取样本单位。
- 等距抽样的最简单形式：

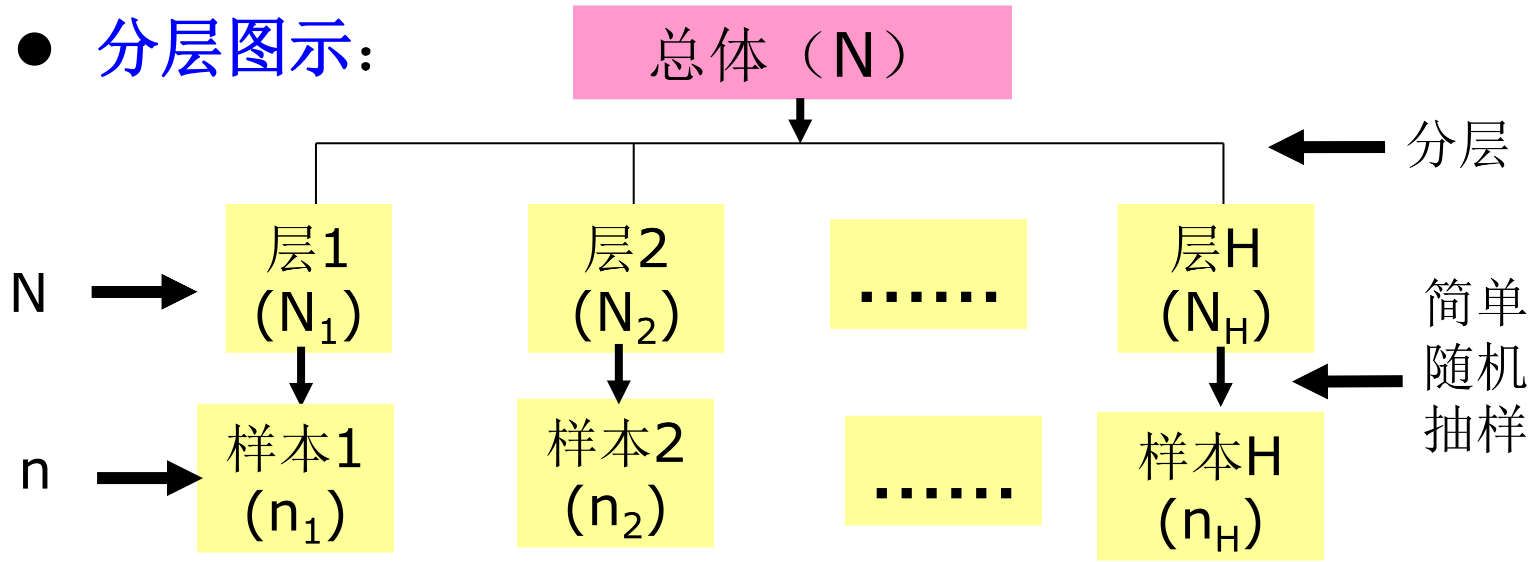


简单等距抽样

- 示例：生产流水线的产品抽检；访问顾客等。

➢ 分层抽样（Stratified Sampling)

- **定义**：根据总体单位的某属性把总体划分为若干层，然后对每层进行简单随机抽样，所有层被抽取总体单位组成样本。
- **分层原则**：层内方差尽量小，层间方差尽量大。
- **分层图示**：

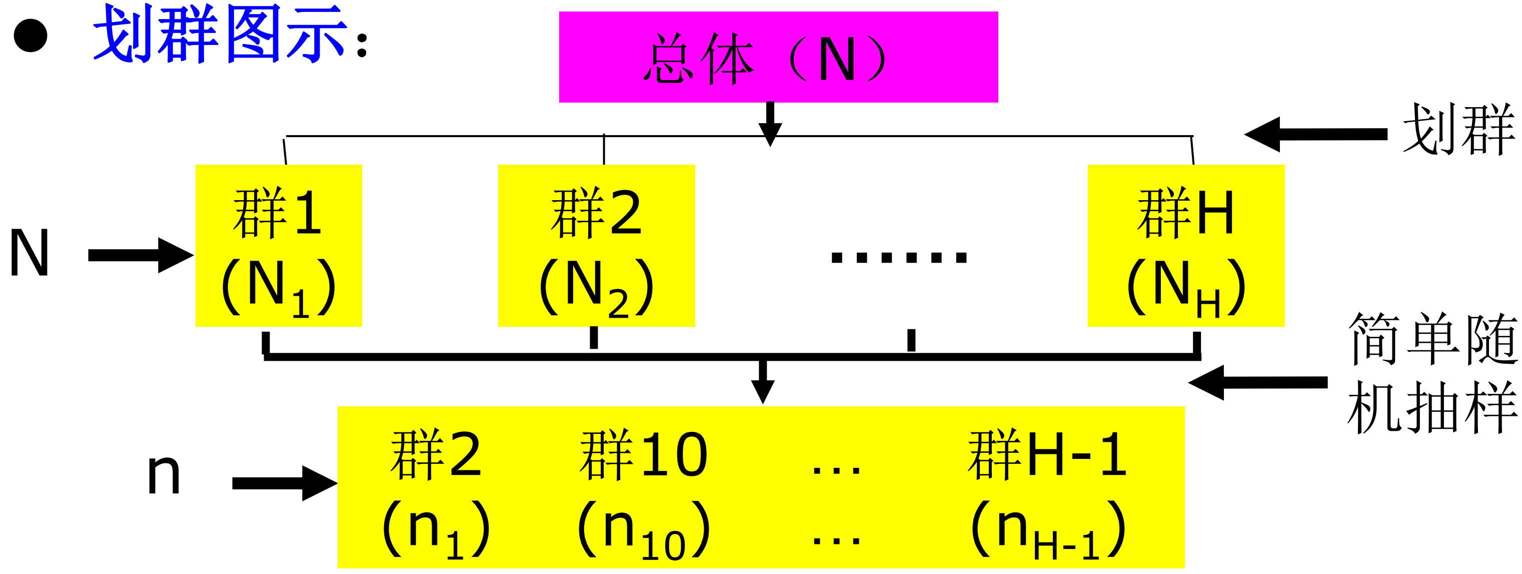


- **示例**：对大学生的调查。

3.6 抽样与抽样分布（续）

➢ 整群抽样（Cluster Sampling)

- **定义**：将总体划分为若干个相互独立的群，然后所有的群实行简单随机抽样，被抽取的群包含的总体单位组成样本。
- **划群原则**：群内方差尽量大，群间方差尽量小。
- **划群图示**：



- **示例**：对大学生的调查。

3.6 抽样与抽样分布（续）

➢ 多阶段抽样（Multi-stage Sampling）

- 多阶段抽样是分层抽样和整群抽样的结合，即先进行多个整群抽样，然后进行分层抽样。
- 示例：从某省1000多万户农户中抽取1000户调查生产性投资情况。可分为以下4个阶段：

（1）从省的所有县中随机抽取5个县；

（2）从抽中的5个县中各随机抽取4个乡；

（3）从抽中的4个乡中各随机抽取5个村；

（4）从抽中的100个村各随机抽取10个户。

3.6 抽样与抽样分布（续）

➢ 抽样调查应用示例：2015年全国1%人口抽样调查

- **调查对象：**以现有人口+本户户籍外出人口作为调查对象，不调查港、澳、台及外籍人员。
- **调查内容：**“住户项目”和“个人项目”两大项。“住户项目”主要调查住户的人口和住房方面的主要内容，共12个调查项目；“个人项目”主要调查个人的性别、年龄、民族、受教育程度、迁移流动、就业、社会保障、婚姻、生育、身体健康状况等指标，共32个调查项目。
- **广东省样本：**采用二阶段、分层、概率比例、整群抽样等方法抽样，全省共抽中乡、镇、街道1620个，中选村（居）委会12184个，中选调查小区13038个，调查的常住人口约326万人，户数约125万户。

3.6 抽样与抽样分布（续）

- **问题：**
  - 分层抽样和整群抽样分别在什么情况下才能取得最佳效果？

- **其他抽样方法——非概率抽样**
  - 方便抽样：按照方便取得为原则来抽取所需样本单位的一种非概率抽样方法，等；
  - 判断抽样：根据调查人员的判断来抽取认为具有代表性的样本单位的一种非概率抽样方法，
  - 示例：对购物中心消费者满度的调查；对包装箱中的桔子质量的检查；对本科生生活习惯的调查等。

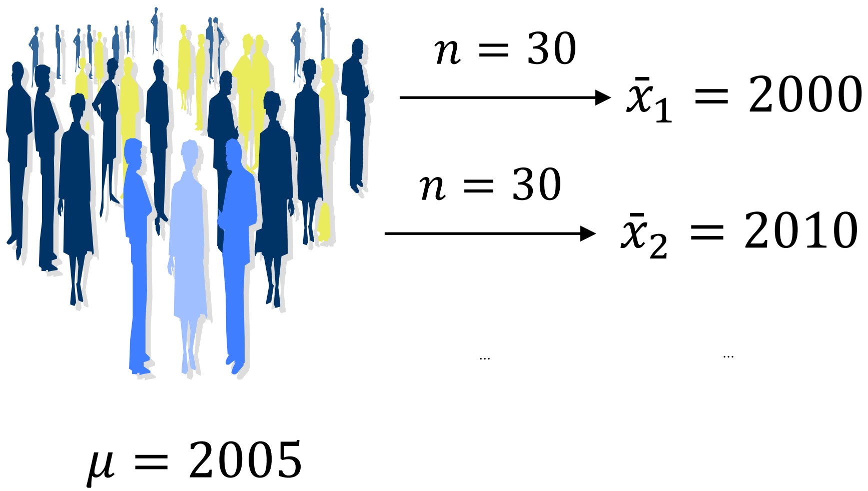
3.6 抽样与抽样分布（续）

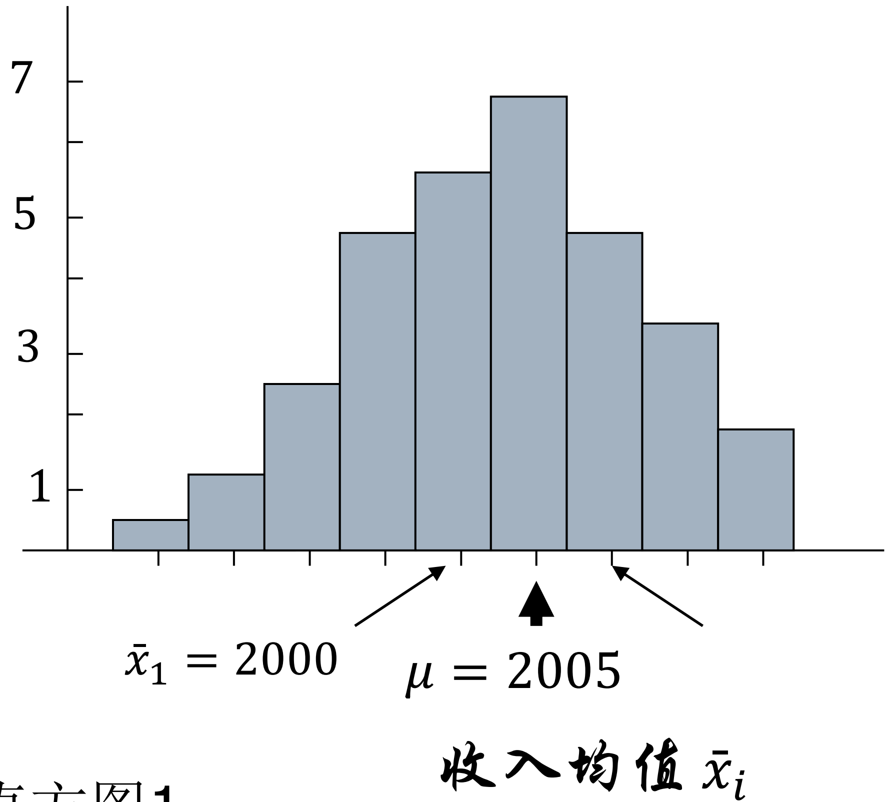
□ 样本统计量是随机变量

- 从一个总体可抽取出多个具有相同样本容量的样本，每个样本都可计算出一个样本均值，因此，从一个总体可计算出多个样本均值。样本均值会随着样本的改变而改变。在简单随机抽样下，每个样本均值 $\bar{x}_i$ 出现的概率是相等的。
- 从某公司中随机抽取30个员工，调查平均收入，得到 $\bar{x}_1 = 2000$ 元；再从该公司中抽取另一个30个员工，又得到 $\bar{x}_2 = 2010$ 元；……，一直重复这样的程序，可得到一系列样本均值的不同取值 $\bar{x}_1$、$\bar{x}_2$、…、$\bar{x}_m$。而且根据这些值可绘制直方图。
- **问题**：不同的样本容量n将会对样本均值的直方图形状产生什么样的影响？

3.6 抽样与抽样分布（续）

- 从公司所有员工中抽出50个大小为30的样本。画出50个样本均值的直方图，长方形高度代表50个样本中有多少样本均值落在长方形底部范围内。





样本均值的直方图1

3.6 抽样与抽样分布（续）

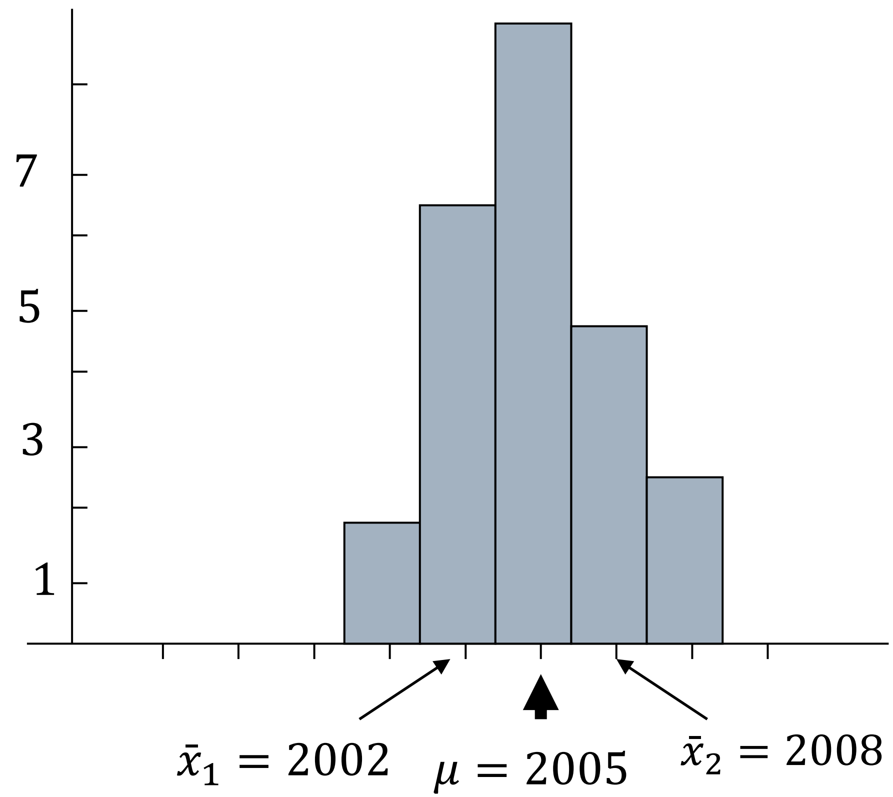
- 同前面一样，改变的只是样本容量由30个增加为100个员工。样本个数仍为50个。与前面的直方图相比，直方图变得更窄（更集中）！更接近于总体均值，样本均值的方差（变异性）更小！

$\mu = 2005$

$\xrightarrow{n = 100} \bar{x}_1 = 2002$

$\xrightarrow{n = 100} \bar{x}_2 = 2008$

… …



样本均值的直方图2

3.6 抽样与抽样分布（续）

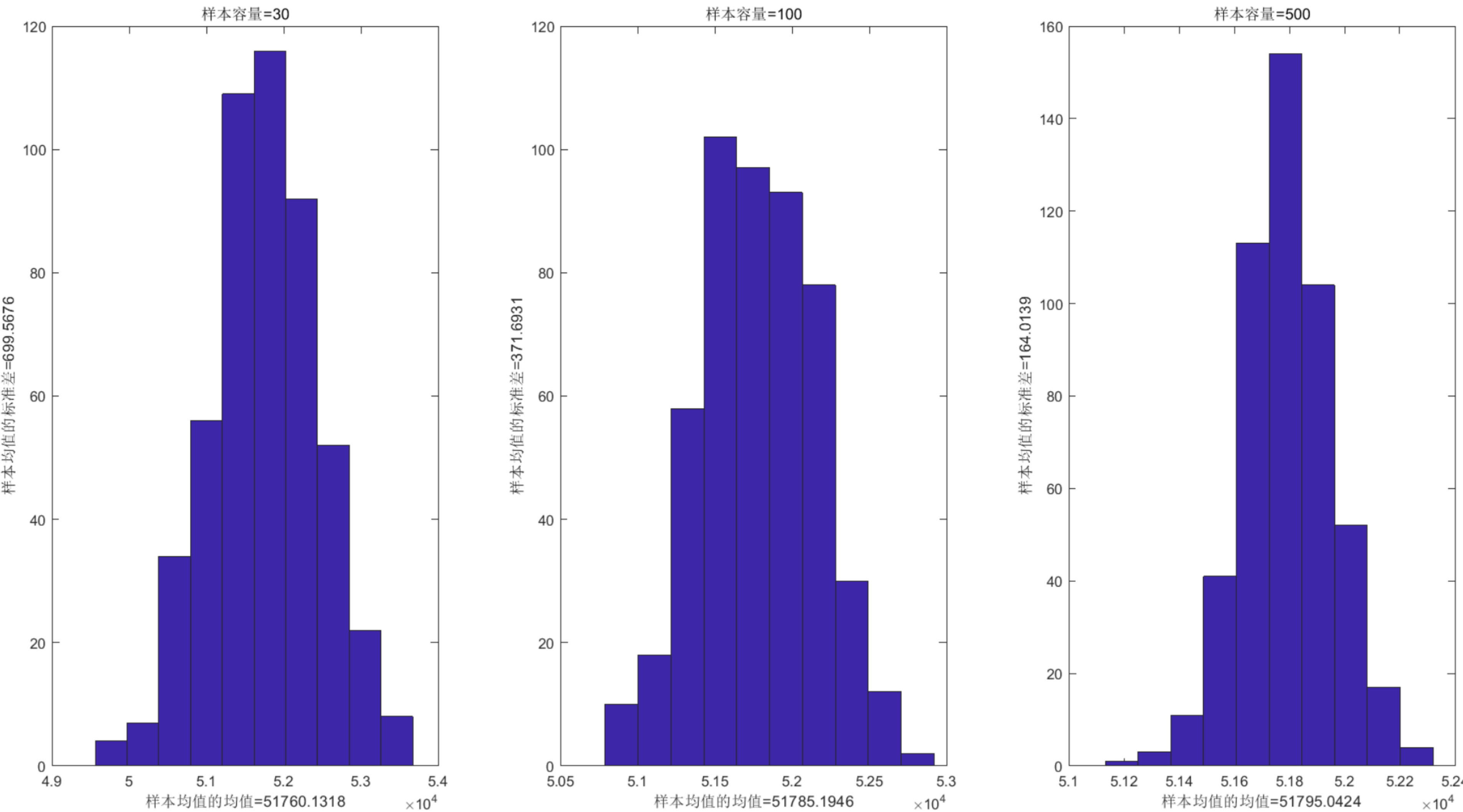
- **样本均值抽样分布实例分析：**EAI管理人员年薪

（1）总体数据描述统计分析：

总体观察值数目N=2500，总体均值μ=51800，总体标准差σ=4000

（3）分析要求：

从总体数据中随机抽取样本容量分别为30,100,500的500个样本，绘制样本均值的直方图，对比三个样本容量下的样本均值的均值和标准差，并思考样本容量对它们的影响以及它们与总体均值和标准差的联系。


样本容量=30
样本均值的标准差=699.5676
样本均值的均值=51760.1318
样本容量=100
样本均值的标准差=371.6931
样本均值的均值=51785.1946
样本容量=500
样本均值的标准差=164.0139
样本均值的均值=51795.0424

样本均值的均值和标准差

设总体X的均值为μ，方差为σ^2，$X_1, X_2, \cdots, X_n$是来自总体的一个样本，则样本均值$\bar{X}$和样本方差S^2有

$$E(\bar{X}) = \mu,$$

$$D(\bar{X}) = \sigma^2 / n,$$

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

3.6 抽样与抽样分布（续）

样本均值的均值

- 样本均值的均值等于总体均值，即 $E(\bar{x}) = \mu$ 。

样本均值的标准差

- 对有限总体，

$$\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1}}\left(\frac{\sigma}{\sqrt{n}}\right)$$

其中，N 为总体单位数，n为样本容量，σ 为总体标准差，通常用样本标准差估计，$\sqrt{(N-n)/(N-1)}$ 为有限总体修正系数。

- 对无限总体(或满足 $n/N \le 0.05$ 的有限总体)，

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

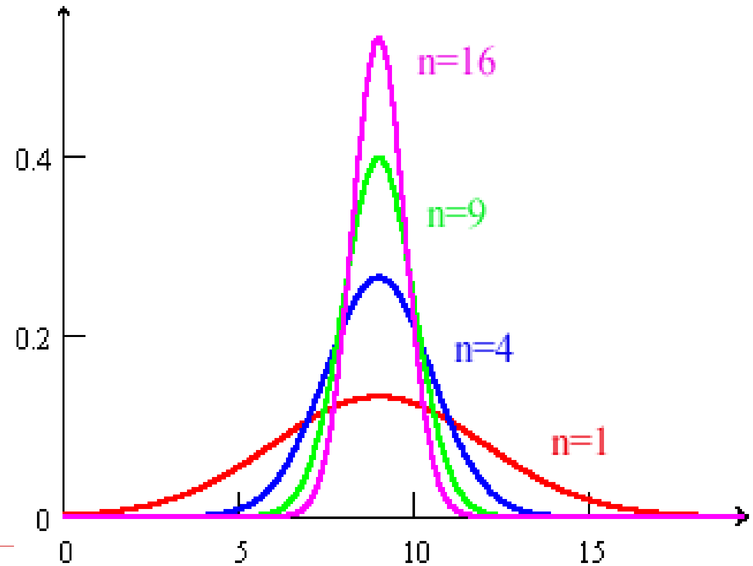
3.6 抽样与抽样分布（续）

样本均值的抽样分布

- 定义
  - 样本均值抽样分布：所有可能取值的概率分布。
- 样本均值的抽样分布：总体分布、总体方差已知

设 $X_1, X_2, \ldots, X_n$ 是来自正态总体 $N(\mu,\sigma^2)$ 的样本，$\bar{X}$ 是样本均值，则有 $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$，

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



当n取不同值时，样本均值 $\bar{x}$ 的分布

3.6 抽样与抽样分布（续）

- 样本均值的抽样分布：总体分布已知、总体方差未知

设 $X_1,X_2,\ldots,X_n$ 是取自正态总体 $N(\mu,\sigma^2)$ 的样本， $\bar{X}$ 和 S^2 分别为样本均值和样本方差，则有

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

3.6 抽样与抽样分布（续）

➢ 样本均值的抽样分布：大样本

- 当样本为大样本（即样本容量n不小于30）时，样本均值的抽样分布都近似为正态分布，即：

$$\bar{x} \sim N(\mu_{\bar{x}}, \sigma_{\bar{x}}^2)$$

或者，

$$\bar{x} \sim N(\mu_{\bar{x}}, s_{\bar{x}}^2)$$

3.6 抽样与抽样分布（续）

➢ **样本均值的抽样分布：小样本，总体服从方差已知的正态分布**

- **当样本为小样本（即样本容量n小于30），而且总体服从方差已知的正态分布时，样本均值的抽样分布近似为正态分布，即：**

$$\bar{x} \sim N(\mu_{\bar{x}}, \sigma_{\bar{x}}^2)$$

3.6 抽样与抽样分布（续）

➢ **样本均值的抽样分布：小样本，总体服从方差未知的正态分布**

- **当样本为小样本（即样本容量n小于30），而且总体服从方差未知的正态分布时，样本均值的抽样分布近似为学生t分布，即：**

$$\frac{\bar{X}-\mu_{\bar{x}}}{s_{\bar{x}}} \sim t(n-1)$$

3.6 抽样与抽样分布（续）

- 样本均值的抽样分布： 小样本，总体分布未知
  - 当总体不服从（或不近似服从）正态分布，样本为小样本时，样本均值的抽样分布形式难以确定，一般需通过增大样本容量来解决。

3.6 抽样与抽样分布（续）

- 示例：计算样本均值落在总体均值某范围的概率

  - EAI公司人事主管部门对2500名经理的年薪进行了调查，发现所有经理的平均年薪及其标准差分别为51800美元和4000美元。假如从2500名经理中随机抽取30名调查，人事主管请你：

（1）计算抽出的30名经理的平均年薪在51300至52300美元的概率是多少？

（2）比较在样本容量分别为30,100,500情况下介于51300和52300之间的概率。

3.6 抽样与抽样分布（续）

➢ 样本方差的抽样分布

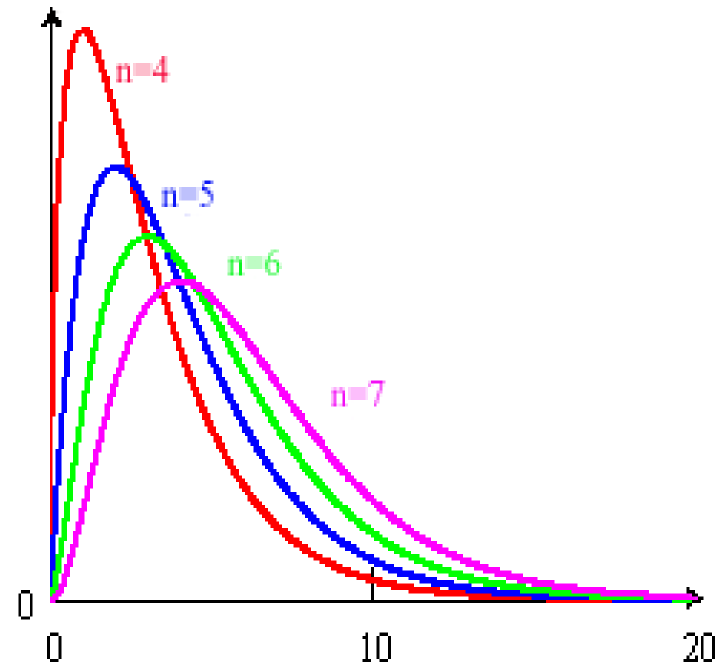
设$X_1,X_2,\ldots,X_n$是来自正态总体 $N(\mu,\sigma^2)$ **的样本，$\bar{X}$和S^2分别为样本均值和样本方差，则有**

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

(2) $\bar{X}$与S^2独立

当n取不同值时，

$$\frac{(n-1)S^2}{\sigma^2}$$ **的分布**



<table>
<tr><th>对象</th><th>目标</th><th>条件</th><th colspan="2">形式</th><th>备注</th></tr>
<tr><td rowspan="6">样本均值</td><td>均值</td><td>无</td><td colspan="2"></td><td rowspan="3"></td></tr>
<tr><td rowspan="2">标准差</td><td>无限总体或n/N≤0.05的有限总体</td><td colspan="2"></td></tr>
<tr><td>有限总体</td><td colspan="2"></td></tr>
<tr><td rowspan="3">抽样分布</td><td></td><td>大样本</td><td>小样本且总体服从正态分布</td><td rowspan="3"></td></tr>
<tr><td>总体标准差已知</td><td></td><td></td></tr>
<tr><td>总体标准差未知</td><td></td><td></td></tr>
</table>

概率分布的主要问题回顾

- 1. 正态分布的“3σ原则”
- 2. 总体分布、样本分布和抽样分布的含义
- 3. 概率抽样与非概率抽样的涵义和具体形式
- 4. 抽样调查的概念、目的和原因
- 5. 样本均值的均值、方差和抽样分布形式