数理统计

Mathematical Statistics

第一节　基本概念

数理统计的分类

描述统计学

对随机现象进行观测、试验，以取得有代表性的观测值

推断统计学

对已取得的观测值进行整理、分析，作出推断、决策，从而找出所研究的对象的规律性

第一节　抽样和抽样分布

一、总体和个体

二、样本、简单随机样本

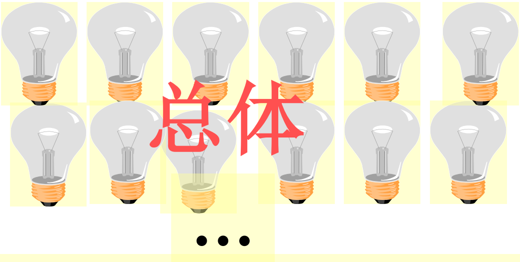
三、抽样分布基本定理

一、总体和个体

一个统计问题总有它明确的研究对象.

研究对象的全体称为总体(母体),

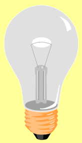
组成总体的每个元素或成员称为个体.



研究某批灯泡的质量

然而在统计研究中，人们关心总体仅仅是关心其每个个体的一项(或几项)数量指标和该数量指标在总体中的分布情况. 这时，每个个体具有的数量指标的全体就是总体.

某批
灯泡的寿命



该批灯泡寿命的
全体就是总体

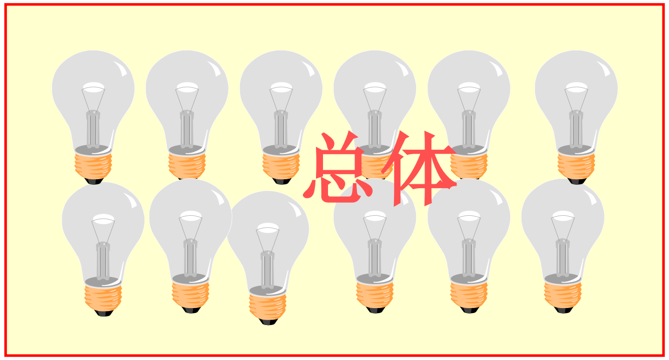
国产轿车每公里耗油
量的全体就是总体

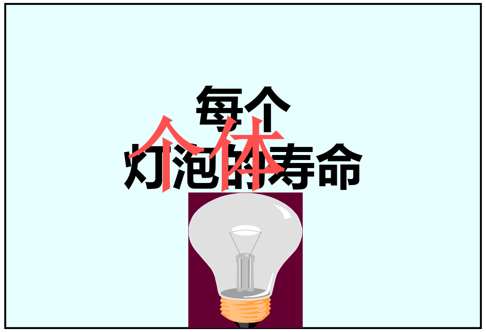
总体:

所研究的对象的某个(或某些)数量指标的全体称为总体,它是一个随机变量(或多维随机变量)，记为X.

X的分布函数和数字特征称为总体分布函数和总体数字特征.

例如:研究某批灯泡的寿命时，总体X是这批灯泡的寿命，而其中每个灯泡的寿命就是个体。





又如:研究某批国产轿车每公里的耗油量时，总体X是这批轿车每公里的耗油量，而其中每辆轿车的耗油量就是个体。

国产轿车每公里耗油量的全体就是总体

类似地，在研究某地区中学生的营养状况时，若关心的数量指标是身高和体重，我们用X和Y分别表示身高和体重，那么此总体就可用二维随机变量(X,Y) 来表示，而每个学生的身高和体重就是个体.

二、样本和简单随机样本

1）抽样和样本

为推断总体分布及各种特征，按一定规则从总体中抽取若干个体进行观察试验，以获得有关总体的信息，这一抽取过程称为“抽样”，所抽取的部分个体称为样本. 样本中所包含的个体数目称为样本容量.

样本的抽取是随机的，每个个体是一个随机变量. 容量为 n 的样本可以看作 n 维随机变量，用 $X_1,X_2,\ldots,X_n$ 表示.

而一旦取定一组样本，得到的是 n 个具体的数 $(x_1, x_2, \ldots, x_n)$，称其为样本的一个观察值，简称样本值 .

2）简单随机样本

由于抽样的目的是为了对总体进行统计推断，为了使抽取的样本能很好地反映总体的信息，必须考虑抽样方法.
最常用的一种抽样方法叫作“简单随机抽样”，它要求抽取的样本满足下面两点:

1. 样本 $X_1, X_2, \ldots, X_n$ 中每一个 X_i 与所考察的总体 X 有相同的分布.

2. $X_1, X_2, \ldots, X_n$ 相互独立.

设 $X_1, X_2, \ldots, X_n$ 是总体 X 的一个简单随机样本，

1) 若 X 为离散型总体，其分布律是 $p(x)$，则 $X_1, X_2, \ldots, X_n$ 的联合分布律为

$$p(x_1)\, p(x_2) \ldots p(x_n)$$

2) 若 X 为连续型总体，其概率密度是 $f(x)$，则 $X_1, X_2, \ldots, X_n$ 的联合分布律为

$$f(x_1)\, f(x_2) \ldots f(x_n)$$

3）总体、样本、样本值的关系

事实上我们抽样后得到的资料都是具体的、确定的值．如我们从某班大学生中抽取10人测量身高，得到10个数，它们是样本取到的值而不是样本．我们只能观察到随机变量取的值而见不到随机变量．

统计是从手中已有的资料 — 样本值，去推断总体的情况 — 总体分布 $F(x)$ 的性质.

总体分布决定了样本取值的概率规律，也就是样本取到样本值的规律，因而可以由样本值去推断总体.

4）经验分布函数

设 $X_1, X_2, \ldots, X_n$ 为取自总体 X 的样本， $x_1, x_2, \ldots, x_n$ 为其观察值. 对于每个固定的 x，设事件 $\{X \le x\}$ 在 n 次观察中出现的次数为 $v_n(x)$， 于是事件 $\{X \le x\}$ 发生的频率为：

$$F_n(x) = \frac{v_n(x)}{n} \qquad -\infty < x < +\infty$$

显然 $F_n(x)$ 为不减右连续函数，且

$$F_n(-\infty) = 0，F_n(+\infty) = 1$$

称 $F_n(x)$ 为样本分布函数或经验分布函数.

定理（格列文科）当 $n\to\infty$ 时，经验分布函数 $F_n(x)$ 依概率 1 关于 x 一致收敛与总体分布函数，即

$$P\{\lim_{n\to\infty}\sup_{-\infty<x<+\infty}|F_n(x)-F(x)|=0\}=1$$

定理表明：当样本容量 n 充分大时，经验分布函数 $F_n(x)$ 几乎一定会充分趋近总体分布函数 $F(x)$，这是**用样本来推断总体的理论依据**.

第二节　　统计量与抽样分布

一、统计量

二、统计学中三个常用分布和上 α 分位点

三、抽样分布定理

一、统计量

由样本值去推断总体情况，需要对样本值进行“加工”，这就要构造一些样本的函数，它把样本中所含的信息集中起来.

定义

如果样本$X_1, X_2, \ldots, X_n$的函数$g(X_1, X_2, \ldots, X_n)$不含有任何的未知参数，则称函数$g(X_1, X_2, \ldots, X_n)$为统计量.

若$x_1, x_2, \ldots, x_n$是相应的样本值，则称函数值$g(x_1, x_2, \ldots, x_n)$为统计量$g(X_1, X_2, \ldots, X_n)$的一个观察值.

例如：$X \sim N(\mu,\sigma^2)$，μ,σ^2 是未知参数，

$X_1, X_2, \cdots, X_n$ 是X 的一个样本，则

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

是统计量，而 $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$ 不是统计量.

若 μ,σ^2 已知，则 $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$ 也是统计量.

几个常用的统计量

它反映了总体均值的信息

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

它反映了总体方差的信息

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

它反映了总体 k 阶矩的信息

样本 k 阶原点矩 $$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$$

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

它反映了总体 k 阶中心矩的信息

$k=1,2,\ldots$

它们的观察值分别为：

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i \qquad s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

$$a_k=\frac{1}{n}\sum_{i=1}^{n}x_i^k \qquad b_k=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^k$$

由大数定律可知：

$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^{\,k}$ 依概率收敛于 $E(X^k)$

例1. 从一批相同的电子元件中随机地抽出8个，测得使用寿命（单位：小时）分别为：2300，2430，2580，2400，2280，1960，2460，2000，试计算样本均值、样本方差及样本二阶矩.

解：

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 2301.25(\text{小时})$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 48126.7857(\text{小时}^2)$$

$$a_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = 5337862.5(\text{小时}^2)$$

二、统计学中三个常用分布和上α分位点

抽样分布

统计量是样本的函数，而样本是随机变量，故统计量也是随机变量，因而就有一定的分布，它的分布称为“**抽样分布**”.

下面介绍三个来自正态总体的抽样分布.

1、 χ^2 分布

定义: 设 $X_1, X_2, \cdots, X_n$ 相互独立,都服从标准正态分布$N(0,1)$,

则称随机变量:

$$\chi^2 = X_1^2 + X_2^2 + \cdots + X_n^2$$

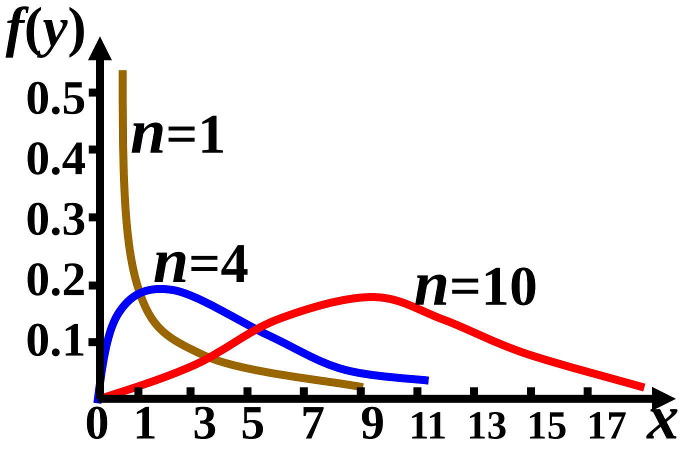
所服从的分布为自由度为 n 的 χ^2分布，记为 $\chi^2 \sim \chi^2(n)$

χ^2分布的概率密度为

$$f(y) = \begin{cases} \dfrac{1}{2^{n/2}\Gamma(n/2)} y^{\frac{n}{2}-1} e^{-\frac{y}{2}} & y > 0 \\ 0 & 其它 \end{cases}$$

其中 $\Gamma(\frac{n}{2})$ 是函数 $\Gamma(s) = \int_0^\infty e^{-t} t^{s-1} dt$ （$s > 0$） 在 $s = \dfrac{n}{2}$ 处的值.

χ^2分布的概率密度图形如下:



显然 χ^2 分布的概率密度图形随自由度的不同而有所改变.

χ^2分布的性质：

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

证明： 设 $\chi^2 = \sum_{i=1}^{n} X_i^2 \quad X_i \sim N(0,1) \;\; i = 1,2,\cdots,n$

$X_1, X_2, \cdots, X_n$ 相互独立,则

$$E(X_i) = 0,\; D(X_i) = 1,\;\; E(X_i^2) = D(X_i) + [E(X_i)]^2 = 1,$$

$$E(\chi^2) = E\left(\sum_{i=1}^{n} X_i^2\right) = \sum_{i=1}^{n} E(X_i^2) = n$$

$$E(X_i^4) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^4 e^{-\frac{x^2}{2}}\mathrm{d}x = 3$$

$$D(X_i^2) = E(X_i^4) - E^2(X_i^2) = 2$$

$$D(\chi^2(n)) = D\left(\sum_{i=1}^{n} X_i^2\right) = 2n$$

性质2. 设 $\chi_1^2 \sim \chi^2(n_1)$，$\chi_2^2 \sim \chi^2(n_2)$，且 χ_1^2 与 χ_2^2 相互独立，则

$$\chi_1^2 + \chi_2^2 \sim \chi^2(n_1 + n_2)$$

这个性质称为 χ^2分布的可加性.

2、t 分布

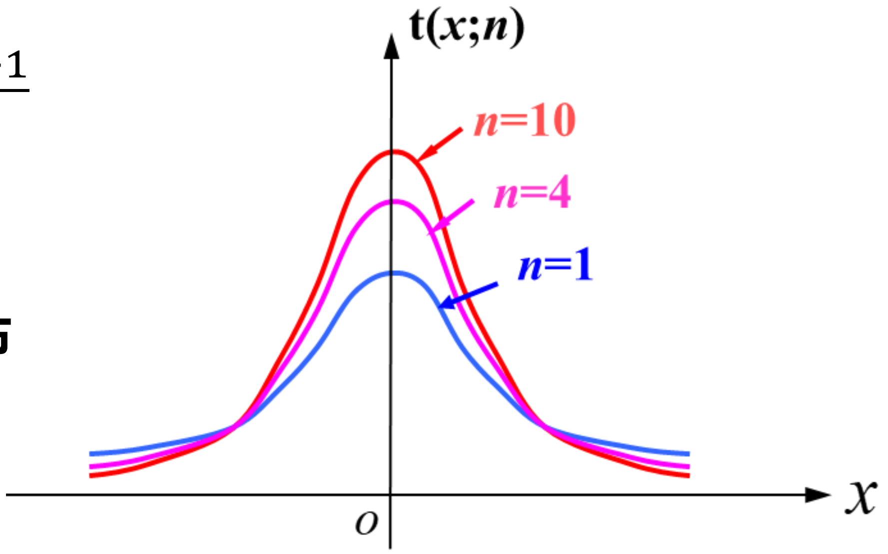
定义： 设 $X \sim N(0, 1)$，$Y \sim \chi^2(n)$，且 X 与 Y 相互独立，

则称变量 $t = \dfrac{X}{\sqrt{Y/n}}$

所服从的分布为自由度为 n 的 t 分布. 记为 $t \sim t(n)$.

t 的概率密度为：

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



t 分布的概率密度函数关于 $t=0$ 对称，且当 n 充分大时 ($n \geq 30$)，其图形与标准正态分布的概率密度函数的图形非常接近. 但对于较小的 n，t 分布与 $N(0,1)$ 分布相差很大.

3、F分布

定义： 设 $X \sim \chi^2(n_1), Y \sim \chi^2(n_2)$， X 与 Y 相互独立，

则称统计量 $F = \dfrac{X/n_1}{Y/n_2}$

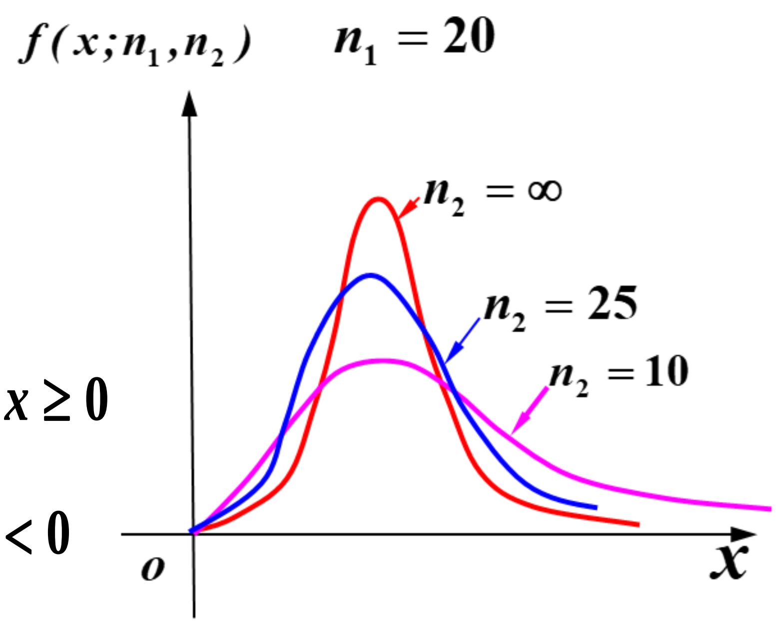
服从自由度为 n_1 及 n_2 的 F 分布，n_1 称为第一自由度，

n_2称为第二自由度，记作 $F \sim F(n_1,n_2)$.

由定义可见，$\dfrac{1}{F} = \dfrac{Y/n_2}{X/n_1} \sim F(n_2, n_1)$

若 $X \sim F(n_1,n_2)$，则 X 的概率密度为

$$\psi(x) = \begin{cases} \dfrac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{2})\,\Gamma(\frac{n_2}{2})}\left(\dfrac{n_1}{n_2}\right)\left(\dfrac{n_1}{n_2}x\right)^{\frac{n_1}{2}-1}\left(1+\dfrac{n_1}{n_2}x\right)^{-\frac{n_1+n_2}{2}} & x \geq 0 \\ 0 & x < 0 \end{cases}$$



4、上α分位点

定义： **设随机变量X的概率密度为 $f(x)$, 对于任意给定的 α($0<\alpha<1$), 若存在实数 x_α, 使得：**

$$P\{X \geq x_\alpha\} = \int_{x_\alpha}^{+\infty} f(x)dx = \alpha$$

则称点 x_α 为该概率分布的上 α 分位点

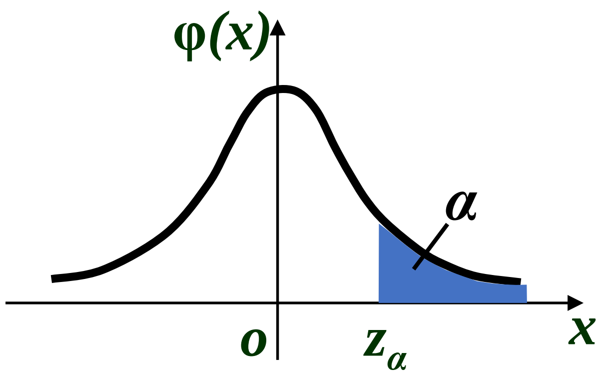
正态分布的上α分位点

对标准正态分布变量 $Z \sim N(0,1)$ 和给定的α，上α分位数是由：

$$P\{Z \geq z_\alpha\} = \int_{z_\alpha}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \alpha$$

即 $P\{Z<z_\alpha\} = 1-\alpha$　　$\Phi(z_\alpha) = 1-\alpha$　　确定点 z_α.

如图：



例如，$\alpha=0.05$，而

$P\{Z\geq 1.645\} = 0.05$

所以，$z_{0.05} = 1.645$.

说明:

1） 除标准正态分布外，χ^2分布、t 分布、F 分布的上 α 分位点都有表可查.

2）对于 χ^2分布，当 n 充分大时（n>45），

$$\chi_\alpha{}^2 \approx \frac{1}{2}(Z_\alpha + \sqrt{2n-1})^2$$

其中 Z_α 是标准正态分布的上 α 分位点

3）对于 t 分布

a) 由其对称性，有： $t_{1-\alpha}(n) = -t_\alpha(n)$

b) 当n充分大时（n>45），$t_\alpha(n) = Z_\alpha$

4）对于 F 分布，有：$F_{1-\alpha}(n_1, n_2) = \dfrac{1}{F_\alpha(n_2, n_1)}$

例2. 查表求下列值： $t_{0.01}(5)$, $t_{0.95}(6)$, $F_{0.1}(10,9)$, $F_{0.9}(28,2)$, $\chi^2_{0.25}(20)$, $z_{0.01}$.

解： $t_{0.01}(5) = 3.3649$

$$t_{0.95}(6) = -t_{0.05}(6) = -1.9432$$

$$F_{0.1}(10,9) = 2.42$$

$$F_{0.9}(28,2) = \frac{1}{F_{0.1}(2,28)} = \frac{1}{2.50} = 0.4$$

$$\chi^2_{0.25}(20) = 23.828 \qquad z_{0.01} = 2.33$$

例3.设总体X和Y相互独立，同服从$N(0,3^2)$分布，而 X_1，X_2，…, X_9和 Y_1，Y_2，…, Y_9分别是来自X和Y的简单随机样本，求统计量

$$U=\frac{X_1+X_2+\cdots+X_9}{\sqrt{Y_1^2+Y_2^2+\cdots+Y_9^2}}$$ 的分布.

解： $\because X_i\sim N(0,9)$ $\quad\therefore \sum_{i=1}^{9}X_i\sim N(0,81)$

$$\therefore \frac{\sum_{i=1}^{9}X_i}{9}\sim N(0,1)$$

$\because Y_i\sim N(0,9)$ $\quad\therefore (\frac{Y_i}{3})^2\sim\chi^2(1)$

$$\therefore \sum_{i=1}^{9}\frac{Y_i^2}{9}=\frac{\sum_{i=1}^{9}Y_i^2}{9}\sim\chi^2(9)$$

$$\therefore U=\frac{\sum_{i=1}^{9}X_i/9}{\sqrt{\sum_{i=1}^{9}Y_i^2/81}}$$

$$=\frac{X_1+X_2+\cdots+X_9}{\sqrt{Y_1^2+Y_2^2+\cdots+Y_9^2}}$$

$$\sim t(9)$$

例4.设总体X服从 $N(0,2^2)$ 分布，而

X_1，X_2，…,X_{15}是来自X的简单随机样本，求

统计量 $Y=\dfrac{X_1^2+X_2^2+\cdots+X_{10}^2}{2(X_{11}^2+X_{12}^2+\cdots+X_{15}^2)}$ 的分布.

解： $\because X_i \sim N(0,2^2)$ $\qquad \therefore (\dfrac{X_i}{2})^2 \sim \chi^2(1)$

$$\therefore \sum_{i=1}^{10}\frac{X_i^2}{4}=\frac{\sum_{i=1}^{10}X_i^2}{4}\sim \chi^2(10)$$

$$\therefore \sum_{i=11}^{15}\frac{X_i^2}{4}=\frac{\sum_{i=11}^{15}X_i^2}{4}\sim \chi^2(5)$$

$$\frac{(X_1^2+X_2^2+\cdots+X_{10}^2)/40}{(X_{11}^2+X_{12}^2+\cdots+X_{15}^2)/20}=\frac{X_1^2+X_2^2+\cdots+X_{10}^2}{2(X_{11}^2+X_{12}^2+\cdots+X_{15}^2)}=Y$$

$$\sim F(10,5)$$

三、抽样分布定理

定理 1 设$X_1,X_2,\ldots,X_n$是取自正态总体 $N(\mu,\sigma^2)$ 的样本，则有

（1）样本均值 $\overline{X} \sim N(\mu,\dfrac{\sigma^2}{n})$

（2）样本均值 $\overline{X}$与样本方差 S^2相互独立。

（3）随机变量

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

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

则有

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

定理 3 (两个总体样本均值差的分布)

设$X \sim N(\mu_1, \sigma^2)$, $Y \sim N(\mu_2, \sigma^2)$, **且$X$与$Y$独立,**

$X_1, X_2, \ldots, X_{n_1}$ **是取自X的样本,** $Y_1, Y_2, \ldots, Y_{n_2}$ **是取自Y的样本,** $\overline{X}$ 和 $\overline{Y}$ **分别是这两个样本的样本均值,**

S_1^2和S_2^2 **分别是这两个样本的样本方差, 则有**

$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

定理 4 (两个总体样本方差比的分布)

设 $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$, 且 X 与 Y 独立,

$X_1, X_2, \ldots, X_{n_1}$ 是取自X的样本, $Y_1, Y_2, \ldots, Y_{n_2}$ 是取自

Y 的样本, $\overline{X}$和$\overline{Y}$ 分别是这两个样本的样本均值,

S_1^2 和 S_2^2 分别是这两个样本的样本方差,则有

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

上述4个抽样分布定理很重要，要牢固掌握.

例6. 设 $X \sim N(72, 100)$，为使样本均值大于70的概率的概率不小于90%,则样本容量至少取多少?

解： 设样本容量为 n，则 $\overline{X} \sim N(72, \frac{100}{n})$

$$P(\overline{X} > 70) = 1 - P(\overline{X} \le 70)$$

$$= 1 - \Phi\left(\frac{70 - 72}{\frac{10}{\sqrt{n}}}\right) = \Phi\left(0.2\sqrt{n}\right)$$

令 $\Phi\left(0.2\sqrt{n}\right) \ge 0.9$ 得 $0.2\sqrt{n} \ge 1.29$

即 $n \ge 41.6025$

所以至少取 $n = 42$

例7. 从正态总体 $X \sim N(\mu, \sigma^2)$ 中，抽取了 $n = 20$ 的样本 $X_1, X_2, \cdots, X_{20}$

① $P\left(0.38165\sigma^2 \le \frac{1}{20}\sum_{i=1}^{20}\left(X_i - \overline{X}\right)^2 \le 1.80955\sigma^2\right)$

② $P\left(0.3717\sigma^2 \le \frac{1}{20}\sum_{i=1}^{20}\left(X_i - \mu\right)^2 \le 1.7085\sigma^2\right)$

解： (1) $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ 即

$$\frac{19S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{20}\left(X_i - \overline{X}\right)^2 \sim \chi^2(19)$$

例7. 从正态总体 $X \sim N(\mu, \sigma^2)$ 中，抽取了

$$⑴\ P\left(0.38165\sigma^2 \le \frac{1}{20}\sum_{i=1}^{20}\left(X_i - \overline{X}\right)^2 \le 1.80955\sigma^2\right)$$

$$⑵\ P\left(0.3717\sigma^2 \le \frac{1}{20}\sum_{i=1}^{20}\left(X_i - \mu\right)^2 \le 1.7085\sigma^2\right)$$

解:

故 $$⑴\ P\left(0.38165\sigma^2 \le \frac{1}{20}\sum_{i=1}^{20}\left(X_i - \overline{X}\right)^2 \le 1.80955\sigma^2\right)$$

$$= P\left(7.633 \le \frac{1}{\sigma^2}\sum_{i=1}^{20}\left(X_i - \overline{X}\right)^2 \le 36.191\right)$$

$$= P\left(\frac{1}{\sigma^2}\sum_{i=1}^{20}\left(X_i - \overline{X}\right)^2 \ge 7.633\right) - P\left(\frac{1}{\sigma^2}\sum_{i=1}^{20}\left(X_i - \overline{X}\right)^2 \ge 36.191\right)$$

$$\overset{\text{查表}}{=} 0.99 - 0.01 = 0.98$$

附： 几种重要随机变量的数学期望和方差

- 一. 二点分布
- 二. 二项分布
- 三. 泊松分布
- 四. 均匀分布
- 五. 正态分布
- 六. 指数分布

一. 二点分布

若随机变量X服从二点分布，其分布律为：

X	0	1
P_k	1-p	p

$$E(X)=p \qquad E(X^2)=p \qquad D(X)=p(1-p)$$

二. 二项分布

随机变量 $X \sim B(n,p)$, 其分布律为:

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

由二项分布定义可知，X 是 n 重贝努利试验中事件 A 发生的次数，且在每次试验中 A 发生的概率为 p，设

$$X_k=\begin{cases}1 & A\text{在第}k\text{次发生}\\ 0 & A\text{在第}k\text{次不发生}\end{cases}, k=1,2,\cdots,n$$

则 X_k 服从二点分布，其分布律为:

X	0	1
P_k	$1-p$	p

$$E(X_k)=p, \quad D(X_k)=p(1-p)$$

$$X = X_1 + X_2 + \cdots + X_n$$

$$E(X) = E(X_1) + E(X_2) + \cdots + E(X_n) = np$$

$$D(X) = D(X_1) + D(X_2) + \cdots + D(X_n) = np(1-p)$$

即：

若随机变量X~B(n , p),则

$$E(X) = np, \quad D(X) = np(1-p)$$

三. 泊松分布

随机变量$X \sim \pi(\lambda)$，其分布律为：

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

$$E(X)=\sum_{k=0}^{\infty} k\frac{\lambda^k e^{-\lambda}}{k!} = \lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}$$

$$=\lambda e^{-\lambda} e^{\lambda} \quad = \lambda$$

$$E(X^2) = E[X(X-1)+X] = E[X(X-1)]+E(X)$$

$$= \sum_{k=0}^{\infty} k(k-1)\frac{\lambda^k e^{-\lambda}}{k!} + \lambda = \lambda^2 e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!} + \lambda$$

$$= \lambda^2 e^{-\lambda} e^{\lambda} + \lambda = \lambda^2 + \lambda$$

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

即： **若随机变量$X \sim \pi(\lambda)$, 则**

$$\boxed{E(X) = \lambda, D(X) = \lambda}$$

四. 均匀分布

设随机变量 X 在区间 (a,b) 上服从均匀分布，

其概率密度为

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

$$E(X)=\int_{-\infty}^{+\infty}xf(x)dx=\int_a^b x\frac{1}{b-a}dx=\frac{1}{b-a}\cdot\frac{x^2}{2}\Big|_a^b=\frac{a+b}{2}$$

$$E(X^2)=\int_a^b x^2\frac{1}{b-a}dx=\frac{b^3-a^3}{3(b-a)}=\frac{a^2+ab+b^2}{3}$$

$$D(X)=E(X^2)-[E(X)]^2=\frac{a^2+ab+b^2}{3}-\frac{a^2+2ab+b^2}{4}=\frac{(b-a)^2}{12}$$

即 **若随机变量 $X\sim U(a,b)$，则**

$$E(X)=\frac{a+b}{2},D(X)=\frac{(b-a)^2}{12}$$

五. 正态分布

随机变量 $X \sim N(\mu,\sigma^2)$，其概率密度为：

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

$$E(X)=\int_{-\infty}^{+\infty}xf(x)dx=\int_{-\infty}^{+\infty}x\cdot\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx \qquad (\text{令 } t=\frac{x-\mu}{\sigma})$$

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

$$=\mu$$

$$D(X)=E\{[X-E(X)]^2\}=\int_{-\infty}^{+\infty}(x-\mu)^2\cdot\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$$

$$=\frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}t^2e^{-\frac{t^2}{2}}dt \quad (\text{令}\ t=\frac{x-\mu}{\sigma})$$

$$=\frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}(-t)de^{-\frac{t^2}{2}}=\frac{\sigma^2}{\sqrt{2\pi}}\left[-te^{-\frac{t^2}{2}}\Big|_{-\infty}^{+\infty}+\int_{-\infty}^{+\infty}e^{-\frac{t^2}{2}}dt\right]$$

$$=\frac{\sigma^2}{\sqrt{2\pi}}\cdot\sqrt{2\pi}=\sigma^2$$

即 $\boxed{\text{若随机变量}X\sim N(\mu,\sigma^2)\text{，则}}$

$$\boxed{E(X)=\mu, D(X)=\sigma^2}$$

六. 指数分布

随机变量X服从参数为λ的指数分布，其概率密度为：

$$f(x)=\begin{cases}\dfrac{1}{\theta}e^{-\frac{x}{\theta}} & x>0\\ 0 & x\le 0\end{cases}$$

$$E(X)=\int_{-\infty}^{+\infty}xf(x)dx=\int_{0}^{+\infty}x\cdot\frac{1}{\theta}e^{-\frac{x}{\theta}}dx=\int_{0}^{+\infty}(-x)\ de^{-\frac{x}{\theta}}$$

$$=(-x)\ e^{-\frac{x}{\theta}}\Big|_{0}^{+\infty}+\int_{0}^{+\infty}e^{-\frac{x}{\theta}}dx=-\theta e^{-\frac{x}{\theta}}\Big|_{0}^{+\infty}=\theta$$

$$E(X^2)=\int_{-\infty}^{+\infty}x^2 f(x)dx=\int_0^{+\infty}x^2\cdot\frac{1}{\theta}e^{-\frac{x}{\theta}}dx$$

$$=\int_0^{+\infty}(-x^2)\ de^{-\frac{x}{\theta}}=(-x^2)\ e^{-\frac{x}{\theta}}\Big|_0^{+\infty}+\int_0^{+\infty}2xe^{-\frac{x}{\theta}}dx$$

$$=\int_0^{+\infty}(-2\theta x)\ de^{-\frac{x}{\theta}}=(-2\theta x)\ e^{-\frac{x}{\theta}}\Big|_0^{+\infty}+2\theta\int_0^{+\infty}e^{-\frac{x}{\theta}}dx$$

$$=-2\theta^2 e^{-\frac{x}{\theta}}\Big|_0^{+\infty}=2\theta^2,\quad D(X)=E(X^2)-[E(X)]^2=\theta^2$$

即

若随机变量X服从参数为λ的指数分布，则

$$E(X)=\theta,\ D(X)=\theta^2$$

例1.已知 $X \sim \pi(3), Y = 2X - 1$, **求** $E(Y), D(Y), E[3(X^2 - 1)]$

解： $X \sim \pi(3)$, $E(X) = 3$, $D(X) = 3$ **则**

$$E(Y) = 2E(X) - 1 = 5$$

$$D(Y) = 4D(X) = 12$$

$$E[3(X^2 - 1)] = 3E(X^2) - 3$$

$$= 3\{D(X) + [E(X)]^2\} - 3 = 33$$

例2. 已知 X 和 Y 相互独立，且X在区间(1，5)上服从均匀分布，$Y \sim N(1,9)$， 求(1) (X,Y)的概率密度;

(2) $E(3X-4Y-2)$，$D(3X-4Y-2)$

解： X 在区间(1，5)上服从均匀分布, $Y \sim N(1,9)$,

$$f_X(x)=\begin{cases}\frac{1}{4} & 1<x<5 \\ 0 & \text{其它}\end{cases},\qquad E(X)=\frac{1+5}{2}=3$$

$$D(X)=\frac{(5-1)^2}{12}=\frac{4}{3}$$

$$f_Y(y)=\frac{1}{3\sqrt{2\pi}}e^{-\frac{(y-1)^2}{18}},\quad -\infty<y<+\infty,$$

$$E(Y)=1,\qquad D(Y)=9$$

由 X 和 Y 相互独立得:

$$f(x,y)=f_X(x)\cdot f_Y(y)$$

$$=\begin{cases}\dfrac{1}{12\sqrt{2\pi}}e^{-\frac{(y-1)^2}{18}}, & 1<x<5,-\infty<y<+\infty\\ 0 & 其它\end{cases}$$

$$E(3X-4Y-2)=3E(X)-4E(Y)-2=3$$

$$D(3X-4Y-2)=9D(X)+(-4)^2D(Y)=156$$