概率论

两类现象

- 自然界与社会生活中的两类现象

确定性现象——结果确定

不确定性现象——结果不确定

随机现象

- 在个别试验中其结果呈现出不确定性，在大量重复试验中其结果又具有统计规律性的现象。

随机试验

- 对随机现象的观察、记录、实验统称为随机试验。
- 它具有以下特性：

• 可以在相同条件下重复进行；

• 每次试验的可能结果不止一个，事先明确试验的所有可能出现的结果；

• 进行试验前不能确定哪个结果会发生。

概率

定义：记 $f_n(A)=\dfrac{n_A}{n}$；

正面朝上

其中 $\boldsymbol{n_A}$ — A发生的次数(频数)；

$\boldsymbol{n}$—总试验次数。

抛硬币

称 $f_n(A)$ 为A在这n次试验中发生的频率。

当重复试验的次数逐渐增大时，频率呈现出稳定性，逐渐稳定于某个常数

$f_n(A)$ 的稳定值p定义为**A**的概率，记为P(A)=p

随机变量

- 投掷一枚硬币，观察出现正反面的情形。试验有两个可能结果——正面**A**、反面**B**

随机变量

入一个变量：

$$X = X(\omega) = \begin{cases} 1, & \omega = A \\ 0, & \omega = B \end{cases}$$

设随机试验为E，其样本空间为**Ω={ω}**，如果对于每个**ω**∈**Ω**，都有一个实数**X(ω)**和它对应，于是就得到一个定义在**Ω**上的实值单值函数**X(ω)**，称**X(ω)**为<u>随机变量</u>。

随机变量的分布函数

定义：设X是一个随机变量，x是任意实数，函数

$F(x)=P\{X\le x\}$ 称为X的**分布函数**。

$$P\{x_1<X\le x_2\}=P\{X\le x_2\}-P\{X\le x_1\}$$
$$=F(x_2)-F(x_1).$$

如果将 X 看作数轴上随机点的坐标，那么分布函数 $F(x)$ 的值就表示 X落在区间 $(-\infty,x]$ 内的概率.

离散随机变量的分布律和分布函数

- 若随机变量**X**的全部可能取值是有限个或可列无限多个，则称这种随机变量为离散型随机变量。

设X为离散型随机变量，则X的所有可能取值x_k $(k=1,2,...,)$，及取各个可能值的概率$P\{X=x_k\}=p_k$ $(k=1,2,...,)$，称为离散型X的分布律.

p_k满足两个条件: a. $p_k \ge 0$ $(k=1,2,...)$ b. $\sum_{k=1}^{\infty} p_k = 1$

离散型随机变量的分布函数

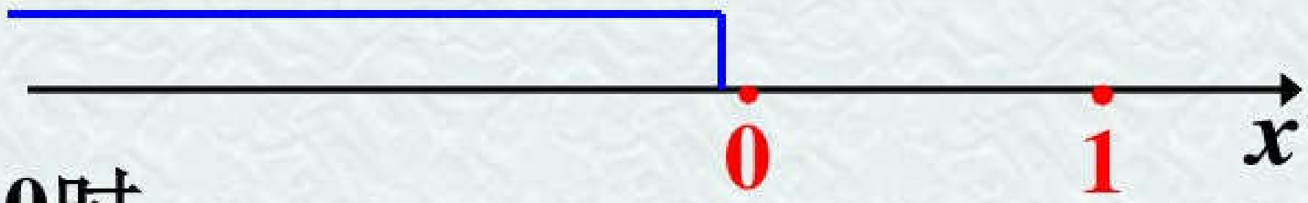
$$F(x)=P\{X\le x\}=\sum_{x_k\le x} p_k = \sum_{x_k\le x} P(X=x_k).$$

例 1 抛掷均匀硬币, 令

$$X=\begin{cases}1, & \text{出正面},\\ 0, & \text{出反面}.\end{cases}$$

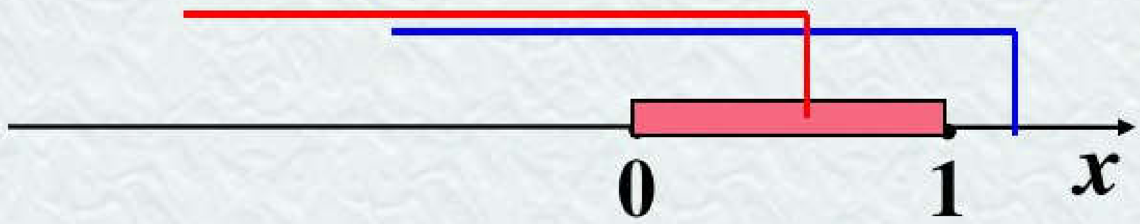
求随机变量 X 的分布函数.

解 $p\{X=1\}=p\{X=0\}=\dfrac{1}{2}$,



当$x<0$时,

$$F(x)=P\{X\le x<0\}=P(\phi)=0$$



当$0 \le x < 1$时,

$$F(x) = P\{X \le x\} = P\{X = 0\} = \frac{1}{2};$$

当$x \ge 1$时,

$$F(x) = P\{X \le x\}$$

$$= P\{X = 0\} + P\{X = 1\}$$

$$= \frac{1}{2} + \frac{1}{2} = 1.$$

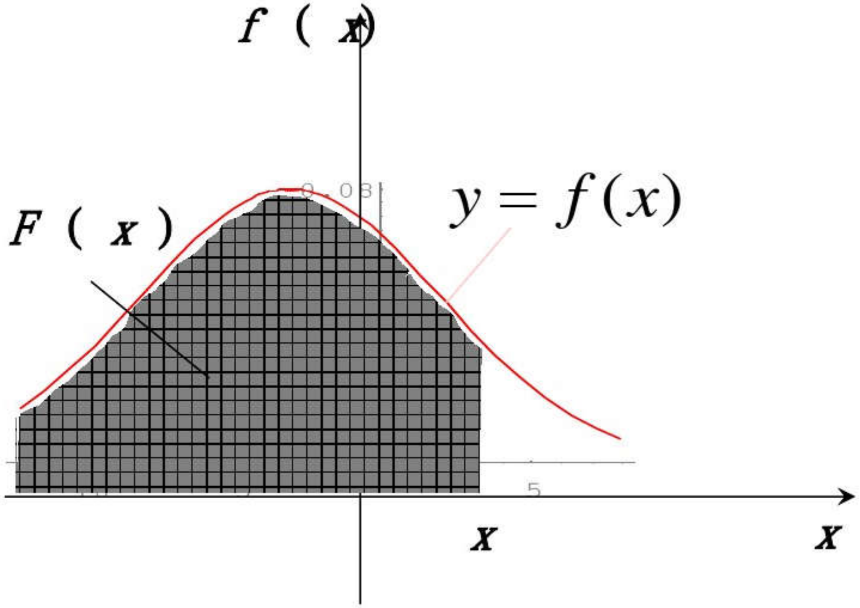
得 $F(x) = \begin{cases} 0, & x < 0, \\ \frac{1}{2}, & 0 \le x < 1, \\ 1, & x \ge 1. \end{cases}$

连续随机变量的概率密度函数

如果对于随机变量X的分布函数$F(x)$，存在非负函数$f(x)$，使对于任意实数x有

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

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

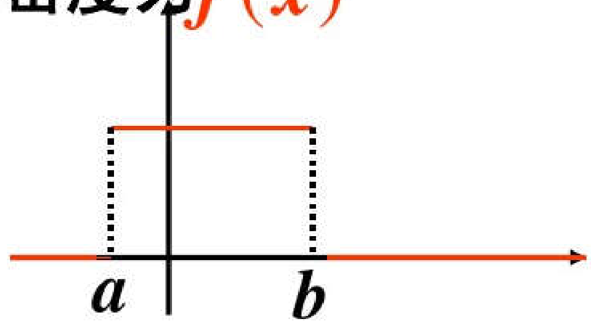


常用的连续型随机变量的分布

（一）均匀分布

如果连续型随机变量 X 的概率密度为

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



则称连续型随机变量X在区间 (a,b) 上服从**均匀分布**。

记为 $X \sim U(a,b)$.

显然有 $f(x)\geq 0$, 且 $\int_{-\infty}^{\infty} f(x)dx=1$.

（二） 指数分布

如果连续型随机变量 X 的概率密度为

$$f(x)=\begin{cases}\dfrac{1}{\theta}e^{-\frac{x}{\theta}} & x>0\\ 0 & \text{其它}\end{cases}$$ 其中 $\theta>0$ 为常数，

则称连续型随机变量 X 服从**指数分布**。

显然有 $f(x)\geq 0$，且 $\int_{-\infty}^{\infty} f(x)dx=1.$

（三） 正态分布

如果连续型随机变量 X 的概率密度为

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

其中 $\mu,\sigma(\sigma>0)$ 为常数，
则称连续型随机变量 X 服从参数为 μ,σ 的**正态分布**
或**高斯(Gauss)分布**，记为 $X\sim N(\mu,\sigma^2)$.

显然有 $f(x)\geq 0$，且 $\int_{-\infty}^{\infty}f(x)dx=1.$

特别，当 $\mu=0,\ \sigma=1$ 时称 X 服从 **标准正态分布**。

数学期望

定义　设离散型随机变量 X的分布律为

$$P\{X = x_k\} = p_k \qquad k = 1,2,......$$

若级数$\sum_{k=1}^{\infty} x_k p_k$绝对收敛，则称级数 $\sum_{k=1}^{\infty} x_k p_k$

为随机变量X的数学期望，简称期望，记为E(X)，即

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

设连续型随机变量X具有概率密度$f(x)$,

若$\int_{-\infty}^{+\infty} x f(x) dx$绝对收敛,则称积分$\int_{-\infty}^{+\infty} x f(x) dx$为$X$的

数学期望，记为$E(X)$，即 $E(X) = \int_{-\infty}^{+\infty} x f(x) dx$

随机变量函数的期望

离散型随机变量函数的数学期望为

若 $Y = g(X)$, 且 $P\{X = x_k\} = p_k$, $(k = 1, 2, \cdots)$

则有 $$E(g(X)) = \sum_{k=1}^{\infty} g(x_k) p_k.$$

若 X 是连续型的, 它的分布密度为 $f(x)$,

则有 $$E(g(X)) = \int_{-\infty}^{+\infty} g(x) f(x) \mathrm{d}x.$$

期望的性质

1. 设C是常数，则有$E(C)=C$.

2. 设X是一个随机变量，C是常数，则有

$$E(CX)=CE(X).$$

3. 设X,Y是两个随机变量，则有

$$E(X+Y)=E(X)+E(Y).$$

4. 设X,Y是相互独立的随机变量，则有

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

方差

定义

设 X 是一个随机变量，若 $E\{[X-E(X)]^2\}$ 存在，则称 $E\{[X-E(X)]^2\}$ 是 X 的方差，记作

$$D(X) \quad 或 \quad \mathrm{Var}(X),$$

即 $$D(X)=\mathrm{Var}(X)=E\{[X-E(X)]^2\},$$

称 $\sqrt{D(X)}$ 为标准差或均方差，记为 $\sigma(X)$.

常用计算公式：

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

方差的性质

1. 设 C 是常数，则有 $D(C)=0.$

2. 设 X 是一个随机变量, C 是常数，则有

$$D(CX)=C^2D(X).$$

3. 设 X,Y 相互独立, $D(X)$, $D(Y)$ 存在,则

$$D(X\pm Y)=D(X)+D(Y).$$

4. $D(X)=0$的充要条件是 X 以概率 1 取常数 C,即

$$P\{X=C\}=1.$$

正态分布： 随机变量 $X \sim N\left(\mu, \sigma^2\right)$，其密度函数为

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

则数学期望为

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

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

$$=\frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} ye^{-\frac{y^2}{2}}dy+\frac{\mu}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}}dy$$

$$=0+\frac{\mu}{\sqrt{2\pi}}\sqrt{2\pi}=\mu.$$

正态分布： 随机变量 $X \sim N\left(\mu, \sigma^2\right)$ ，其密度函数为

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

方差为

$$D(X)=\int_{-\infty}^{+\infty}\left(x-EX\right)^2 f(x)dx$$

$$=\int_{-\infty}^{+\infty}\left(x-\mu\right)^2\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx \quad （令 \ y=\frac{x-\mu}{\sigma} ）$$

$$=\int_{-\infty}^{+\infty}\frac{\sigma^2}{\sqrt{2\pi}}y^2e^{-\frac{y^2}{2}}dy=-\sigma^2\int_{-\infty}^{+\infty}\frac{y}{\sqrt{2\pi}}d\left(e^{-\frac{y^2}{2}}\right)$$

$$=-\frac{\sigma^2}{\sqrt{2\pi}}ye^{-\frac{y^2}{2}}\Bigg|_{-\infty}^{+\infty}+\sigma^2\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy$$

$$=\sigma^2.$$

正态分布常用性质

$$若\ X \sim N(\mu,\sigma^2),\ 则\ Z = \frac{X-\mu}{\sigma} \sim N(0,1).$$

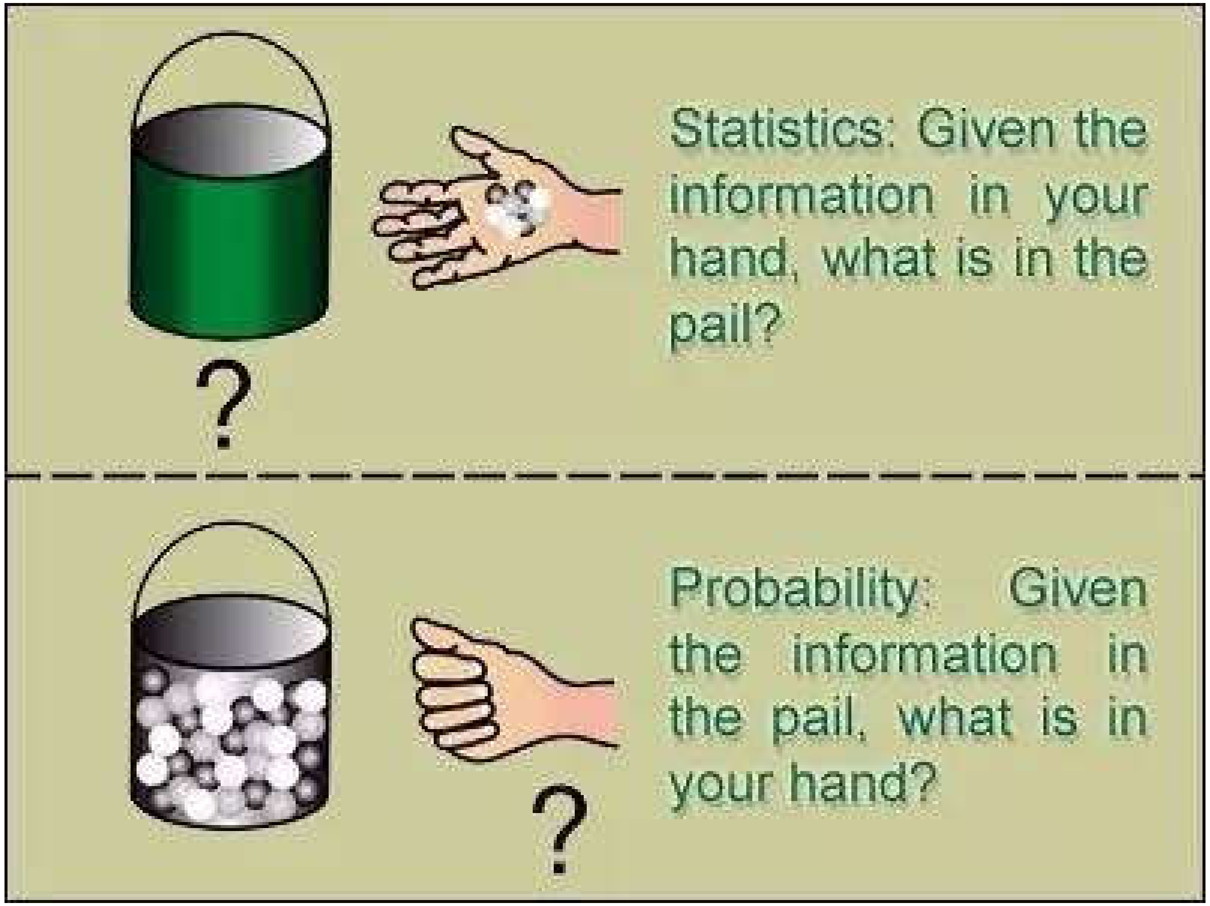
一般,设X,Y相互独立且$X \sim N(\mu_1,\sigma_1^2),Y \sim N(\mu_2,\sigma_2^2)$.则$Z = X+Y$仍然服从正态分布,且有$Z \sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$.

有限个相互独立的正态随机变量的线性组合仍然服从正态分布,即正态分布具有可加性.

数理统计

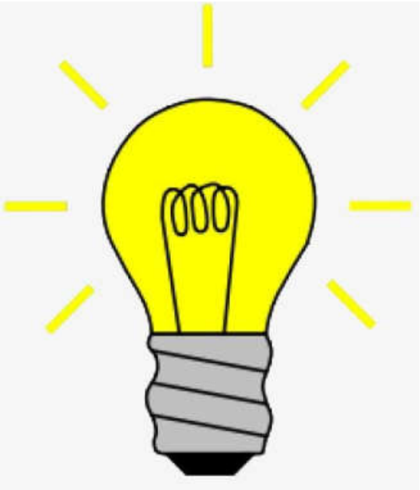
- 数理统计学 是一门以数据为基础的科学, 可以定义为收集数据, 分析数据和由数据得出结论的一组概念、原则和方法。
- 数理统计就是通过对随机现象有限次的观测或试验所得数据进行归纳，找出这有限数据的内在数量规律性，并据此对整体相应现象的数量规律性做出推断或判断的一门学科。

统计学VS概率论



?
Statistics: Given the information in your hand, what is in the pail?
?
Probability: Given the information in the pail, what is in your hand?

- 例：若规定灯泡寿命低于1000小时者为次品，如何确定次品率？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，这是数理统计学研究的问题之一。



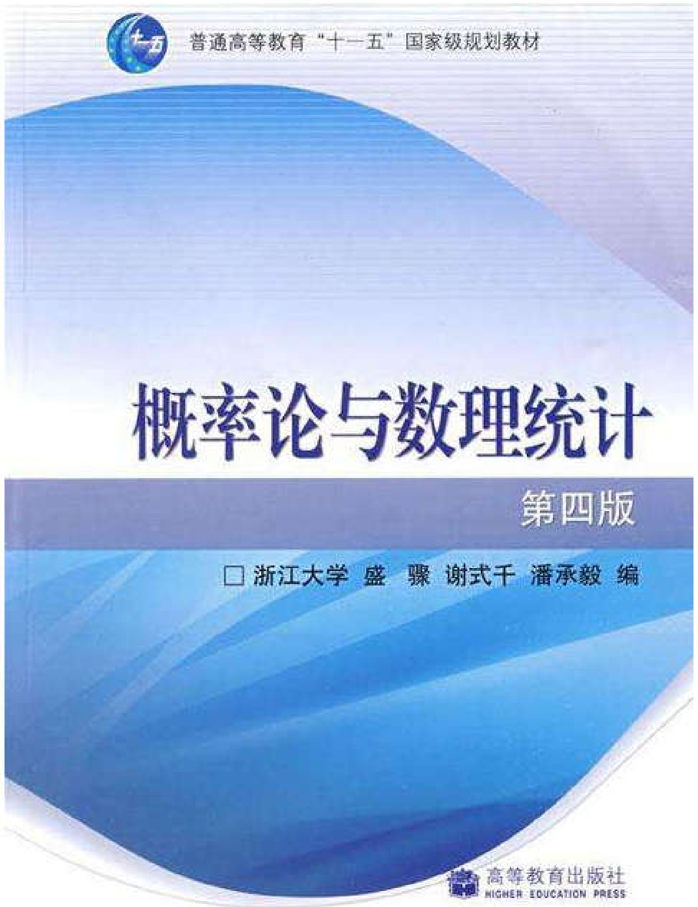
例： 设某种清漆的9个样品，其干燥时间（以小时计）分别为：

6.0 5.7 5.5 6.5 7.0 5.8 5.2 6.1 5.0

根据以往经验，干燥时间的总体服从正态分布N(6.0, 0.36),现根据样本检验均值是否与以往有显著差异？

课程内容

- 样本及抽样分布
- 参数估计
- 假设检验
- 方差分析及回归分析



考核方式

- 课堂练习
- 作业（3-4次）
- 期末考试（概率论50%+数理统计50%）