Chpt.2 Random Variables & Probability Distributions

第二章 随机变量及其分布

Review

独立事件：A_1，A_2，…，A_n，其任意多个事件的积事件的概率都等于各个事件概率的积，则A_1，A_2，…，A_n相互独立。

- 独立、互斥往往是根据实际意义去判断
- 利用独立的特性，把和事件的概率转化为对立事件--积事件的概率加以计算

全概率公式：设S为随机试验E的样本空间，$B_1, B_2, ---, B_n$为S的一个划分，且 $P(B_i)>0, i=1,2, \cdots, n$。则任意事件A的概率为

$$P(A)=P(A|B_1)P(B_1)+P(A|B_2)P(B_2)+\cdots+P(A|B_n)P(B_n)$$

Review

[Bayes定理]：$B_1, B_2, \cdots, B_n$为样本空间S的一个划分，事件A 的概率$P(A)>0, P(B_i)>0 (i=1,2,\cdots n)$，则：

$$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{\sum_{j=1}^{n} P(A \mid B_j)P(B_j)} \qquad i = 1,2,\cdots,n$$

- Bayes定理是在观察到事件A已发生的条件下，寻找导致A发生的每个原因的概率. 所以也称为逆概率公式。
- $P(B_i)$和$P(B_i \mid A)$分别称为先验概率和后验概率.

基于Bayes理论的决策、判别

如何确定一个新开信用卡客户的信誉?

基本思路： 根据过去存在的信用卡客户的基本信息与实际信誉建立一个模型，而后判断新开客户的信誉。属于模式分类问题，采取的是机器学习思想（Machine Learning）。

假设已有的大量信用卡客户记录（包含客户基本信息、客户类别标签）构成一个样本集S，共把客户分成了M类（如金牌、银牌、铜牌、一般、黑名单），记为 $S=\{c_1, \dots, c_i, \dots, c_M\}$，样本集S被划分为$c_1$、$c_2$、 …、$c_M$。每类的概率（先验）为$P(c_i)$，$i=1, 2, \dots, M$。

现在的问题是，对一个没有标示类别的样本Y（新客户信息），判断其属于哪一类?

基于Bayes理论的决策、判别

当样本集非常大时，可以认为

$$P(c_i)=c_i\text{类样本数}/\text{总样本数}$$

对于一个待分类样本Y，根据Bayes定理，可得其属于c_i类的后验概率$P(c_i/Y)$:

$$P(c_i/Y)=P(Y/c_i)\cdot P(c_i)/P(Y)$$

则可以看$P(c_1/Y)$、 $P(c_2/Y)$ 、……、$P(c_M/Y)$ 哪个大，就断定Y属于哪一类。或者说$X\in c_K$

$$P(c_K/Y)=\underset{i=1,2,\cdots}{Max}P(c_i/Y)$$

$$=\underset{i=1,2,\cdots}{Max}P(Y/c_i)P(c_i)\Big/P(Y)$$

事实上只要计算 $\underset{i=1,2,\cdots}{Max}P(Y/c_i)P(c_i)$ 就可。

式(2)是最大后验概率判决准则.

基于Bayes理论的决策、判别

注意到 $P(Y/c_i)P(c_i)$ 的计算中包含了 $P(Y/c_i)$

假设属性集合为x_1、x_2、…、x_n，那么在已经得到的数据集S中，我们可以统计得到

$$P(x_1/c_i)\cdots\cdots P(x_n/c_i) \qquad (i=1,2,\cdots,n)$$

假设属性之间是独立的，则知道

$$P(X/c_i)=P(x_1/c_i)\cdots P(x_n/c_i)$$

我们可以用原有集合中的x_1、x_2、……、x_n伴随每个c_i出现的概率

$P(X/c_i)$ 来估计 $P(Y/c_i)$

概率研究中的Case Study方法是否能解决所有问题？

前已介绍一些概率模型、很多例子，都有代表性；

但是可以找出更多的例子进行研究，永无止境！

这种方法属于Case Study。

存在问题：不同的case之间有无本质、统一的东西？

进一步，不同的概率现象之间有什么样的联系？

2.1 Random Variables

2.1.1 Introduction

那么，概率函数是否可用?

概率函数定义在样本空间S上，$\forall A \in S$有$P(A)$与之对应；但由于样本空间的多样性，导致样本描述的多样性，使抽取其共性变得困难。

对于数值函数我们已经有足够多的研究成果

自然希望把一般的概率问题 ⟶ 数值函数问题。

事件空间 ⟷ 实数集合

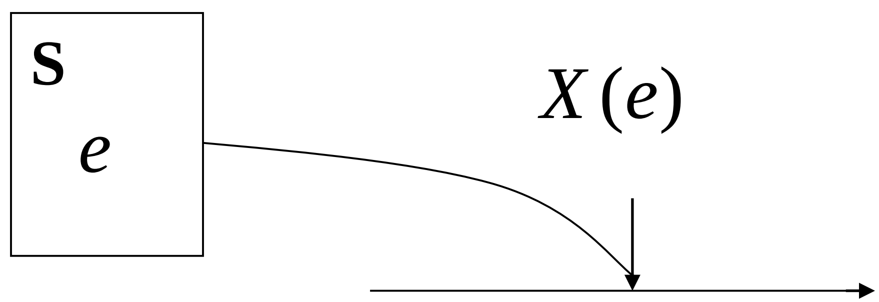
研究事件 ⟶ 研究数量

2.1.2 随机变量的定义

随机试验E，其样本空间为S＝{e}，如果对每个e∈S ，有一个实数X(e)与之对应，就得到一个定义在S上的单值实函数X＝X(e)，它是一个因变量，且因e的随机性而具有随机性，称为随机变量。

$$S \xrightarrow{X(.)} R_X$$

$$X\text{：}\ e \in S \to X(e) \in R_X$$



2.1 Random Variables

通过随机变量可以描述事件，随机变量本身取值有一定的概率，区别于其他普通函数。对于一个实数集合L，随机变量X，$\{X \in L\}$表示的是一个事件A：

$$\{X \in L\} \Leftrightarrow \{e \mid X(e) \in L\} = A$$

因此，

$$P\{X \in L\} = P\{e \mid X(e) \in L\}$$

Example 2.1 某射手每次射击命中率为0.02，现在不断地射击，直到命中目标为止。

可以定义" X＝命中目标所需的射击次数"

$$R_x = \{1,2,\cdots\}$$

$$L_1 = \{100\}$$

$$\{X \in L_1\} \Leftrightarrow \{e \mid \text{射击到第100次命中}\}$$

$$L_2 = \{1,2,\cdots 100\}$$

$$\{X \in L_2\} \Leftrightarrow \{e \mid \text{不多于100次命中目标}\}$$

有些试验结果本身就是数量描述的，构成了随机变量。

例如：高炮在一定的条件下射击，用ρ表示弹着点与目标之间的距离，ρ就是随机变量；如以目标点为原点，建立直角坐标系，则弹着点（x,y）中的x、y均是随机变量，(x,y)本身是一个二维随机变量。

2.1 Random Variables

Example 2.2 盒子中有N个白球、M个黑球，从中抽取k个(k≤M+N)，则以下量为随机变量

X＝取得的白球数

Y＝取得的黑球数

Z＝两种球个数的差

Example 2.3 某公共汽车站每5分钟有一辆汽车通过，如果某人到达该车站的时刻是随机的。那么，一般来说，他等车的时间X是一个随机变量。

注意：此处随机变量X的取值是一个连续区间[0，5]

2.2.1 离散型随机变量的定义

若随机变量X可能取的值为有限个或可列无限个， 则称X为离散型随机变量(discrete random variable)。

Example 2.4

[1] 上面例子中的射击命中次数X；

[2] 在[0，1]区间上方随机抛球，观察落到有理点X上的情形。

有理点为：

0, 1

1/2

1/3, 2/3

1/4, 2/4, 3/4

………

2.2.2 概率分布律

对离散型随机变量涉及到两点：[1] 随机变量所有可能的取值；[2] 取每个值得概率。

如果X所有可能的取值为x_k (k=1,2,…)， X取 x_k (即事件 A={x=x_k}) 的概率记为

$$P\{x = x_k\} = p_k (k = 1,2,\cdots)$$

称之为离散随机变量X的概率分布或分布律。

其直观的表示是列表

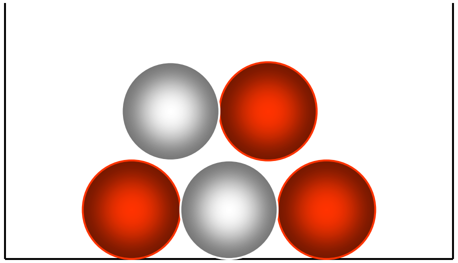
$$\begin{pmatrix} x_1 & x_2 & \cdots & x_n & \cdots \\ p(x_1) & p(x_2) & \cdots & p(x_n) & \cdots \end{pmatrix}$$

Example 2.5 如右图所示，从盒中任取3个球。取到的白球数X是一个随机变量。X可能取的值是0,1,2。取每个值的概率为:

$$P(X=0)=\frac{C_3^3}{C_5^3}=\frac{1}{10}$$

$$P(X=1)=\frac{C_3^2C_2^1}{C_5^3}=\frac{6}{10}$$

$$P(X=2)=\frac{C_3^1C_2^2}{C_5^3}=\frac{3}{10}$$



其分布列为

X	0	1	2
p_i	0.1	0.6	0.3

Example 2.6 设随机变量X的分布列为

$$\begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ \dfrac{a-1}{4} & \dfrac{a+1}{4} & 0.1 & 0.2 & 0.1 \end{pmatrix}$$

求$P\{-1 \le X \le 2\}$

解 (1) 由 $$\frac{a-1}{4}+\frac{a+1}{4}+0.1+0.2+0.1=1$$

解得 $a=1.2$

故分布列为 $$\begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ 0.05 & 0.55 & 0.1 & 0.2 & 0.1 \end{pmatrix}$$

(2) $$P\{-1 \le X \le 2\} = \sum_{-1 \le X \le 2} p(x_i)$$

$$= 0.55 + 0.1 + 0.2 + 0.1$$

$$= 0.95$$

Example 巴斯卡分布 在伯努里试验中，每次成功的概率为p。记直至得到第r次成功时的试验次数为X，求X的分布列。

解　P{X＝k}＝P{前k-1次试验中有r-1次成功，有k-r次不成功，且第k次成功}

$$= C_{k-1}^{r-1} p^{r-1}(1-p)^{k-r}\, p$$

$$= C_{k-1}^{r-1} p^{r}(1-p)^{k-r}$$

$$(k = r, r+1, r+2, \cdots\cdots)$$

2.2.3 常见的概率分布

（一）退化分布（单点分布）

设随机变量X只取一个常数值c， 即

$$P(X=c)=1$$

称它为退化(degenerate)分布，又称为单点分布。

事实上，{X＝c}是一概率为1的事件，X可以看作一个常数，但有时我们宁愿把它看作（退化的）随机变量。

（二）0-1分布（两点分布/Bernoulli分布）

随机变量X只取两个值0和1，满足分布：

$$P(X=0)=1-p \qquad p>0$$

$$P(X=1)=p$$

称X服从参数为p的 0-1分布（也称伯努里分布）。

2.2.3 常见的概率分布

Example 2.7 2000件产品，有1990件合格品，10 件次品。从中随机抽一件，规定e_1＝{取得的是合格品}，e_2＝{取得的是次品}

定义随机变量 $X(e_1)=0, X(e_2)=1$

$$p=10/2000=0.05$$

X服从参数0.05的0-1分布，如定义。

[1] 伯努里试验具有广泛性： 电路‘断’与‘不断’，产品‘合格’与‘不合格’， 种子‘发芽’与‘不发芽’，掷硬币得‘正面’ 与‘反面’，…

[2] 任一伯努里试验（具有广泛性）的结果都可用伯努里分布描述

（三）二项分布

对于n重伯努里（Bernoulli）试验，假设每次成功的概率是p，定义随机变量X描述n次试验中事件A可能发生的次数k

$$P\{X=k\}=b(k,n,p)$$

$$=C_n^k p^k(1-p)^{n-k} \qquad (k=0,1,2,\cdots n)$$

称X服从参数为n和p的二项分布(Binomial distribution)，记为X~b(n,p)或X~B(n,p).

（三）二项分布

二项分布是概率论中最重要的分布之一，应用很广

[在此，进一步体会把一般概率问题抽象为一个随机变量，用分布加以描述的意义]

(1)考察某地n个人是否患某种非流行性疾病，患病人数X服从二项分布. 检查一人是否患某种非流行性疾病是一次伯努里试验，各人是否生这病可认为相互独立，并可近似认为患病的概率p相等。

(2) 保险公司对某种灾害 (自行车被盗，火灾，...) 保险 各人发生此种灾害与否可认为相互独立，并假定概率相等。设一年间一人发生此种灾害的概率为p， 则在参加此种保险的n人中发生此种灾害的人数X服从二项分布。

（三）二项分布

(3) 碰运气能否通过英语四级考试

$$P\{X=k\}=b(k,85,0.25)$$

$$P\{X\geq 51\}=b(51,85,0.25)+b(52,85,0.25)+\cdots+b(85,85,0.25)$$

$$\approx 8.74\times 10^{-12}$$

(4) 机房内n台同类计算机，在一年内每台损坏的概率为p，则在一年时间内损坏的机器数X服从二项分布。

如此，我们要问这样的分布具有什么样的性质呢？

比如：我们看看一年内最多会坏多少台机器？

为保证工作，最多备多少台机器？

平均会坏多少？一般备多少台机器（平均）？

二项分布的重要性质

[1] $b(k,n,p)=b(n-k,n,1-p)$

这从二项分布的概率以及 $C_n^k = C_n^{n-k}$ 立即可得。

也可以这样理解：n次试验中，事件 {k次成功} 与事件 {n-k次不成功} 是同样的，而 {n-k次不成功} 的概率即为

$b(n-k,n,1-p)$

很多情况下二项分布的计算很复杂，有时备有相应的计算表格，但只限于$p \leq 0.5$的情况，当$p>0.5$时就可利用上式来计算。

[2] 增减性以及最可能成功（发生）次数

对固定的n、p，由于

$$\frac{b(k,n,p)}{b(k-1,n,p)}=\frac{C_n^k p^k(1-p)^{n-k}}{C_n^{k-1}p^{k-1}(1-p)^{n-k+1}}$$

$$=\frac{(n-k+1)p}{k(1-p)}$$

$$=1+\frac{(n+1)p-k}{k(1-p)}$$

上式是否大于1，主要看(n+1)p−k的正负，或者说看(n+1)p与k的大小问题。

当 $k<(n+1)p$ 时，$\dfrac{b(k,n,p)}{b(k-1,n,p)}>1$，$b(k,n,p)$单调递增；

当 $k>(n+1)p$ 时，$\dfrac{b(k,n,p)}{b(k-1,n,p)}<1$，$b(k,n,p)$单调递减；

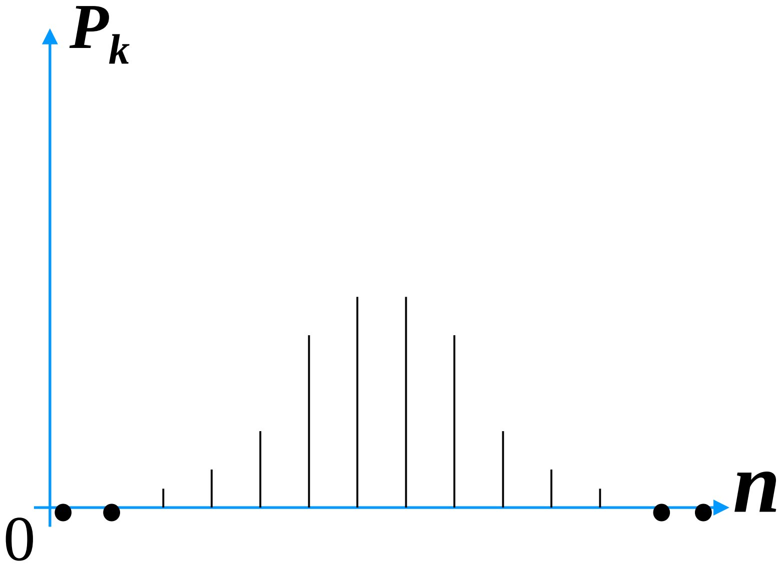
[2] 增减性以及最可能成功（发生）次数

(n+1)p是整数且k=(n+1)p时, $b(k,n,p)=b(k-1,n,p)$ 达最大值;

我们称m=(n+1)p或(n+1)p−1为最可能成功次数;

当(n+1)p不是整数时，最可能成功次数 $m=[(n+1)p]$

直观推断：概率p是统计得到的，现在做n次试验，最可能的成功次数应该是np附近。



n=13, p=0.5

Example 2.8 邮局开设多少服务窗口合理

某居民区有n人，拟设一个邮局，开设m个服务窗口，每个窗口都办理所有业务。窗口数m太小，则经常派长队；窗口数m太大不经济。

假定在每一个指定时刻，小区n个人中每人是否去邮局是独立的，每人去邮局办理业务的概率都是p。

问题：” 在营业中任意时刻保证每个窗口的排队人数（包括正在被服务的那个人）不超过s”这个事件的概率不小于α（一般取α＝0.80, 0.85, 0.90），则至少需开设多少个窗口？

分析：假设在邮局办事的人数为X，有k个人在邮局的概率为

$$P\{X=k\}=C_n^k p^k(1-p)^{n-k} \qquad k=0,1,2,\cdots,n$$

Example 2.8 邮局开设多少服务窗口合理

“在营业中任意时刻每个窗口的排队人数（包括正在被服务的那个人）不超过s” 这件事相当于m个窗口总的人数X：{ X≤sm } 这个事件。其发生的概率为

$$P\{X \le sm\} = \sum_{k=0}^{sm} P\{X = k\} = \sum_{k=0}^{sm} C_n^k p^k (1-p)^{n-k}$$

题目要求此概率大于α，即

$$\sum_{k=0}^{sm} C_n^k p^k (1-p)^{n-k} \ge \alpha$$

找一个自然数m，使得上述不等式成立，此m即为问题的答案。

[3] $n\to\infty$ 时的渐近性质

假定p与n有关，记作p_n。 考虑 n $\to\infty$的情况，有下面的定理：

[泊松(Poisson)定理] 如果存在正常数 λ，当 $n\to\infty$ 时有 $np_n\to\lambda$，则

$$\lim_{n\to\infty} b(k,n,p)=\frac{\lambda^k}{k!}e^{-\lambda} \qquad k=0,1,2,\cdots$$

[3] $n \to \infty$ 时的渐近性质

Proof: 记$\lambda_n = np_n$

$$b(k,n,p) = C_n^k (p_n)^k (1-p_n)^{n-k}$$

$$= \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}\left(\frac{\lambda_n}{n}\right)^k\left(1-\frac{\lambda_n}{n}\right)^{(n-k)}$$

$$= \frac{\lambda_n^k}{k!}\,\frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k}\left(1-\frac{\lambda_n}{n}\right)^n \Big/ \left(1-\frac{\lambda_n}{n}\right)^k$$

$$\downarrow \qquad \downarrow \qquad \downarrow \qquad \downarrow$$

$$\boxed{\frac{\lambda^k}{k!}} \qquad \boxed{1} \qquad \boxed{e^{-\lambda}} \qquad \boxed{1}$$

$$\to \frac{\lambda^k}{k!} e^{-\lambda} \qquad (n \to \infty)$$

[3] $n\to\infty$ 时的渐近性质

$$\lim_{n\to\infty} b(k,n,p)=\frac{\lambda^k}{k!}e^{-\lambda} \qquad k=0,1,2,\cdots$$

定理的条件 $np_n\to\lambda$ 意味着当 n很大时，p_n 必定很小.

因此，利用泊松定理，对于二项分布，当 n 很大，p 很小时有以下近似式：

$$C_n^k p^k (1-p)^{n-k} \approx \frac{\lambda^k e^{-\lambda}}{k!}$$

其中 $\lambda = np$

实际计算中，$n\geq 100$, $np\leq 10$ 时近似效果就很好.

Example 2.9 为保证设备正常工作，要配备适量的维修人员 .

设共有300台设备，每台的工作相互独立，发生故障的概率都是0.01. 在通常情况下，一台设备的故障可由一人来处理 .

问：至少应配备多少维修人员，才能保证当设备发生故障时不能及时维修的概率小于0.01?

分析：

设X为300台设备同时发生故障的台数，300台设备独立工作，每台出故障概率p=0.01 . 可看作n=300的伯努里概型.

可见， **X~B(n,p)，n＝300, p＝0.01**

设需配备N个维修人员, 所求问题就是满足$P(X>N) < 0.01$ 或 $P(X\leq N)\geq 0.99$ 的最小的N.

$$P\{X > N\} = \sum_{k=N+1}^{300} C_{300}^{k}(0.01)^k(0.99)^{300-k}$$

$$\approx \sum_{k=N+1}^{300} \frac{3^k e^{-3}}{k!}$$

$$\approx \sum_{k=N+1}^{\infty} \frac{3^k e^{-3}}{k!}$$

我们求满足 $\sum_{k=N+1}^{\infty} \frac{3^k e^{-3}}{k!} < 0.01$ 的最小的N.

查泊松分布表得

$$\sum_{k=9}^{\infty} \frac{e^{-3}3^k}{k!} \approx 0.0038, \qquad \sum_{k=8}^{\infty} \frac{e^{-3}3^k}{k!} \approx 0.012$$

$K = N+1 \geq 9$即可，即$N \geq 8$.　即至少需配备8个维修人员.

（四）泊松分布(Poisson)

从上面的泊松定理可引入另一类重要的分布。

设随机变量X可取一切非负整数值，取这些值的概率为：

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

其中λ>0是它的参数，称X服从参数为λ的泊松分布，简记 $X \sim \pi(\lambda)$

容易知道：

$$P\{X = k\} > 0 \qquad (k = 0,1,2\cdots)$$

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

泊松分布的背景

考虑在白天一分钟内某电话交换台接到的呼叫数X的分布，X可能取值范围为0,1,2...。那么，X服从何种分布？

我们采用分割-求和-取极限的方法来剖析。

[1] 首先把一分钟分割成n个足够小的区间（对等划分），假设每个区间只可能呼叫一次，对应的呼叫概率均为 p_n ，可以视n个小区间的呼叫为n重Bernoulli试验

[2] 在其中有k次呼叫的概率为 $b(k,n,p)=C_n^k(p_n)^k(1-p_n)^{n-k}$

假设在一分钟内的呼叫次数的统计平均数为λ，则 $np_n=\lambda$

[3] 取n足够的大，$n>>\lambda$ ，由Bernoulli定理，我们知道

$$b(k,n,p)\to\frac{\lambda^k}{k!}e^{-\lambda}\qquad(n\to\infty)$$

（四）泊松分布(Poisson)

Remark 1 这说明电话交换台在一分钟内呼叫的次数X满足参数为λ的泊松分布X~$\pi(\lambda)$。 λ的含义恰恰是在一分钟内的呼叫次数的统计平均数

Remark 2 这也说明泊松分布是n重Bernoulli分布的极限。当n足够大时, p较小，可以近似有

$$b(k,n,p) \sim \frac{\lambda^k}{k!}e^{-\lambda}$$

2.2.3 常见的概率分布

（五）几何分布

单次试验中事件A 发生概率为p，现重复多次，直到A出现为止。定义X为事件A首次发生时所进行的试验次数，则X为一随机变量，可能的取值为

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

称X服从参数为 p 的几何分布

几何分布的无记忆性

假定在前m次试验中，事件A一直没有发生，则从第m+1次开始直到成功的次数Y也服从同样的几何分布(实验次数Y与m无关, 好像把前面m次失败‘忘记’了)。

（证明见《概率论引论》，汪仁宫，pp.57）

2.3 Distribution Function of Random Variable

2.3.1 WHY and HOW

WHY：分布列不能表示很多的随机变量之分布，比如机器的寿命,无法一一罗列这些值及其概率，有必要引入概率分布的新的表示法。

HOW：定义区间$(-\infty, x]$上的概率

- 事实上，我们经常关心的是一个随机变量落入一个区间、空间的概率，如机器寿命大于 T 的概率，产品中次品数少于 k 件的概率。
- 注意到随机变量取值为实数，可以说分布在实数轴上的任意的区间 $(x_1, x_2]$ 都可以用 $(-\infty, x]$ 来表示，进一步的由于任意波雷尔集B是左开右闭区间的（有限或可列）并、（有限或可列）交、逆产生的集合，故此可以定义在区间 $(-\infty, x]$ 上概率。

2.3.2 定义

设随机变量X，对任意的实数x，随机变量X取值落入区间$(-\infty, x]$内的概率为F(x)

$$F(x) = P(X \le x) \qquad -\infty < x < +\infty$$

称为随机变量X的分布函数 (distribution function)。

[1] 对确定的随机变量X, 其分布函数F(x)是唯一确定的，它是实变量x的函数，因此我们可以利用实变函数论这一有力工具来研究随机变量。

[2] 有了分布函数，则对任一波雷尔集A，概率P(A)可以用分布函数来表示。

$$P(a < X \le b) = P\{(-\infty, b]\} - P\{(-\infty, a]\}$$
$$= F(b) - F(a)$$

2.3.3 分布函数的特性

由于分布函数已经是一个普通函数，就可以用数学分析、实变函数来研究。至此已经达到目的：

概率事件 $\Rightarrow$ 随机变量(统一描述) $\Rightarrow$ 分布函数(函数工具)

(1) $F(x)$是一个不减的函数， $\forall x_1 \le x_2, F(x_1) \le F(x_2)$

(2) $0 \le F(x) \le 1$

(3) $F(-\infty) = \lim\limits_{x \to -\infty} F(x) = 0$ $\quad$ $F(\infty) = \lim\limits_{x \to \infty} F(x) = 1$

(4) 函数的右连续性 $F(x+0) = \lim\limits_{y \to x+0} F(y)$

（严格证明见：《概率论》，王仁宫，pp.38，pp.66）

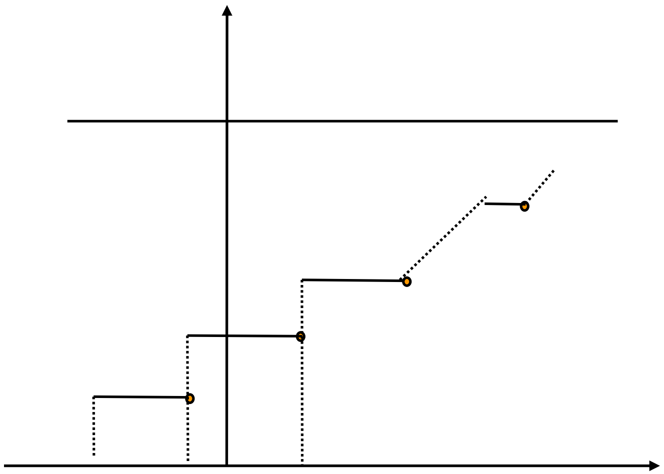
2.3 Distribution Function of Random Variable

Example 2.7 离散随机变量的分布函数，随机变量X的分布列为

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_k & \cdots \\ p(x_1) & p(x_2) & \cdots & p(x_k) & \cdots \end{pmatrix}$$

且 $x_1 < x_2 < \cdots\cdots < x_k < \cdots\cdots$ ，则X的分布函数为 $F(x) = P(X \le x) = \sum_{x_k \le x} p_k$

$$F(x) = \begin{cases} 0, & x < x_1, \\ p(x_1), & x_1 \le x < x_2, \\ \cdots & \cdots \\ \sum_{i \le k} p(x_i), & x_k \le x < x_{k+1}, \\ \cdots & \cdots \end{cases}$$



它是间断的分段函数，在x_k(k=1，2， …)各有一跳跃，跃度为$p(x_k)$，在每一段 [x_k，x_{k+1})中都是常数，呈阶梯形。

比如，随机变量X的分布律为

X	-1	2	3
p	1/4	1/2	1/4

X的分布函数为F(x)

$$F(x)=\begin{cases} 0 & x<-1 \\ P(X=-1) & -1\le x<2 \\ P\{X=-1\}+P\{X=2\} & 2\le x<3 \\ 1 & x\ge 3 \end{cases}$$

2.3 Distribution Function of Random Variable

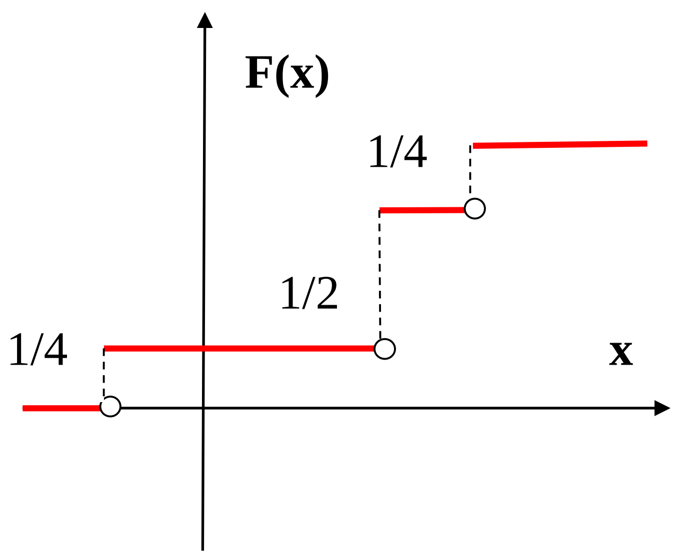
$$F(x)=\begin{cases}0 & x<-1\\ \dfrac{1}{4} & -1\le x<2\\ \dfrac{3}{4} & 2\le x<3\\ 1 & x\ge 3\end{cases}$$

$$P\left\{X\le\frac{1}{2}\right\}=F\left\{\frac{1}{2}\right\}=\frac{1}{4}$$

$$P\left\{\frac{3}{2}<X\le\frac{5}{2}\right\}=F\left(\frac{5}{2}\right)-F\left(\frac{1}{2}\right)=\frac{1}{2}$$

$$P\{2\le X\le 3\}=F(3)-F(2)+P(X=2)$$

$$=\frac{3}{4}$$

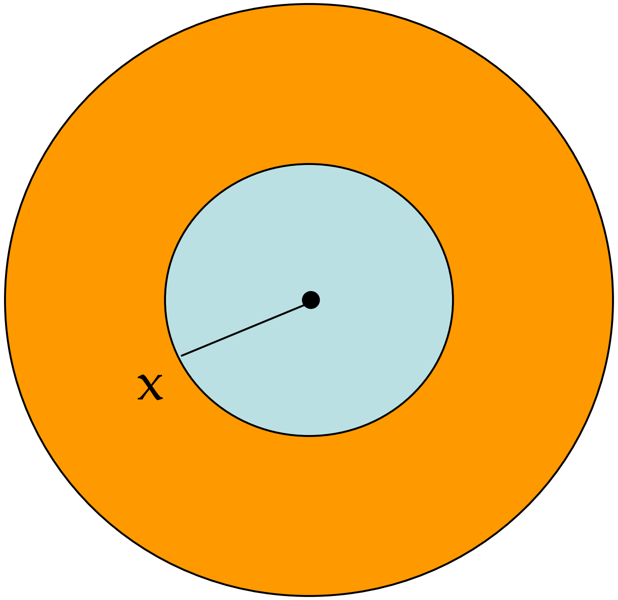


Example （pp.50 例2）

一个半径为2米的圆盘形靶子，设击中靶上任何一同心圆上的点的概率与该圆的面积成正比，并设射击都能中靶，以X表示弹着点与圆心的距离。

求：随机变量X的分布函数

解：如图，



[1] 若$x<0$，则$\{X\le x\}$是不可能事件

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

[2] 若$0\le x\le 2$, 由假设知道$\{0\le X\le x\}$的概率正比于圆的面积πx^2

$$P\{0\le X\le x\}=k\pi x^2$$

其中k为常数

当x＝2时，有 $P\{0\leq X\leq x\}=k\pi 4=1$

求得 $k\pi=1/4$

于是有 $P\{0\leq X\leq x\}=x^2/4$

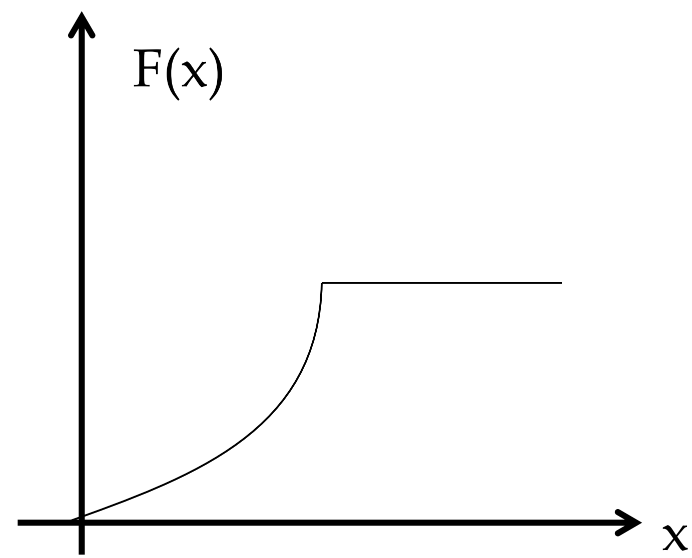
$F(x)=P\{X\leq x\}=P(X<0)+P\{0\leq X\leq x\}=x^2/4$

[3] 若x≥2，由题意{X ≤ x}是必然事件

$F(x)=P\{X\leq x\}=1$

综合以上知道：

$$F(x)=\begin{cases}0 & x<0\\ \dfrac{x^2}{4} & 0\leq x<2\\ 1 & x\geq 2\end{cases}$$



同时分布函数可以写为积分形式

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

$$f(t)=\begin{cases}\dfrac{t}{2} & 0<t<2\\ 0 & \text{其他}\end{cases}$$

上节课复习

Question

1. What is random variables?
2. Why define random variable?
3. What is Distribution Function?
4. What is the difference between Probability Function and Distribution Function?

Review

随机变量：样本空间为$S=\{e\}$，如果对每个 $e\in S$，有一个实数$X(e)$与之对应，就得到一个定义在S上的单值实函数$X=X(e)$。X就称为**随机变量**。

$$P\{X\in L\}=P\{D\},\qquad D=\{e\,|\,X(e)\in L\}$$

分布函数：设随机变量 X，对任意实数x，X 取值落入区间内的概率为$F(x)$

$$F(x)=P(X\le x),\quad -\infty<x<+\infty$$

$F(x)$表示的是什么？它表示的是这样一个随机事件的概率值：

事件：$A=\{\,e\,|\,X(e)\le x\,\}$

$F(x)=P\{A\}\equiv P\{\,e\,|\,X(e)\le x\}$

如果是在一点 x 的概率是多少呢？

2.4.1 连续型随机变量及密度函数

猜测连续型随机变量的分布函数（依据）：

[1] 概率定义一再地引用几何面积的概念；

[2] 对于一个随机变量 X 取得某个区间的所有值时，从离散随机变量的分布函数为概率的和。

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

Definition 1 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 满足

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

则称 X 为连续型(continuous)随机变量，称 $f(x)$ 为 X 的概率密度函数，简称为**密度函数(density function)**。

2.4.2 连续型随机变量的特性

连续型随机变量的分布函数$F(x)$ 具有下列数学性质:

(1) 由实变函数论可以证明，$F(x)$ 必定处处连续；在 $f(x)$ 的连续点，$F(x)$可导，且 $F'(x) = f(x)$.

(2) $\int_{-\infty}^{+\infty} f(x)dx = 1.$

(3) 随机变量X落在任意区间(a，b]的概率:

$$P\{a < x \le b\} = F(b) - F(a)$$
$$= \int_a^b f(x)dx.$$

良好的性质

(4) 特别，对任一常数c，$P\{X = c\} = 0.$

Remark 1：对连续型随机变量，计算在一点的概率是没有意义的，这也是不能用分布列描写连续型随机变量的理由之一。

Remark 2：对连续型随机变量$\{X=C\}$是一个可能发生的事件，因此一事件A的概率为0并不表明$A=\varnothing$（如在一个区间上方抛点，点可以落到任意一点 x，但是落到任意一点的概率为0）；同样若$P(A)=1$，也并不表明$A=S$（样本空间）。

Remark 3：除了离散型，连续型以外，随机变量还有其它类型吗？

$$F(x)=\begin{cases} 0, & x<0, \\ (1+x)/2, & 0\le x<1, \\ 1, & x>1. \end{cases}$$

是分布函数，它不是离散型的，也不是连续型的(因为 $F(x)$ 本身就不连续)。

2.4.3 常见的连续型随机变量

[一] 均匀(Uniform)分布

考虑在一个区间[a,b]上方抛点，假设点落入任意一个等长度区间的概率（可能性）是相同的（或者说点落入某个区间的可能性大小只与区间长度有关，与区间端点无关），用随机变量 X 表示区间的端点 $(-\infty, x]$，则 X 的分布函数为

$$F(x)=P\{X\le x\}=\begin{cases}0, & x\le a\\ \dfrac{x-a}{b-a}, & a<x<b\\ 1, & x\ge b\end{cases}$$

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

称具有上述概率密度的连续型随机变量 X 为在区间 $[a, b]$ 上服从**均匀分布**，简记作 $X\sim U(a, b)$。

2.4.3 常见的连续型随机变量

[一] 均匀(Uniform)分布

Example 考察一个数据，它在小数点n位后四舍五入，则其真值x与其近似值$\hat{x}$之间的误差$\varepsilon=\hat{x}-x$，一般假定服从$\left[-0.5\times10^{-n},0.5\times10^{-n}\right]$上的均匀分布。

利用均匀分布理论可以对大量运算后的数据进行误差分析。这对于使用计算机解题时是很重要的，因为计算机的字长总是有限的。

Example 列车从车站每40分钟开出一次，乘客到站时刻是机会均等的，问等候超过10分钟的概率是多少？

以T表示等候时间，它是一个随机变量。$P\{t_1<T\le t_1+\Delta t\}$指在区间$[t,t+\Delta t]$等到车的概率.

T取值范围为[0, 40]，其服从均匀分布 $\boldsymbol{U}$ [0, 40]。

$$P\{T\ge10\}=1-P\{T\le10\}=\int_{10}^{40}\frac{1}{40}dt=\frac{3}{4}$$

[二] 指数(Exponential)分布

随机变量 X 的密度函数为（当然，我们说它应该满足分布密度）

$$f(x)=\begin{cases}\dfrac{1}{\theta}e^{-x/\theta}=\dfrac{1}{\theta}Exp\left(-\dfrac{x}{\theta}\right), & x>0\\ 0, & else\end{cases}$$

其中θ>0为常数，则称X服从参数为θ的**指数分布**(Exponential)。

其分布函数为

$$F(x)=\begin{cases}1-e^{-x/\theta}, & x>0\\ 0, & else\end{cases}$$

指数分布常用来描述产品的寿命。

[二] 指数分布

Example 若是用了t 小时的电子管，在以后的Δt小时内失效的概率为 $\lambda\Delta t+o(\Delta t)$，其中$\lambda$是不依赖于$t$ 的正常数，求电子管在 T 小时内失效的概率（假定电子管寿命为0的概率是0）。

解： 设 X 为电子管的寿命。注意，对于成批的电子管而言，X 是一个随机变量。按题意要找出 $P\{X\le T\}$，即要找出 X 的分布函数 $F(t)$，求 $F(t)$ 在 T 点的值。

从问题给出的条件入手，其可以表示为：

$$P\{t<X<t+\Delta t \mid t<X\}=\lambda\Delta t+o(\Delta t)$$

$$\frac{P\{(t<X<t+\Delta t)\bigcap(t<X)\}}{P\{t<X\}}=\lambda\Delta t+o(\Delta t)$$

$$\frac{P\{(t<X<t+\Delta t)\}}{P\{t<X\}}=\lambda\Delta t+o(\Delta t)$$

[二] 指数分布

$$\frac{F(t+\Delta t)-F(t)}{1-F(t)}=\lambda\Delta t+o(\Delta t)$$

$$\frac{F(t+\Delta t)-F(t)}{\Delta t}=\lambda\big(1-F(t)\big)+o(\Delta t)/\Delta t$$

$$F'(t)=\lambda\big(1-F(t)\big)$$

解此线性微分方程组（注意到初始条件$F(0)=0$），解得

$$F(t)=1-e^{-\lambda t}$$

X 有密度函数 $$f(t)=\begin{cases}\lambda e^{-\lambda t} & t>0\\ 0 & else\end{cases}$$

电子管在T小时内失效的概率为 $1-e^{-\lambda T}$.

注意：寿命←→失效的对立性

寿命为T小时的概率为 $1-F(T)=e^{-\lambda T}$.

[二] 指数分布

指数分布具有类似几何分布的“**无记忆性**”：使用s小时后元件的寿命不低于t的概率与初始时使用寿命为t的概率相等。

事实上，设随机变量ξ服从参数为λ的指数分布，则对于任意的$s>0$, $t>0$,

$$P\{(\xi>s+t)\cap(\xi>s)\}$$

$$P(\xi>s+t\,|\,\xi>s)=P(\xi>s+t,\xi>s)/P(\xi>s)$$

$$=e^{-\lambda(s+t)}/e^{-\lambda s}$$

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

$$=P(\xi>t)$$

[三] 正态分布

若随机变量 X 的密度函数为

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

就称 X 服从**正态(Normal)分布**，记作 $X\sim N(\mu,\sigma^2)$。其中, $\sigma>0$, $-\infty<\mu<\infty$。

我们来证明上述定义的 $f(x)$ 的确是密度函数。

显然 $f(x)>0$;

进一步证明 $\int_{-\infty}^{\infty} f(x)=1$，即证明

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

[三] 正态分布

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

$$\left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{t^2}{2}}dt\right]^2=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-\frac{t^2}{2}}e^{-\frac{s^2}{2}}dtds$$

$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-\frac{t^2+s^2}{2}}dtds$$

上述二重积分可用极坐标表示成

$$\frac{1}{2\pi}\int_{0}^{2\pi}d\theta\int_{0}^{\infty}e^{-\frac{1}{2}r^2}rdr=\int_{0}^{\infty}e^{-\frac{1}{2}r^2}rdr=1.$$

也即 $\int_{-\infty}^{\infty}f(x)=1.$

2.4 Distribution for Continuous Random Variables

Remark 1： 正态分布是概率论中最重要的一种分布，与二项分布、泊松分布并称为**三大分布**。

正态分布是19世纪初高斯（Gauss）在研究测量误差时首次引进的，故正态分布又称误差分布或高斯分布。

Remark 2： 对正态分布应用很广，一般说来，若影响某一数量指标的随机因素很多，而每一因素所起的作用又不很大，则这个数量指标服从正态分布。例如进行测量时，由于仪器精度、人的视力、心理因素、外界干扰等多种因素影响，测量结果大致服从正态分布，测量误差也服从正态分布。

Remark 3： 正态分布具有良好的性质，一定条件下，很多分布可用正态分布来近似表达；另一些分布又可以通过正态分布来导出，因此，正态分布在理论研究中也相当重要。

1. 正态分布的特性

正态分布的密度曲线是一条关于纵轴对称的钟形曲线。特点是“两头小，中间大，左右对称”。

[1] 曲线 $f(x)$ 关于 $x=\mu$对称。

$$f(\mu-x)=f(\mu+x) \qquad P\{\mu-h<X<\mu\}=P\{\mu<X<\mu+h\}$$

[2] 当$x=\mu$时，$f(x)$取得最大值 $\dfrac{1}{\sqrt{2\pi}\sigma}$

当$x<\mu$时，$f(x)$单调递增；

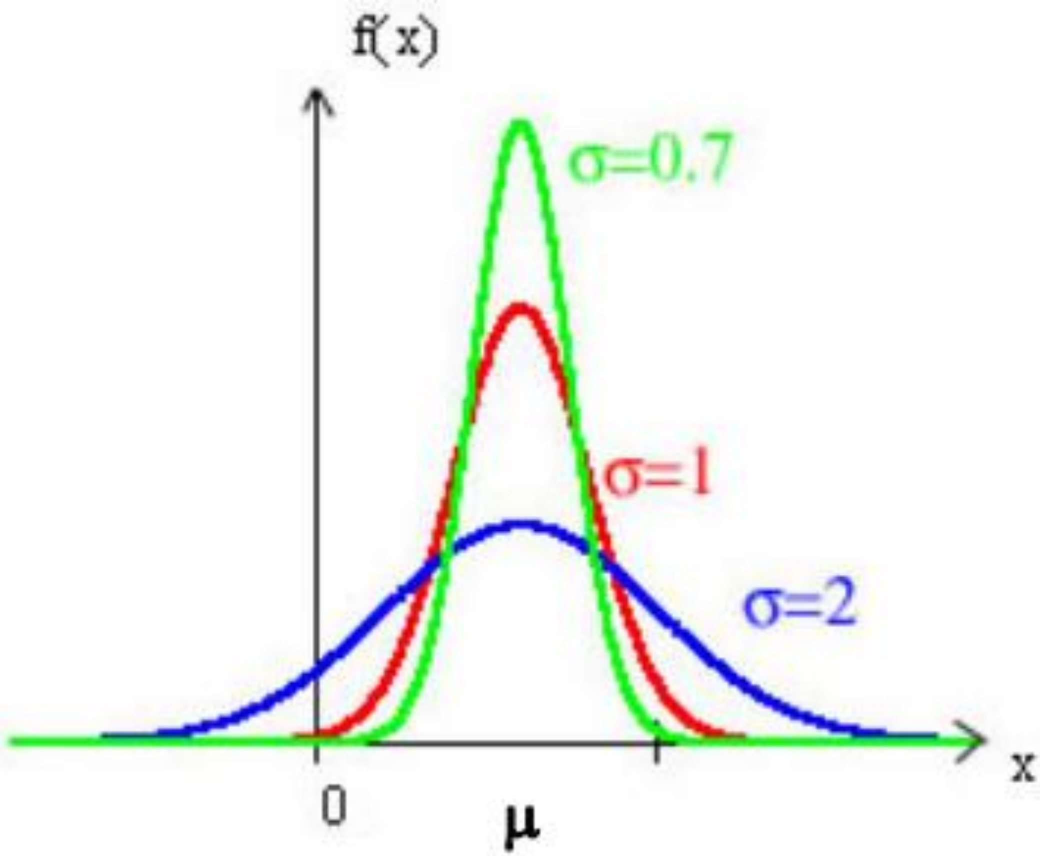
当$x>\mu$时，$f(x)$单调递减；

$x\to\infty$ 时，$f(x)\to 0.$

这些表明，x 离 μ 越远，$f(x)$的值越小。对于同样长度的区间，当区间离 μ 越远时，X 落在此区间的概率越小。

1. 正态分布的特性

[3] σ^2（方差）越大，最高点越低，$f(x)$ 的图形越扁平，说明 X 分布比较分散，X 取值离开 μ 点远的概率也越大；

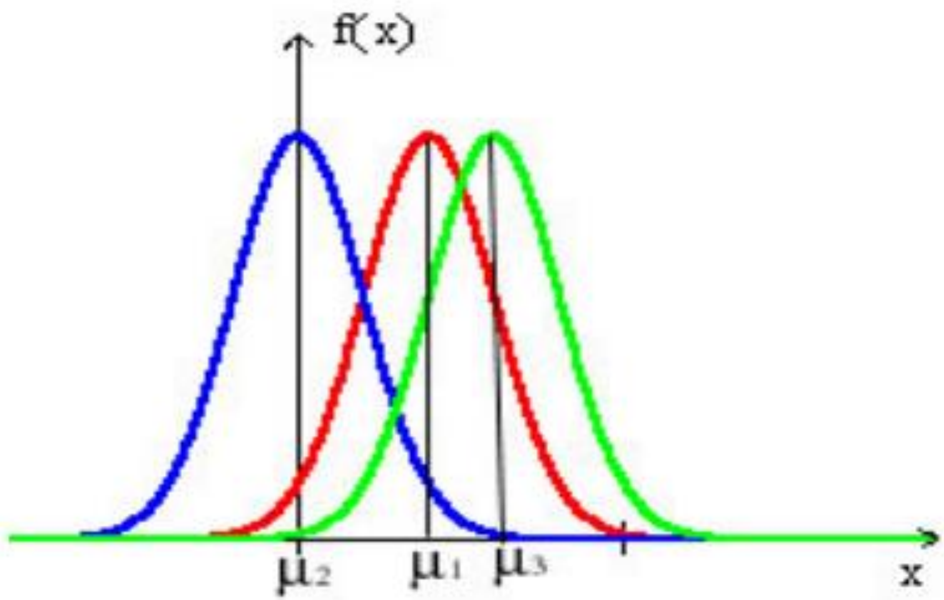


σ^2（方差）越小，最高点越高，说明 X 分布比较集中；$f(x)$的图形越陡峭，X 取值越集中在点 μ 附近。

1. 正态分布的特性

[4] 固定σ，改变μ的值，则图形沿着x轴平移，而不改变其形状。。

可见正态分布的概率密度曲线的位置完全由参数μ决定，μ称为**位置参数**。



[1] μ、σ 分别称为均值与均方差；

[2] 考研：四科成绩分布，希望是均值尽量高（总分过线），方差也要尽量小（别偏科）。

甲：总分307，英语60，政治65，数学85， 专业97

乙：总分312，英语46，政治65，数学101，专业100

丙：总分264，英语51，政治51，数学81， 专业81

1. 正态分布的特性

[5] 当 $\mu=0, \sigma=1$时，称为**标准正态分布**(standardized normal distribution)，它的密度曲线关于纵轴对称，其密度及分布函数特别记为 $\varphi(x)$ 和 $\Phi(x)$：

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \qquad -\infty < x < \infty$$

容易知道 $\Phi(-x)=1-\Phi(x)$.

2. 正态分布的计算

[1] 标准正态分布概率的计算

标准正态分布的计算不容易，有专门的表格供查阅。

当 $x \ge 0$ 时, 每隔一定数值，可以查到对应的分布函数 $\Phi(x)$的值；

在这些数值之间，用线性插值法求得相应的函数值；

(附录中这个间隔是纵向排列的，间隔是0.1，插补为0.01，横向排列）。

当 $x < 0$ 时, 注意到

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

结合 $-x > 0$时的 $\Phi(-x)$ 表就可算出 $x<0$时 $\Phi(x)$ 的值。

Example 设 $X \sim N(0, 1)$。

(1) 计算 $P\{-1<X<3\}$;

(2) 已知 $P\{X<\lambda\}=0.9755$，求λ。

解： (1) $P(-1<X<3)=P(X<3)-P(X<-1)$

$$=\Phi(3)-\Phi(-1)$$

$$=\Phi(3)-\left(1-\Phi(1)\right)$$

$$=0.9987+0.8413-1$$

$$=0.8400$$

2. 正态分布的计算

(2) $\Phi(\lambda)=0.9755$，求λ。

我们倒查表格，看看0.9755对应的值是多少。

但是，表格没有这样的值, 只有相近的。

$$\Phi(1.96)=0.9750<\Phi(\lambda)<\Phi(1.97)=0.9756$$

由于$\Phi(x)$是单调不减的，故λ在1.96与1.97之间

$$1.96<\lambda<1.97$$

在这个小的范围内，我们把λ与$\Phi(\lambda)$的关系近似看作线性：

$$\frac{\lambda-1.96}{1.97-1.96}\approx\frac{\Phi(\lambda)-\Phi(1.96)}{\Phi(1.97)-\Phi(1.96)}$$

$$\lambda\approx 1.96+\frac{\Phi(\lambda)-\Phi(1.96)}{\Phi(1.97)-\Phi(1.96)}(1.97-1.96)$$

$$\approx 1.968$$

以上思路称为**线性插值法**。

3. 一般正态分布的概率计算

对一般的 $X \sim N(\mu, \sigma^2)$，可以变换为标准正态分布加以计算。

记$Y=(X-\mu)/\sigma$ (称为X的**标准化随机变量**)，则 Y 服从 $N(0, 1)$。

事实上 Y 的分布函数 $F_y(y)$

$$
\begin{aligned}
F_y(y) &= P\{Y \le y\} \\
&= P\{X \le \sigma y + \mu\} \\
&= \int_{-\infty}^{\sigma y+\mu} \frac{1}{\sqrt{2\pi}\sigma} Exp\left(-(t-\mu)^2/2\sigma^2\right) dt
\end{aligned}
$$

$$\Big\Downarrow \; z = \frac{t-\mu}{\sigma}$$

$$= \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} Exp\left(-\frac{z^2}{2}\right) dz$$

3. 一般正态分布的概率计算

由此可知： $X \sim N(\mu, \sigma^2)$ ，则

$$P\{X \le x\} = P\left\{\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right\}$$

$$= P\left\{Y \le \frac{x-\mu}{\sigma}\right\}$$

$$= \Phi\left\{\frac{x-\mu}{\sigma}\right\}$$

对于任意的区间 $(x_1, x_2]$

$$P\{x_1 < X \le x_2\} = P\left\{\frac{x_1-\mu}{\sigma} < \frac{X-\mu}{\sigma} \le \frac{x_2-\mu}{\sigma}\right\}$$

$$= \Phi\left(\frac{x_2-\mu}{\sigma}\right) - \Phi\left(\frac{x_1-\mu}{\sigma}\right)$$

3. 一般正态分布的概率计算

Example 汽车设计手册中指出：人的身高服从正态分布 $X \sim N(\mu, \sigma^2)$。根据各国统计资料，可得各国、各民族男子身高的 μ 和 σ。对于中国人，$\mu = 1.75$，$\sigma = 0.05$。现要求上下车时要低头的人不超过0.5%，车门需要多高？

解： 设大巴士车车门高位h，X 为乘客的身高，则 $X \sim N(1.75, 0.05^2)$，根据题意

$$P\{X \le h\} \ge 99.5\%$$

$$P\left\{\frac{X-1.75}{0.05} \le \frac{h-1.75}{0.05}\right\} \ge 99.5\%$$

$$\Phi\left(\frac{h-1.75}{0.05}\right) \ge 99.5\%$$

$$\frac{h-1.75}{0.05} \ge 2.58$$

$$h \ge 1.879$$

如此，知道车门应该高1.88米。

3. 一般正态分布的概率计算

Example从南郊某地乘车到北区火车站有两条路可走，第一条路较短，但交通拥挤，所需时间T_1服从 $N(50, 100)$ 分布；第二条路线略长，但意外阻塞较少，所需时间T_2服从 $N(60, 16)$。

(1) 若有70分钟可用，问应走哪一条路？

(2) 若只有65分钟可用，又应走哪一条路？

分析： 应该走在允许时间内有较大概率赶到火车站的路线。

解： (1) 若有70分钟可用

走第一条路线及时赶到的概率。

$$P\{T_1 \le 70\} = \Phi\left(\frac{70-50}{10}\right) = \Phi(2) = 0.9772$$

走第二条路线及时赶到的概率。

$$P\{T_2 \le 70\} = \Phi\left(\frac{70-60}{4}\right) = \Phi(2.5) = 0.9938$$

在这种场合，应走第二条路线。

(2) 若只有65分钟可用，又应走哪一条路？

走第一条路线及时赶到火车站的概率为

$$P\{T_1 \le 65\} = \Phi\left(\frac{65-50}{10}\right) = \Phi(1.5) = 0.9332$$

走第二条路线及时赶到的概率为

$$P\{T_2 \le 65\} = \Phi\left(\frac{65-60}{4}\right) = \Phi(1.25) = 0.8944$$

因此在这种场合，应走第一条路线更为保险。

Remark： 此处的μ和σ的含义分别是走完这段路程所需要的平均时间和均方差。

3. 一般正态分布的概率计算

Example 设 $X \sim N(\mu, \sigma^2)$，

求 $P\{|X-\mu|<\sigma\}, P\{|X-\mu|<2\sigma\}, P\{|X-\mu|<3\sigma\}$

解： $Y=(X-\mu)/\sigma$，则 $Y \sim N(0, 1)$，故由标准正态分布函数 $\Phi(y)$，

可以求得 $P\{|X-\mu|<k\sigma\} = P\left\{-k < \dfrac{X-\mu}{\sigma} < k\right\}$

$$= \Phi(k) - \Phi(-k)$$

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

$$P\{|X-\mu|<\sigma\} = 2\Phi(1) - 1 \approx 0.6826$$

$$P\{|X-\mu|<2\sigma\} = 2\Phi(2) - 1 \approx 0.9544$$

$$P\{|X-\mu|<3\sigma\} = 2\Phi(3) - 1 \approx 0.9973$$

说明正态随机变量的99.73 %的值落在(μ-3σ, μ+3σ)之中，落在该区间之外的概率几乎为零，这情况被实际工作者称为“**3σ原则**”。

2.5 随机变量的函数及其分布-I

人们经常碰到随机变量的函数。例如分子运动动能 $T = mv^2/2$ 是分子运动速度—随机变量 v 的函数；

数理统计中经常用到 $\chi^2(n)$ 分布，相应的随机变量

$$\chi^2 = \xi_1^2 + \xi_2^2 + \cdots + \xi_n^2$$

其中各 ξ_i 相互独立，都服从 $N(0, 1)$。 χ^2是$\xi_1, \ldots, \xi_n$的函数。

一般地，若 X 是随机变量， $y = g(x)$是普通的实函数，则 $Y = g(X)$ 是 X 的函数。

产生两个问题：

(1) Y 是随机变量吗？

(2) 如果是，Y 的分布与 X 的分布有什么关系？

2.5.1 离散型随机变量的函数

Example (pp.62 例1) 假设随机变量 X 具有如下分布律

X	-1	0	1	2
p_n	0.2	0.3	0.1	0.4

求： $Y=(X-1)^2$ 的分布。

解： Y 的的可能取值为0,1,4, 是有限个，只须算出对应的概率。由于

$$P\{Y=0\} = P\{X=1\} = 0.1,$$

故$P\{Y=1\} = P\{X=0\}+P\{X=2\} = 0.7$

$$P\{Y=4\} = P\{X = -1\} = 0.2$$

类似可得其它概率。我们得到 Y 的分布：

Y	0	1	4
p_n	0.1	0.7	0.2

一般，设 X 有分布列 $P\{X = x_i\} = p\{x_i\}$，$i=1,2,\ldots$， 则

$Y=f(X)$ 有分布列 $\boxed{P\{Y=y_j\} = \sum_{f(x_i)=y_j} p(x_i), \quad j=1,2,\cdots}$

例2 设 $X \sim B(n_1, p)$, $Y \sim B(n_2, p)$，X、Y相互独立，

求： $Z = X + Y$ 的分布。

解： X, Y 各可取值 $0, 1, \ldots, n_1$ 和 $0, 1, \ldots, n_2$, 则 Z 可取值$0, 1, \ldots, n_1+n_2$, 得

$$P\{Z = s\} = \sum_{k=0}^{s} P\{X = k, Y = s-k\}$$

$$= \sum_{k=0}^{s} P\{X = k\} P\{Y = s-k\}$$

$$= \sum_{k=0}^{s} C_{n_1}^{k} p^{k} (1-p)^{n_1-k} C_{n_2}^{s-k} p^{s-k} (1-p)^{n_2-(s-k)}$$

$$= p^{s} (1-p)^{n_1+n_2-s} \sum_{k=0}^{s} C_{n_1}^{k} C_{n_2}^{s-k}$$

$$= C_{n_1+n_2}^{s} p^{s} (1-p)^{n_1+n_2-s}$$

$\boxed{X+Y \sim B(n_1+n_2, p)}$

Remark1： 上面公式的推导用到了组合数的性质。

$$\sum_{k=0}^{s} C_{n_1}^{k} C_{n_2}^{s-k} = C_{n_1+n_2}^{s}$$

Remark2： $X+Y \sim B(n_1+n_2, p)$ 这个事实显示了二项分布一个很重要的性质：**两个独立的二项分布，当它们的第二参数相同时，其和也服从二项分布，它的第一参数恰为这两个二项分布第一参数的和。**

- 这性质称为二项分布的<u>**再生性**</u>(或<u>**可加性**</u>(additive property))
- 从X, Y的概率意义来看，这结果是非常明显的：X 和Y 分别是 n_1和 n_2重贝努里试验中成功的次数，两组试验合起来，$Z=X+Y$应该就是n_1+n_2重贝努里试验中成功的次数。

2.5.2 一维连续型随机变量的函数的分布

设 X 的密度函数为 $f_X(x)$，我们要求出$Y=g(X)$的分布函数 $F_Y(y)$ 。事实上

$$F_Y(y)=P\left(Y\le y\right)=P\left(g\left(X\right)\le y\right)$$

$$=P\left(X\in D\right)$$

而 $D=\{x\mid g(x)\le y\}$是一维的波雷尔集，故

$$F_Y(y)=P\left(X\in D\right)=\int_{x\in D} f_X(x)dx$$

至于Y是不是连续型随机变量，它的密度函数是什么，在一般场合无法作出决定，但在某些特殊而又常见的场合，我们可以直接导出 Y 的密度函数 $f_Y(y)$。

Example $X \sim N(0,1)$, 求 $Y=X^2$ 的密度。

解： 先求分布函数 $F_Y(y)$，再微分求得 $f_Y(y)$。

$$F_Y(y)=P\left(Y\le y\right)$$

$$=P\left(X^2\le y\right)$$

$$=\begin{cases} 0 & y\le 0 \\ P\left(-\sqrt{y}<X<\sqrt{y}\right) & y>0 \end{cases}$$

$$=\begin{cases} 0 & y\le 0 \\ \int_{-\sqrt{y}}^{\sqrt{y}} f_X(x)dx & y>0 \end{cases}$$

$$f_Y(y)=\begin{cases} 0 & y\le 0 \\ \dfrac{1}{2\sqrt{y}}\left[f_X(\sqrt{y})+f_X(-\sqrt{y})\right] & y>0 \end{cases}$$

当 $X \sim N(0,1)$ 时，$Y = X^2$

$$f_Y(y) = \begin{cases} \dfrac{1}{2\sqrt{y}}\left(\dfrac{1}{\sqrt{2\pi}}e^{-y/2} + \dfrac{1}{\sqrt{2\pi}}e^{-y/2}\right) & y > 0 \\ 0 & y \le 0 \end{cases}$$

$$= \begin{cases} \dfrac{1}{\sqrt{2\pi}} y^{-1/2} e^{-y/2} & y > 0 \\ 0 & y \le 0 \end{cases}$$

此时称 Y **服从自由度为1的$\chi^2(1)$ 分布**，它也是 **Γ-分布**的一种。

2.5.2 一维连续型随机变量的函数的分布

某些情形下，我们可以直接导出 Y 的密度函数 $f_Y(y)$ 。

定理1 X 为连续型随机变量，有概率密度函数

$$f_X(x)(-\infty < x < \infty)$$

若$g(X)$严格单调且处处可导，则$Y=g(X)$也是连续型随机变量。若令其中 $h(y)$ 是 $g(x)$ 的反函数，则Y的密度函数为

$$\boxed{f_Y(y)=\begin{cases}|h'(y)|\, f_X[h(y)] & y\in g(x)\text{的值域}\\ 0 & y\in \text{其它}\end{cases}}$$

证： 不妨设 $g(x)$ 严格单调增加，且$-\infty < x < +\infty$时, $\alpha < g(x) < \beta$。显然若 $y \le \alpha$，则$F_Y(y) = 0$；若 $y > \beta$，则$F_Y(y) = 1$，两种情形下都有$f_Y(y) = 0$。当 $\alpha < y < \beta$时，$\{ Y \le y \} = \{ g(X) \le y \} = \{ X \le h(y) \}$，故

$$F_Y(y) = P(Y \le y) = P\left(X \le h(y)\right) = \int_{-\infty}^{h(y)} f_X(x)dx$$

$F_Y(y)$ 对 y 求导，得到 Y 的概率密度函数

$$f_Y(y) = \begin{cases} h'(y) f_X\left[h(y)\right] & y \in g(x)\text{的值域} \\ 0 & y \in \text{其他} \end{cases}$$

当 $y = f(x)$为严格单调减少时，类似可证：

$$f_Y(y) = \begin{cases} -h'(y) f_X\left[h(y)\right] & y \in g(x)\text{的值域} \\ 0 & y \in \text{其他} \end{cases}$$

二者结合，可得定理的结论。

2.5.2 一维连续型随机变量的函数的分布

Example 设 $X\sim N(\mu, \sigma^2)$，求$Y=aX+b$的密度函数 $(a\neq 0)$。

解： $y = g(x) = ax+b$ 满足上述定理1中的条件，$h(y) = (y-b)/a$

Y 的密度

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

$$f_Y(y)=f_X(h(y))|h'(y)|$$

$\boxed{\text{结论：}\ Y\sim N\left(a\mu+b,(|a|\sigma)^2\right)}$

$$=\frac{1}{\sqrt{2\pi}\sigma}Exp\left(-\left(\frac{y-b}{a}-\mu\right)^2\Big/2\sigma^2\right)\times\frac{1}{|a|}$$

$$=\frac{1}{\sqrt{2\pi}|a|\sigma}Exp\left(-\left[y-(a\mu+b)\right]^2\Big/2(a\sigma)^2\right)$$

$$=\frac{1}{\sqrt{2\pi}|a|\sigma}e^{-\left[y-(a\mu+b)\right]^2/2(a\sigma)^2}$$