概率论基础知识

（补充读物）

第一章 随机事件及其概率 2
一 随机事件 2
二 事件的概率 3
三 条件概率 7
第二章 随机变量及其分布函数 9
一 随机变量及其分布函数 9
二 离散型随机变量及其分布律 12
三 连续型随机变量及其概率密度 14
四 随机变量函数的分布 20
第三章 随机变量的数字特征 22
一 数学期望 22
二 方差 26
三 标准差 30
四 协方差及相关系数 30

高顿财经 CPA 研究院编制

第一章 随机事件及其概率

一 随机事件

§1 几个概念

1、**随机实验**:满足下列三个条件的试验称为随机试验;（1）试验可在相同条件下重复进行;（2）试验的可能结果不止一个，且所有可能结果是已知的;（3）每次试验哪个结果出现是未知的;随机试验以后简称为试验，并常记为 E。

例如：E_1：掷一骰子，观察出现的总数;

E_2：上抛硬币两次，观察正反面出现的情况;

E_3：观察某电话交换台在某段时间内接到的呼唤次数。

2、**随机事件**：在试验中可能出现也可能不出现的事情称为随机事件：常记为 A，B，C……

例如，在 E_1 中，A 表示“掷出 2 点”，B 表示“掷出偶数点”均为随机事件。

3、**必然事件与不可能事件**：每次试验必发生的事情称为必然事件，记为Ω。每次试验都不可能发生的事情称为不可能事件，记为Φ。

例如，在 E_1 中，“掷出不大于 6 点”的事件便是必然事件，而“掷出大于 6 点”的事件便是不可能事件,以后，随机事件，必然事件和不可能事件统称为事件。

4、**基本事件**：试验中直接观察到的最简单的结果称为基本事件。

例如，在 E_1 中，“掷出 1 点”，“掷出 2 点”，……，“掷出 6 点”均为此试验的基本事件。

由基本事件构成的事件称为复合事件，例如，在 E_1 中“掷出偶数点”便是复合事件。

5、**样本空间**：从集合观点看，称构成基本事件的元素为样本点，常记为 e.

例如，在 E_1 中，用数字 1，2，……，6 表示掷出的点数，而由它们分别构成的单点集{1}，{2}，…{6}便是 E_1 中的基本事件。在 E_2 中，用 H 表示正面，T 表示反面，此试验的样本点有（H，H），（H，T），（T，H），（T，T），其基本事件便是{（H，H）}，{（H，T）}，{（T，H）}，{（T，T）}显然，任何事件均为某些样本点构成的集合。

例如， 在 E_1 中“掷出偶数点”的事件便可表为{2，4，6}。试验中所有样本点构成的集合称为样本空间。记为Ω。

例如，

在 E_1 中，Ω={1，2，3，4，5，6}

在 E_2 中，Ω={（H，H），（H，T），（T，H），（T，T）}

在 E_3 中，Ω={0，1，2，……}

例 1，一条新建铁路共 10 个车站，从它们所有车票中任取一张，观察取得车票的票种。

此试验样本空间所有样本点的个数为 $N_\Omega=P^2_{10}=90$.（排列：和顺序有关，如北京至天津、天津至北京）

若观察的是取得车票的票价，则该试验样本空间中所有样本点的个数为 $N_\Omega=\binom{10}{2}=45$ (组合)

例 2．随机地将 15 名新生平均分配到三个班级中去，观察 15 名新生分配的情况。此试验的样本空间所有样本点的个数为

$$N_\Omega=\binom{15}{5}\binom{10}{5}\binom{5}{5}\text{或者}N_\Omega=\frac{15!}{5!5!5!}$$

第一种方法用组合+乘法原理；第二种方法用排列

二 事件的概率

§1 概率的定义

所谓事件 A 的概率是指事件 A 发生可能性程度的数值度量，记为 P（A）。规定 P(A)≥0，P（Ω）=1。

1、古典概型中概率的定义

古典概型：满足下列两条件的试验模型称为古典概型。

（1）所有基本事件是有限个；　（2）各基本事件发生的可能性相同；

例如：掷一匀称的骰子，令 A={掷出 2 点}={2}，B={掷出偶数总}={2，4，6}。此试验样本空间为

Ω={1，2，3，4，5，6}，于是，应有 1=P（Ω）=6P（A），即 P（A）= $\frac{1}{6}$。

而 P（B）=3P（A）= $\frac{3}{6}=\frac{B\text{所含的基本事件数}}{\text{基本事件总数}}$

定义 1：在古典概型中，设其样本空间Ω所含的样本点总数，即试验的基本事件总数为 N_Ω 而事件 A 所含的样本数，即有利于事件 A 发生的基本事件数为 N_A，则事件 A 的概率便定义为：

$$P(A)=\frac{N_A}{N_\Omega}=\frac{A\text{包含基本事件数}}{\text{基本事件总数}}$$

例 1，将一枚质地均匀的硬币一抛三次，求恰有一次正面向上的概率。

解：用 H 表示正面，T 表示反面，则该试验的样本空间

Ω={（H，H，H）（H，H，T）（H，T，H）（T，H，H）（H，T，T）（T，H，T）（T，T，H）（T，T，T）}。

可见 $N_{\Omega}=8$ 令 A={恰有一次出现正面}，则 A={（H，T，T）（T，H，T）（T，T，H）}

可见，令 $N_A=3$ 故 $P(A)=\frac{N_A}{N_\Omega}=\frac{3}{8}$

例 2，**（取球问题）**袋中有 5 个白球，3 个黑球，分别按下列三种取法在袋中取球。

（1）**有放回地取球**：从袋中取三次球，每次取一个，看后放回袋中，再取下一个球；

（2）**无放回地取球**：从袋中取三次球，每次取一个，看后不再放回袋中，再取下一个球；

（3）**一次取球**：从袋中任取 3 个球。*在以上三种取法中均求 A={恰好取得 2 个白球}的概率。*

解：（1）有放回取球 $N_{\Omega}=8\times8\times8=8^3=512$ (袋中八个球，不论什么颜色，取到每个球的概率相等)

$N_A=\binom{3}{2}5\times5\times3=\binom{3}{2}5^2 3^1=225$ （先从三个球里取两个白球，第一次取白球有五种情况，第二次取白球还有五种情况<注意是有放回>，第三次取黑球只有三种情况） $P(A)=\frac{N_A}{N_\Omega}=\frac{225}{512}=0.44$

（2）无放回取球 $N_\Omega=8\times7\times6=A_8^3=336$

$N_A=\binom{3}{2}5\times4\times3=\binom{3}{2}A_5^2A_3^1=180$ 故 $P(A)=\frac{N_A}{N_\Omega}=\frac{180}{336}=0.54$

（3）一次取球 $N_\Omega=\binom{8}{3}=\frac{8!}{3!5!}=56$ $N_A=\binom{5}{2}\binom{3}{1}=30$

故 $P(A)=\frac{N_A}{N_\Omega}=\frac{\binom{5}{2}\binom{3}{1}}{\binom{8}{3}}=\frac{30}{56}=0.54$

属于取球问题的一个实例：

设有 100 件产品，其中有 5%的次品，今从中随机抽取 15 件，则其中恰有 2 件次品的概率便为

$$P=\binom{5}{2}\binom{95}{13}\Big/\binom{100}{5}=0.1377$$

（属于一次取球模型）

例 3（**分球问题**）将 n 个球放入 N 个盒子中去，试求恰有 n 个盒子各有一球的概率（n≤N）。

解： 令 A={恰有 n 个盒子各有一球}，先考虑基本事件的总数

$$N_\Omega=\underbrace{N\cdot N\cdots N}_{n}=N^n$$

先从 N 个盒子里选 n 个盒子，然后在 n 个盒子里 n 个球全排列

$$N_A=\binom{N}{n}n!$$

故 $P(A)=\dfrac{\binom{N}{n}n!}{N^n}$

属于分球问题的一个实例：

全班有 40 名同学，向他们的生日皆不相同的概率为多少？令 A={40 个同学生日皆不相同}，则有

$$N_\Omega=365^{40},N_A=\binom{365}{40}40!$$

(可以认为有 365 个盒子，40 个球)

故 $P(A)=\dfrac{\binom{365}{40}40!}{365^{40}}\approx 0.109$

例 4（**取数问题**）

从 0，1，……,9 共十个数字中随机的<u>不放回</u>的接连取四个数字，并按其出现的先后排成一列，求下列事件的概率：（1） 四个数排成一个偶数；（2） 四个数排成一个四位数；（3） 四个数排成一个四位偶数；

解：令 A={四个数排成一个偶数}，B={四个数排成一个四位数}，C={四个数排成一个四位偶数}

$$N_\Omega=A_{10}^4=10\times9\times8\times7;\ N_A=\binom{5}{1}A_9^3=5\times9\times8\times7;故P(A)=\frac{5\times9\times8\times7}{10\times9\times8\times7}=0.5$$

$$N_B=A_{10}^4-A_9^3=10\times9\times8\times7-9\times8\times7,\ 故P(B)=\frac{10\times9\times8\times7-9\times8\times7}{10\times9\times8\times7}=0.9$$

$$N_C=\binom{5}{1}A_9^3-\binom{4}{1}A_8^2=5\times9\times8\times7-4\times8\times7,\ 故P(C)=\frac{5\times9\times8\times7-4\times8\times7}{10\times9\times8\times7}=0.456$$

例 5（**分组问题**）将一幅 52 张的朴克牌平均地分给四个人，分别求有人手里分得 13 张黑桃及有人手里有

4 张 A 牌的概率各为多少？

解：令 A={有人手里有 13 张黑桃}，B={有人手里有 4 张 A 牌}

$$N_\Omega=\binom{52}{13}\binom{39}{13}\binom{26}{13}\binom{13}{13},N_A=\binom{4}{1}\binom{39}{13}\binom{26}{13}\binom{13}{13}$$

于是 $$P(A)=\frac{N_A}{N_\Omega}=\frac{4\binom{39}{13}\binom{26}{13}\binom{13}{13}}{\binom{52}{13}\binom{39}{13}\binom{26}{13}\binom{13}{13}}=\frac{4}{\binom{52}{13}}=6.3\times10^{-12}$$

$$N_B=\binom{4}{1}\binom{48}{9}\binom{39}{13}\binom{26}{13}\binom{13}{13}$$，故 $$P(B)=\frac{N_B}{N_\Omega}=\frac{\binom{4}{1}\binom{48}{9}\binom{39}{13}\binom{26}{13}\binom{13}{13}}{\binom{52}{13}\binom{39}{13}\binom{26}{13}\binom{13}{13}}=\frac{4\binom{48}{9}}{\binom{52}{13}}=0\ 01$$

不难证明，古典概型中所定义的概率有以下三条**基本性质**：

1° P（A）≥0

2° P（Ω）=1

3° 若 A_1，A_2，……，A_n 两两互不相容，则 $P(\bigcup_{i=1}^{n}A_i)=\sum_{i=1}^{n}P(A_i)$

2、概率的统计定义

频率：在 n 次重复试验中，设事件 A 出现了 n_A 次，则称：$f_n(A)=\frac{n_A}{n}$ 为事件 A 的频率。频率具有一定的稳定性。示例见下例表

试验者	抛硬币次数 n	正面（A）出现次数 n_A	正面（A）出现的频率 $f_n(A)=\frac{n_A}{n}$
德·摩尔根	2048	1061	0．5180
浦丰	4040	2148	0．5069
皮尔逊	12000	6019	0．5016

皮尔逊	24000	12012	0．5005
维尼	30000	14994	0．4998

定义 2：在相同条件下，将试验重复 n 次，如果随着重复试验次数 n 的增大，事件 A 的频率 $f_n(A)$越来越稳定地在某一常数 p 附近摆动，则称常数 p 为事件 A 的概率，即 P（A）=p

不难证明频率有以下基本性质：

1° $f_n(A) \geq 0$　　2° $f_n(\Omega) = 1$

3° 若 A_1，A_2，……，两两互不相容，则 $f_n(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} f_n(A_i)$

3、概率的公理化定义 （数学定义）

定义 3：设某试验的样本空间为Ω，对其中每个事件 A 定义一个实数 P（A），如果它满足下列三条公理：

1° P（A） ≥0（非负性）　　2° P（Ω）=1（规范性）

3° 若 A_1，A_2，……，A_n……两两互不相容，则 $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$（可列可加性，简称可加性）

则称 P（A）为 A 的概率

三 条件概率

§1 条件概率的概念及计算

在已知事件 B 发生条件下，事件 A 发生的概率称为事件 A 的条件概率，记为 P（A/B）。条件概率 P（A/B）与无条件概率 P（A）通常是不相等的。

例 1：某一工厂有职工 500 人，男女各一半，男女职工中非熟练工人分别为 40 人和 10 人，即该工厂职工人员结构如下：

人数	男	女	总和
非熟练工人	40	10	50
其他职工	210	240	450
总和	250	250	500

现从该厂中任选一职工，令 A= {选出的职工为非熟练工人}，B= {选出的职工为女职工}

显然，$P(A)=\frac{50}{500},P(B)=\frac{250}{500},P(AB)=\frac{10}{500}$；而

$$P(A/B)=\frac{10}{250}=\frac{10/500}{250/500}=\frac{P(AB)}{P(B)}，\quad P(B/A)=\frac{10}{50}=\frac{10/500}{50/500}=\frac{P(AB)}{P(A)}$$

定义 1 设 A、B 为两事件，如果 P（B）>0，则称 $P(A/B)=\frac{P(AB)}{P(B)}$ 为在事件 B 发生的条件下，事件 A 的条件概率。同样,如果 P（A）>0，则称 $P(B/A)=\frac{P(AB)}{P(A)}$ 为在事件 A 发生条件下，事件 B 的条件概率。

条件概率的计算通常有两种办法：

(1)由条件概率的*含义*计算（通常适用于古典概型），　(2)由条件概率的*定义*计算。

例 2：一盒子内有 10 只晶体管，其中 4 只是坏的，6 只是好的，从中无放回地取二次晶管，每次取一只，当发现第一次取得的是好的晶体管时，向第二次取的也是好的晶体管的概率为多少？

解：　令　A={第一次取的是好的晶体管}，B={第二次取的是好的晶体管}

按条件概率的含义立即可得：　$P(B/A)=\frac{5}{9}$

按条件概率的定义需先计算：$P(A)=\frac{6}{10}=\frac{3}{5},P(AB)=\frac{6\times5}{10\times9}=\frac{1}{3}$；于是 $P(B/A)=\frac{P(AB)}{P(A)}=\frac{1/3}{3/5}=\frac{5}{9}$

例 3：某种集成电路使用到 2000 小时还能正常工作的概率为 0.94,使用到 3000 小时还能正常工作的概率为 0.87 .有一块集成电路已工作了 2000 小时,向它还能再工作 1000 小时的概率为多大?

解：令　A={集成电路能正常工作到 2000 小时}，B={集成电路能正常工作到 3000 小时}

已知:：P(A)=0.94,　P(B)=0.87 且 $B\subset A$ ，既有 AB=B 于是 P(AB)=P(B)=0.87

按题意所要求的概率为：$P(B/A)=\frac{P(AB)}{P(A)}=\frac{0.87}{0.94}=0.926$

第二章 随机变量及其分布函数

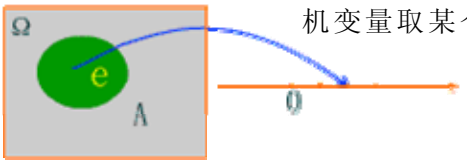
一 随机变量及其分布函数

§1 随机变量的概念

为了对各种各样不同性质的试验能以统一形式表示实验中的事件，并能将微积分等数学工具引进概率论。我们需引入随机变量的概念。

随机变量：设试验的样本空间为 Ω，在 Ω 上定义一个单值实函数 X=X(e), e∈Ω, 对试验的每个结果 e, X＝X(e)有确定的值与之对应。由于实验结果是随机的，那 X＝X（e）的取值也是随机的，我们便称此定义在样本空间 Ω 上的单值实函数 X＝X（e）为一个随机变量。

引进随机变量后，试验中的每个事件便可以通过此随机变量取某个值或在某范围内取值来表示了。（见图）



通俗讲，随机变量就是依照试验结果而取值的变量。

例 1 向靶子（见图）射击一次，观察其得分，规定

击中区域Ⅰ得 2 分

击中区域Ⅱ得 1 分

击中区域Ⅲ得 0 分

样本空间Ω＝｛Ⅰ，Ⅱ，Ⅲ｝。定义随机变量 X 表示射击一次的得分

即

于是，

$$X = X(e) = \begin{cases} 2, & e = \mathrm{I}, \\ 1, & e = \mathrm{II}, \\ 0, & e = \mathrm{III}. \end{cases}$$

$$A = \{\text{击中区域I}\} = \{e : X(e) = 2\} \overset{\text{简记}}{=\!=} \{X = 2\}$$

$$B = \{\text{中靶}\} = \{\text{击中区域I或击中区域II}\} = \{e : X(e) = 2\text{或}X(e) = 1\} \overset{\text{简记}}{=\!=} \{X = 2\text{或}X = 1\}$$

例 2 观察某电话交换台，在时间 T 内接到的呼唤次数。 样本空间Ω＝｛0，1，2，……｝。可定义随机变量 X 就表示在时间 T 内接到的呼唤次数。于是，

A＝｛接到呼唤次数不超过 1 0 次｝＝｛X≤1 0｝

B＝｛接到呼唤次数介于 5 至 1 0 次之间｝＝｛5≤X≤1 0｝ ,,

例 3 从一批灯泡中任取一个灯泡作寿命试验。观察所测灯泡的寿命（单位：小时） 样本空间Ω＝［0，＋∞］。可定义随机变量 X 表示所测得灯泡的寿命于是，

A＝｛测得灯泡寿命大于 5 0 0 （小时)｝＝｛X>500｝

B＝｛测得灯泡寿命不超过 5 0 0 0 （小时)｝＝｛X≤5 0 0 0｝。

不具明显数量性质的试验也可以定义随机变量表示试验中每个事件。

例 4 将一枚硬币上抛一次，观察正，反面出现的情况。 试验的样本空间Ω＝｛H， T｝，H－正面， T－反面。 可定义随机变量 X 表示上抛 1 次硬币正面出现的次数，即

$$X = X(e) = \begin{cases} 1, & e = H, \\ 0, & e = T, \end{cases}$$

于是， A＝｛出现正 面｝＝｛X＝1｝。 用随机变量表示事件常见形式有

$$\{X \le x\}$$

$$\{X > x\} = \overline{\{X \le x\}}$$

$$\{x_1 < X \le x_2\} = \{X \le x_2\} - \{X \le x_1\}$$

等等（这里 X 为随机变量， x， x_1， x_2 等为实数）

§2 分布函数

定义　设 X 为随机变量，对任意实数 x，则称函数 F（x）=P{X≤x} 为随机变量 X 的分布函数。

例 1　机房内有两台设备，令 X 表示某时间内发生故障的设备数，并知 P{X=0}=0.5， P{X=1}=0.3，P{X=2}=0.2，求 X 的分布函数 F（x）。

解：由于 X 的可能取值为 0，1，2 故应分情况讨论：

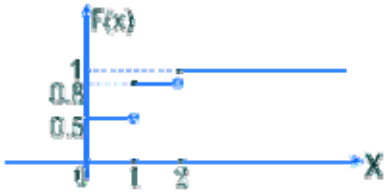
（1）　当 x<0 时，F（x）=P{X≤x}=0

（2）　当 0≤x<1 时，F（x）=P{X≤x}=P{X=0}=0.5

（3）　当 1≤x<2 时，F（x）=P{X≤x}=P{X=0}+P{X=1}=0.5+0.3=0.8

（4）　当 x≥2 时，F（x）=P{X≤x}=P{X=0}+P{X=1}+P{X=2}=0.5+0.3+0.2=1

总之，
$$F(x) = \begin{cases} 0, & x < 0 \\ 0.5, & 0 \le x < 1 \\ 0.8, & 1 \le x < 2 \\ 1, & x \ge 2 \end{cases}$$



例 2 向一半径为 2 米的圆形靶子射击，假设击中靶上任何一同心圆的概率为该同心圆的面积成正比，且每次射击必中靶。令 X 表示弹着点到靶心距离，求 X 的分布函数 F（x）。

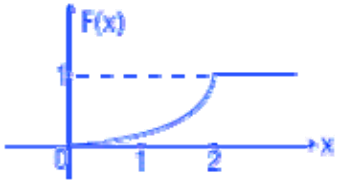
解： 当 x<0 时，F（x）=P{X≤x}=0

当 0≤x≤2 时，F（x）=P{X≤x}=P{击中半径为 x 的同心圆}=λπx^2

特别，当 x=2 时，1=F{2}=λπ4，解得 λ=1/4π，代入上式便得

$F(x)=\frac{x^2}{4}$　　当 x>2 时，F（x）=P{X≤x}=1

总之，$F(x)=\begin{cases}0, & x<0,\\ \frac{x^2}{4}, & 0\le x\le 2,\\ 1, & x>2,\end{cases}$



性质　1。F（x）是单调不减的，即对任意 $x_1<x_2$，有 F（x_1）≤F（x_2）；

2。0≤F（x）≤1 且 F（-∞）=0，F（+∞）=1；

3。F（x）为右连续的，即对任意 x，有 F（x+0）= F（x）。

可以证明（略）以上三条性质是分布函数所具有的三条基本共同特性。

利用分布函数可求随机变量落在某些区间上的概率，如

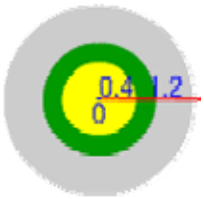
$$P\{X\le a\}=F(a)$$

$$P\{X>a\}=p\{\overline{X\le a}\}=1-P\{X\le a\}=1-F(a)$$

$P\{a<X\le b\}=P\{X\le b\}-P\{X\le a\}=F(b)-F(a)$ 等等。

例 3 在前面打靶的例子中，已知 X 表示弹着点到靶心距离，并求得其分布函数为

$$F(x)=\begin{cases}0, & x<0,\\ \frac{x^2}{4}, & 0\le x\le 2,\\ 1, & x>2,\end{cases}$$



于是便可以利用此分布函数，求出击中靶上环形区域（见图）的概率

$$P\{0.4<X\le 1.2\}=F(1.2)-F(0.4)=\frac{(1.2)^2}{4}-\frac{(0.4)^2}{4}=0.32$$

随机变量分类：

随机变量
- 离散型
- 非离散型
  - 连续型
  - 混合型
  - 奇异型等

二 离散型随机变量及其分布律

§1 离散型随机变量及其分布律的概念

定义：如果随机变量 X 的所有可能取值为有限个或可列个，则称随机变量 X 为离散型随机变量。

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

设 X 的所有可能取值为 x_1，x_2，……x_n，……，则称下列一组概率 $P\{X=x_i\}=\rho_i$，i=1,2,……，n,…… 为 X 的分布律。分布律也常常写成表格形式

性质： 1。$p_i \geqslant 0$, 一切 I； 2。$\sum_i p_i = 1$

例 1 设袋中装着分别标有-1，2，2，2，3，3 数字的六个球，现从袋中任取一球，令 X 表示取得球上所标的数字，求 X 的分布律。

解：X 的可能取值为 -1，2，3，且容易求得 $P\{X=-1\}=\frac{1}{6}, P\{X=2\}=\frac{3}{6}=\frac{1}{2}, P\{X=3\}=\frac{2}{6}=\frac{1}{3}$ 故 X 的分布律为

X	-1	2	3
p	1/6	1/2	1/3

例：相同条件下，独立的向目标射击 4 次，设每次击中目标的概率为 0.8，求击中目标次数 X 的分布律

解： X 的可能取值为 0，1，2，3，4 利用二项概率公式便可求得

$$P\{X=0\}=(0.2)^4=0.0016$$

$$P\{X=1\}=\binom{4}{1}(0.8)^1(0.2)^3=0.0256$$

$$P\{X=2\}=\binom{4}{2}(0.8)^2(0.3)^2=0.1536$$

$$P\{X=3\}=\binom{4}{3}(0.8)^3(0.3)^1=0.4096$$

$$P\{X=4\}=(0.8)^4=0.4096$$

X 的分布律为

X	0	1	2	3	4
p	0.0016	0.0256	0.1536	0.4096	0.4096

例 2 社会上定期发行某种奖券，每券一元，中奖率为 p，某人每次买 1 张奖券，如果没有中奖便继续买一张，直到中奖为止。求该人购买奖券次数 X 的分布律。如果中奖率为 1%，问他至少应买多少张奖券才能以不少于 99%的概率中奖。

解：（1） 令 A_i={第 i 次购买的奖券中奖}，i=1,2,……

且 $P(A_i)=P, P(\overline{A_i})=1-p, A_1, A_2, \cdots$ 是相互独立的。于是

$$P\{X=1\}=P(A_1)=p$$
$$P\{X=2\}=P(\overline{A_1}A_2)=P(\overline{A_1})P(A_2)=(1-p)p$$
$$P\{X=3\}=P(\overline{A_1}\,\overline{A_2}A_3)=P(\overline{A_1})P(\overline{A_2})P(A_3)=(1-p)^2p$$
$$\cdots\cdots\cdots\cdots$$
$$P\{X=i\}=P(\overline{A_1}\,\overline{A_2}\cdots\overline{A_{i-1}}A_i)=P(\overline{A_1})P(\overline{A_2})\cdots P(\overline{A_{i-1}})P(A_i)=(1-p)^{i-1}p$$
$$\cdots\cdots\cdots\cdots$$

X 的分布律为

X	1	2	3	……	i	……
p	p	(1-p)p	$(1-p)^2p$	……	$(1-p)^{i-1}p$	……

（2）设 n 为所需购买的奖券数，按题意 P{X≤n}≥99%

即 $P\{X\le n\}=\sum_{i=1}^{n}P\{X=i\}=\sum_{i=1}^{n}(1-p)^{i-1}p=p\cdot\frac{1-(1-p)^n}{1-(1-p)}$

$=1-(1-p)^n\ge 99\%$ 即 $(1-p)^n\le 0.01$ 即 $(0.99)^n\le 0.01$，解得

$n-1\ge\frac{\ln 0.01}{\ln 0.99}\approx 456.96$ 取 $n\ge 457$

例 4 某产品 40 件，其中有次品 3 件，现从中任取 3 件，（1）求取出的 3 件产品中所含次品数 X 的分布律；（2）求取出产品中至少有一件次品的概率；（3）求出 X 的分布函数 F（x），并作其图形。

解：（1）X 的可能取值为 0，1，2，3，且有

$$P\{X=0\}=\frac{\binom{37}{3}}{\binom{40}{3}}=0.7865 \quad P\{X=1\}=\frac{\binom{3}{1}\binom{37}{2}}{\binom{40}{3}}=0.2022 \quad P\{X=2\}=\frac{\binom{3}{2}\binom{37}{1}}{\binom{40}{3}}=0.0112$$

$$P\{X=3\}=\frac{\binom{3}{3}}{\binom{40}{3}}=0.0001$$

于是 X 的分布律为

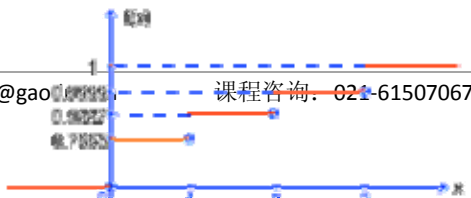
X	0	1	2	3
P	0.7865	0.2022	0.0112	0.0001

（2）任取 3 件产品中至少含有一件次品的概率为

P{X≥1}=P{X=1}+P{X=2}+P{X=3}=0.2022+0.0112+0.0001=0.2135 或 P{X≥1}=1－P{X＜1＝1－P{X=0}=1－0.7865=0.2135

（3）由分布函数定义不难求得 X 的分布函数为

$$F(x)=\begin{cases}0, & x<0\\ 0.7865, & 0\le x<1\\ 0.9887, & 1\le x<2\\ 0.9999, & 2\le x<2\end{cases}$$



离散型随机变量其分布函数的图形有如下特点：

(1)阶梯形；（2）仅在其可能取值处有跳跃；（3）其跃度为此随机变量在该处取值的概率。

一般，若 X 的分布律为 $P\{X=x_i\}=p_i, i=1,2,\cdots\cdots$，则 X 落在区间 I 内的概率便为 $P\{X\in I\}=\sum_{x_i\in I}p_i$

从而，X 的分布函数与分布律的关系便为 $F(x)=\sum_{x_i\le x}p_i$

X	0	1
p	q	p

三 连续型随机变量及其概率密度

§1 连续型随机变量及其概率密度的概念

所谓**连续型随机变量**是指此随机变量的可能取值至少应充满某个区间且其分布函数应当是连续的，连续型随机变量 X 有以下**特点**：

(1) **对任意实数 x，** $p\{X=x\}=0$ 事实上，$0\le p\{X=x\}=p\{x-\Delta x<X\le x\}=F(x)-F(x-\Delta x)\xrightarrow[\Delta x\to 0+0]{}F(x)-F(x)=0$；故 $p\{X=x\}=0$

(2) $p\{a<X\le b\}=p\{a\le X\le b\}=p\{a\le X<b\}=p\{a<X<b\}$

下面建立连续型随机变量 X 在实数 x 处的**概率密度（概念的引入）**

首先，考虑 X 落在区间 $(x,x+\Delta x]$ 内的概率 $p\{x<X\le x+\Delta x\}=F(x+\Delta x)-F(x)$

其次，求出 X 落在区间 $(x,x+\Delta x]$ 内的平均概率密度 $\frac{p\{x<X\le x+\Delta x\}}{\Delta x}=\frac{F(x+\Delta x)-F(x)}{\Delta x}$

最后，令 $\Delta x\to 0$ 便得到 X 在 x 处的概率密度 $\lim_{\Delta x\to 0}\frac{p\{x<X\le x+\Delta x\}}{\Delta x}=\lim_{\Delta x\to 0}\frac{F(x+\Delta x)-F(x)}{\Delta x}=F'(x)$

令 $f(x)=F'(x)\ge 0$，从而便有 $F(x)=\int_{-\infty}^{x}f(t)dt$

1. 定义 设 $F(x)$ 为随机变量 X 的分布函数，如果存在非负函数 $f(x)$ 使得对任意实数 x, 有

$F(x)=\int_{-\infty}^{x} f(t)dt$，**则称 X 为连续型随机变量，** $f(x)$ **为 X 的概率密度。**

性质 $1^{\circ}\ f(x)\geq 0$, 一切 x； $2^{\circ}\ \int_{-\infty}^{\infty} f(x)dx=1$

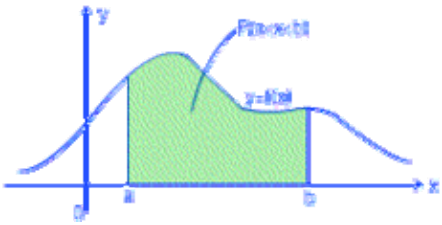
事实上由于 $F(x)=\int_{-\infty}^{x} f(t)dt$， $1=F(+\infty)=\int_{-\infty}^{\infty} f(t)dt$

2. 一个重要结果 $p\{a<x<b\}=\int_{a}^{b} f(x)dx$

事实上， $p\{a<x<b\}=p\{a<x\leq b\}=F(b)-F(a)=\int_{-\infty}^{b} f(x)dx-\int_{-\infty}^{a} f(x)dx$

$$=\int_{-\infty}^{a} f(x)dx+\int_{a}^{b} f(x)dx-\int_{-\infty}^{a} f(x)dx=\int_{a}^{b} f(x)dx$$

3. 几何解释



（1） $f(x)\geq 0$，表明密度曲线 $y=f(x)$ 在 x 轴上方；

（2） $\int_{-\infty}^{\infty} f(x)dx\geq 0$ 表明密度曲线 $y=f(x)$ 与 x 轴所夹图形的面积为 1；

（3） $p\{a<x<b\}=\int_{a}^{b} f(x)dx$ 表明 X 落在区间（a,b）内的概率等于以区间（a,b）为底，以密度曲线 $y=f(x)$ 为顶的曲边梯形面积。

4. 关 系 ： $F(x)=\int_{-\infty}^{x} f(t)dt$, $f(x)=F'(x)$ （在 $f($

例 1：已知连续型随机变量 X 的概率密度 $f(x)=\begin{cases} kx+1, 0\leq x\leq 2 \\ 0, \quad 其它 \end{cases}$ 为求系数 k 及分布函数 F（x），并计算概率 P{1.5<X<2.5}

解: (1) 因为 $\int_{-\infty}^{\infty} f(x)dx=1$ 故 $1=\int_{0}^{2}(kx+1)dx=\left(k\frac{x^2}{2}+x\right)\Big|_{0}^{2}=2k+2$ 解得 k =-1/2.

于是 X 的概率密度为 $f(x)=\begin{cases}-\frac{1}{2}x+1, 0\le x\le 2\\ 0, \quad 其它\end{cases}$

（2）当 $x<0$时，$F(x)=0$

当 $0\le x\le 2$时，$F(x)=\int_{-\infty}^{x}f(t)dt=\int_{0}^{x}\left(-\frac{1}{2}t+1\right)dt=\left(-\frac{1}{4}t^2+t\right)\Big|_0^x=-\frac{1}{4}x^2+x$

当 $x>2$时，$F(x)=\int_{-\infty}^{x}f(t)dt=\int_{0}^{2}\left(-\frac{1}{2}t+1\right)dt=1$

总之，$F(x)=\begin{cases}0, & x<0\\ -\frac{x^2}{4}+x, & 0\le x\le 2\\ 1, & x>2\end{cases}$

（3）$p\{1.5<x<2.5\}=F(2.5)-F(1.5)=1-\left(-\frac{(1.5)^2}{4}+1.5\right)=0.0625$

例 2.一种电子管的使用寿命为 X 小时，其概率密度为 $f(x)=\begin{cases}\frac{100}{x^2}, & x\ge 100\\ 0, & x<100\end{cases}$ 某仪器内装有三个这样电子管，试求使用 150 小时内只有一个电子管需要换的概率。

解：首先计算一个电子管使用寿命不超过 150 小时的概率，此概率为

$$p\{X\le 150\}=\int_{100}^{150}\frac{100}{x^2}dx=-\frac{100}{x}\Big|_{100}^{150}=1-\frac{100}{150}=\frac{1}{3}$$

令 Y 表示工作 150 小时内损坏的电子管数，则 $Y\sim B\left(3,\frac{1}{3}\right)$ *Y* 服从二项分布

于是，此仪器工作 150 小时内仅需要更换一个电子管的概率 $p\{Y=1\}=\binom{3}{1}\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^2=\frac{4}{9}=0.44$

§2 几个重要分布

1. 均匀分布　如果随机变量 X 的概率密度为 $f(x)=\begin{cases}\frac{1}{b-a}, & a\le x\le b\\ 0, & 其它\end{cases}$ 则称 X 在区间[a, b]上服从均匀分布，记为 X~U[a, b]；其分布函数为 $F(x)=\begin{cases}0, & x<a\\ \frac{x-a}{b-a}, & a\le x\le b\\ 1, & x>b\end{cases}$

实际背景：如果实验中所定义的随机变量 X 仅在一个有限区间[a,b]上取值，且在其内取值具有“等可能”性，则 X~U[a,b]。

例 2.某公共汽车从上午 7:00 起每隔 15 分钟有一趟班车经过某车站,即 7:00,7:15,7:30,…时刻有班车到达此车站,如果某乘客是在 7:00 至 7:30 等可能地到达此车站候车,问他等候不超过 5 分钟便能乘上汽车的概率。

解：设乘客于 7 点过 X 分钟到达车站，则 X~U[0,30],即其概率密度为 $f(x)=\begin{cases}\frac{1}{30}, & 0\le x\le 30\\ 0, & 其它\end{cases}$

于是该乘客等候不超过 5 分钟便能乘上汽车的概率为

$$p\{10\le X\le 15或25\le X\le 30\}=p\{10\le X\le 15\}+p\{25\le X\le 30\}$$
$$=\int_{10}^{15}\frac{1}{30}dx+\int_{25}^{30}\frac{1}{30}dx=\frac{5}{30}+\frac{5}{30}=\frac{1}{3}$$

2. 指数分布 如果随机变量 X 的概率密度为 $f(x)=\begin{cases}\lambda e^{-\lambda x}, & x\ge 0\\ 0, & x<0\end{cases}$ 其中 $\lambda>0$，则称 X 服从参数为 λ 的指数分布，记为 X~E(λ)，其分布函数为 $F(x)=\begin{cases}1-e^{-\lambda x}, & x\ge 0\\ 0, & x<0\end{cases}$

实际背景：在实践中，如果随机变量 X 表示某一随机事件发生所需等待的时间，则一般 X~E(λ)。

例如，某电子元件直到损坏所需的时间（即寿命）；随机服务系统中的服务时间；在某邮局等候服务的等候时间等等均可认为是服从指数分布。

例 3.设随机变量 X 服从参数为 λ=0.015 的指数分布，（1）求 $p\{X>100\}$；（2）若要使 $p\{X>x\}<0.1$，问 x 应当在哪个范围内？

解：由于 X~E(0.015)即其概率密度为 $f(x)=\begin{cases}0.015e^{-0.015x}, & x\geq 0\\ 0, & x<0\end{cases}$

于是，（1） $p\{X>100\}=\int_{100}^{+\infty}f(x)dx=\int_{100}^{+\infty}0.015e^{-0.015x}dx=\left(-e^{-0.015x}\right)\Big|_{100}^{+\infty}$
$=e^{-1.5}=0.223$

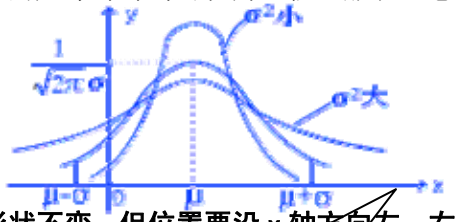
（2）要使 $p\{X>x\}<0.1$，即 $\int_{x}^{+\infty}f(t)dt=\int_{x}^{+\infty}0.015e^{-0.015t}dt=\left(-e^{-0.015t}\right)\Big|_{x}^{+\infty}=e^{-0.015x}<0.1$

取对数，便得 $-0.015x<\ln 0.1$ 于是便解得 $x>\frac{-\ln 0.1}{0.015}=153.5$

3. 正态分布（高斯分布）

△如果随机变量 X 的概率密度为 $f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. $-\infty<x<+\infty$ 其中 $\mu,\sigma^2(\sigma>0)$ 为常数，则称 X 服从参数 (μ,σ^2) 的正态分布，记为 X~N (μ,σ^2)。

△实际背景：在实践中，如果随机变量 X 表示许许多多均匀微小随机因素的总效应，则它通常将近似地服从正态分布，如：测量产生的误差；弹着点的位置；噪声电压；产品的尺寸等等均可认为近似地服从正态分布。



△正态密度曲线： 参数 μ,σ^2 对密度曲线的影响

(1) 当 σ^2 不变 μ 改变时，密度曲线 $y=f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ 形状不变，但位置要沿 x 轴方向左，右平移。（μ 实际上就是落在曲边梯形内部的平均概率）

(2) 当 μ 不变 σ^2 改变时， σ^2 变大，曲线变平坦； σ^2 变小，曲线变尖窄

△分布函数： $F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt

（积分是存在的，但是不能用初等函数表示）

△ 标准正态分布： 称 $\mu=0,\sigma^2=1$ 的正态分布 N(0,1) 为标准正态分布，其概率密度为 $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ ， $-\infty<x<+\infty$；分布函数为 $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt$ （其值有表可查）

公式 $\Phi(-x)=1-\Phi(x)$

变量替换，积分限变化

证：$\Phi(-x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-x}e^{-\frac{t^2}{2}}dt \overset{u=-t}{=} -\frac{1}{\sqrt{2\pi}}\int_{\infty}^{x}e^{-\frac{u^2}{2}}du=\frac{1}{\sqrt{2\pi}}\int_{x}^{+\infty}e^{-\frac{u^2}{2}}du$

$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{-\frac{u^2}{2}}du-\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{u^2}{2}}du=1-\Phi(x)$

例 5.设 X~ N(0,1) 求 $p\{-1<x<2\}$及$p\{|x|<1\}$

解：$p\{-1<x<2\}=\Phi(2)-\Phi(-1)=\Phi(2)-[1-\Phi(1)]=\Phi(2)+\Phi(1)-1=0.97725+0.8413-1=0.81855$

$p\{|x|<1\}=p\{-1<x<1\}=\Phi(1)-\Phi(-1)=\Phi(1)-[1-\Phi(1)]$
$=2\Phi(1)-1=2\times0.8413-1=0.6826$

例 6.设 X~N(0,1),要使 $p\{|x|\geq\lambda\}=0.05$, 问 λ 应为何值？

解：由于 $p\{|x|\geq\lambda\}=1-p\{|x|<\lambda\}=1-p\{-\lambda<x<\lambda\}$

$=1-[\Phi(\lambda)-\Phi(-\lambda)]=1-[2\Phi(\lambda)-1]=2-2\Phi(\lambda)=0.05$即 $\Phi(\lambda)=0.975$反查表，便得 $\lambda=1.96$

6. 一般正态分布与标准分布的关系：

若 $X\sim N(\mu,\sigma^2)$，其分布函数为 F(X)，则有 $F(x)=\Phi\left(\frac{x-\mu}{\sigma}\right)$

证：$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt \overset{u=\frac{t-\mu}{\sigma}}{=\!=\!=} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{x-\mu}{\sigma}}e^{-\frac{u^2}{2}}du$

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

7.正态变量落在区间内的概率：

如果 $X\sim N(\mu,\sigma^2)$ **则** $P\{a<x<b\}=\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)$

事实上，由 $F(x)=\Phi\left(\frac{x-\mu}{\sigma}\right)$ **立即可得** $P\{a<x<b\}=F(b)-F(a)=\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)$

例 7. 设 $X \sim N(2.3,4)$ 试求 $P\{2 < X < 4\}$

解： $p\{2 < x < 4\} = \Phi\left(\frac{4-2.3}{2}\right) - \Phi\left(\frac{2-2.3}{2}\right) = \Phi(0.85) - \Phi(-0.15) = \Phi(0.85) - [1 - \Phi(0.15)]$

$= 0.8023 + 0.5596 - 1 = 0.3619$

例 8 从某地乘车前往火车站搭火车，有两条路可走（1）走市区路程短，但交通拥挤，所需时间 $X_1 \sim N(50,100)$ ，(2)走郊区路程长，但意外阻塞少，所需时间 $X_2 \sim N(60,16)$ 。 问若有 70 分钟可用，应走哪条路线？

解：走市区及时赶上火车的概率为

$$P\{0 \le X_1 \le 70\} = \Phi\left(\frac{70-50}{10}\right) - \Phi\left(\frac{0-50}{10}\right) = \Phi(2) - \Phi(-5) \approx \Phi(2) = 0.97725$$

走郊区及时赶上火车的概率为 $P\{0 \le X_2 \le 70\} = \Phi\left(\frac{70-60}{4}\right) - \Phi\left(\frac{0-60}{4}\right)$ $= \Phi(2.5) - \Phi(-12.5) \approx \Phi(2.5) = 0.9938$；故应走郊区路线。

如果还有 65 分钟可用情况又如何呢？

同样计算，走市区及时赶上火车的概率为 $P\{0 \le X_1 \le 65\} = \Phi\left(\frac{65-50}{10}\right) - \Phi\left(\frac{0-50}{10}\right)$ $\approx \Phi(1.5) = 0.9332$

而走郊区及时赶上火车的概率便为 $P\{0 \le X_2 \le 65\} = \Phi\left(\frac{65-60}{4}\right) - \Phi\left(\frac{0-60}{4}\right)$ $\approx \Phi(1.25) = 0.8944$

此时便应改走市区路线。

四 随机变量函数的分布

§1 离散型随机变量的情况

所谓随机变量 X 的函数$Y=g(x)$是指 Y 也是一个随机变量，且每当 X 取值为 x 时， Y 的取值便为 $y=g(x)$

例如，车床车轴，若令 X 表示车出轴的直径，Y 表示车出轴的横断面积，则$Y=\frac{\pi}{4}X^2$

问题：已知 X 的分布，求$Y=g(X)$的分布。

例 1 设离散型随机变量 X 的分布律为

X	-1	0	1	2	5/2
P	2/10	1/10	1/10	3/10	3/10

求（1）Y=X-1，（2）$Y=-2X^2$ 的分布律

解：（1）由随机变量函数的概念便可由 X 的可能值求出 Y 的可能值，见下表：

Y=X-1	-2	-1	0	1	3/2
X	-1	0	1	2	5/2
P	2/10	1/10	1/10	3/10	3/10

于是便得 Y 的分布律

Y=X-1	-2	-1	0	1	3/2
P	2/10	1/10	1/10	3/10	3/10

（2）$Y=-2X^2$ 的可能值由下表给出

$Y=-2X^2$	-2	0	-2	-8	-25/2
X	-1	0	1	2	5/2
P	2/10	1/10	1/10	3/10	3/10

由于 Y 的值有相同的，即-2 ，因此应将其合并，相应的概率应按概率的可加性进行相加，即

$$P\{Y=-2\}=P\{X=-1,\text{或}X=1\}=P\{X=-1\}+P\{X=1\}=\frac{2}{10}+\frac{1}{10}=\frac{3}{10}$$

最后，得 Y 的分布律为

$Y=-2X^2$	-25/2	-8	-2	0
P	3/10	3/10	3/10	1/10

第三章 随机变量的数字特征

一 数学期望

§4.1.1 离散型随机变量的数学期望

在概率论和统计学中，一个离散性随机变量的期望值（或数学期望、或均值，亦简称期望）是试验中每次可能结果的概率乘以其结果的总和。换句话说，期望值是随机试验在同样的机会下重复多次的结果计算出的等同“期望”的平均值。需要注意的是，期望值并不一定等同于常识中的“期望”——“期望值”也许与每一个结果都不相等。（换句话说，期望值是该变量输出值的平均数。期望值并不一定包含于变量的输出值集合里。）

例 1：全班 40 名同学，其年龄与人数统计如下：

年龄	18	19	20	21	∑
人数	5	15	15	5	40

该班同学的平均年龄为：$\bar{a}=\frac{18\times5+19\times15+20\times15+21\times5}{40}=18\times\frac{5}{40}+19\times\frac{15}{40}+20\times\frac{15}{40}+21\times\frac{5}{40}=19.5$

若令 x 表示从该班同学中任选一同学的年龄，则 x 的分布律为

x	18	19	20	21
p	$\frac{5}{40}$	$\frac{15}{40}$	$\frac{15}{40}$	$\frac{5}{40}$

于是，x 取值的平均值，即该班同学年龄的平均值为

$$E(x)=\bar{a}=18\times\frac{5}{40}+19\times\frac{15}{40}+20\times\frac{15}{40}+21\times\frac{5}{40}=\sum_i x_i p_i=19.5$$

定义 1：设 x 为离散型随机变量，其分布律为 $P\{x=x_i\}=p_i,\ i=1,2,\cdots$ 如果级数 $\sum_i x_i p_i$ 绝对收敛，则此级数为 x 的数学期望（或均值）既为 E(X)，即 E(X)= $\sum_i x_i p_i$

意义：E(X)表示 X 取值的（加权）平均值

例 2：甲、乙射手进行射击比赛，设甲中的环数位 X1，乙中的环数为 X2，已知 X1 和 X2 的分布律分别为：

X_1	8	9	10
P	0.3	0.1	0.6

X_2	8	9	10
P	0.2	0.5	0.3

问谁的平均中环数高？

解：甲的平均中环数为 $E(X_1)=8\times0.3+9\times0.1+10\times0.6=9.3$

乙的平均中环数为 $E(X_2)=8\times0.2+9\times0.5+10\times0.3=9.1$

可见 $E(X_1) > E(X_2)$，即甲的平均中环数高于乙的平均中环数。

例 3：设 $X \sim \pi(\lambda)$，求 E(X)

解：由于 $X \sim \pi(\lambda)$，其分布律为 $P\{X=k\}=\frac{\lambda^k}{k!}e^{-\lambda}$，k=0,1,2…,所以

$$E(X)=\sum_k x_k p_k=\sum_{k=0}^{\infty}k\cdot\frac{\lambda^k}{k!}e^{-\lambda}=\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!}e^{-\lambda}=\lambda\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda}\quad(令m=k-1)=\lambda\sum_{m=0}^{\infty}\frac{\lambda^m}{m!}e^{-\lambda}=\lambda$$

例 4：一无线电台发出呼唤信号被另一电台收到的概率为 0.2，发方每隔 5 秒拍发一次呼唤信号，直到收到对方的回答信号为止，发出信号到收到回答信号之间需经 16 秒钟，求双方取得联系时，发方发出呼唤信号的平均数？

解：令 X 表示双方取得联系时，发方发出呼唤信号的次数。
X 的分布律为

X	4	5	6	7	… n …
P	0.2	0.8 × 0.2	$(0.8)^2\times0.2$	$(0.8)^3\times0.2$	$(0.8)^{n-4}\times0.2$

于是，双方取得联系时，发方发出的呼唤信号的平均数为

$$E(X)=\sum_{n=4}^{\infty}n\cdot(0.8)^{n-4}\times0.2=0.2\times(0.8)^{-3}\cdot\sum_{n=4}^{\infty}n\cdot(0.8)^{n-1}$$

由于 $\sum_{n=4}^{\infty}nx^{n-1}=\sum_{n=4}^{\infty}(x^n)'=(\sum_{n=4}^{\infty}x^n)'=(\frac{x^4}{1-x})'$，求导数 $=\frac{4x^3(1-x)+x^4}{(1-x)^2}=\frac{x^3(4-3x)}{(1-x)^2}$

将 x=0.8 代如上式，便得 $\sum_{n=4}^{\infty}n(0.8)^{n-1}=\frac{(0.8)^3(4-3\times0.8)}{(0.2)^2}=\frac{(0.8)^3\times1.6}{(0.2)^2}$

将此结果代入原式便得：$E(X)=0.2\times(0.8)^{-3}\frac{(0.8)^3\times1.6}{(0.2)^2}=\frac{1.6}{0.2}=8$（次）

§4.1.2 连续型随机变量的数学期望

定义：设X为连续型随机变量，其概率密度为$f(x)$,如果积分 $\int_{-\infty}^{+\infty}xf(x)dx$ 绝对收敛，则称此积分为 X 的数学期望，记为 E(X)，即 $E(X)=\int_{-\infty}^{+\infty}xf(x)dx$

例5：设$X\sim U[a,b]$，求$E(X)$

解：由于$X \sim U[a,b]$，即其概率密度$f(x)=\begin{cases}\frac{1}{b-a}, a\le x\le b\\ 0, \quad 其它\end{cases}$

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

例6：设$X\sim N(\mu,\sigma^2)$，求$E(X)$

解：由于$X\sim N(\mu,\sigma^2)$，即其概率密度为$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

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

$=\frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}te^{-\frac{t^2}{2}}dt+\mu\cdot\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{-\frac{t^2}{2}}dt=\frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{-\frac{t^2}{2}}d\frac{t^2}{2}+\mu=-\frac{\sigma}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\Big|_{-\infty}^{+\infty}+\mu=\mu$

例 7：设风速 V 是一个随机变量，且 V~U[0,a]，又设飞机的机翼上所受的压力 W 是风速 V 的函数：$W=kv^2$ 这里 a,k 均为已知正数。试求飞机机翼上所受的平均压力 E(W)。

解：首先求出W的概率密度$f_w(z)$（然后方可求出$E(w)=\int_{-\infty}^{+\infty}zf_w(z)dz$）

W 的分布函数为 $F_w(z)=P\{W\le z\}=P\{kV^2\le z\}=P\left\{-\sqrt{\frac{z}{k}}\le V\le\sqrt{\frac{z}{k}}\right\}(z>0)=F_v(\sqrt{\frac{z}{k}})-F_v(-\sqrt{\frac{z}{k}})$

两边求导，使得 $f_w(z)=F_w'(z)=F_v'(\sqrt{\frac{z}{k}})-F_v'(-\sqrt{\frac{z}{k}})=f_v(\sqrt{\frac{z}{k}})\frac{1}{2\sqrt{kz}}+f_v(-\sqrt{\frac{z}{k}})\frac{1}{2\sqrt{kz}}$

$=\frac{1}{2\sqrt{kz}}\left[f_v(\sqrt{\frac{z}{k}})+f_v(-\sqrt{\frac{z}{k}})\right]$ 由于$V\sim U[0,a]$，即其概率密度$f_v(v)=\begin{cases}\frac{1}{a}, 0\le v\le a\\ 0, \quad 其它\end{cases}$

故$f_v(-\sqrt{\frac{z}{k}})=0$，即有$f_w(z)=\frac{1}{2\sqrt{kz}}f_v(\sqrt{\frac{z}{k}})=\begin{cases}\frac{1}{2\sqrt{kz}}\cdot\frac{1}{a}, 0\le\sqrt{\frac{z}{k}}\le a\\ 0, \quad 其它\end{cases}=\begin{cases}\frac{1}{2a\sqrt{kz}}, 0\le z\le ka^2\\ 0, \quad 其它\end{cases}$

进而便可求得 W 的数学期望

$$E(W)=\int_{-\infty}^{+\infty}zf_w(z)dz=\int_0^{ka^2}z\cdot\frac{1}{2a\sqrt{kz}}dz(令z=kv^2 dz=2kvdv)$$

$=\int_0^a kv^2\frac{1}{2a\sqrt{k^2z^2}}2kvdv=\int_0^a kv^2\cdot\frac{1}{a}dv=\frac{1}{3}ka^2$

由此运算过程可以看到，不必求出 W 的概率密度 fw(z)，而根据 V 的概率密度 fv(v)也可直接求出 W 的数学期望值，即 $E(W)=E[kV^2]=\int_{-\infty}^{+\infty}kv^2f_v(v)dv=\int_0^a kv^2\cdot\frac{1}{a}dv=\frac{1}{3}ka^2$

随机变量函数的数学期望值

1.一维随机变量函数的数学期望

定理 1：设 X 为随机变量，Y=g(X)，

(1) 如果 X 为离散型随机变量，其分布律为 $P\{X=x_i\}=p_i, i=1,2,\cdots$**，且级数** $\sum_i g(x_i)p_i$ **绝对收敛，则有** $E(Y)=E[g(x)]=\sum_i g(x_i)p_i$

(2) 如果 X 为连续型随机变量，其概率密度为 $f(X)$，且积分 $\int_{-\infty}^{\infty} g(x)f(x)dx$ **绝对收敛，则有**

$$E(Y)=E[g(x)]=\int_{-\infty}^{\infty} g(x)f(x)dx$$

证略

例 8：已知 X 的分布律为

X	-1	0	1/2	1	2
P	1/3	1/6	1/6	1/12	1/4

求：$E(x^2)$

解：$E(x^2)=\sum_i x_i^2 p_i=(-1)^2\times\frac{1}{3}+0^2\times\frac{1}{6}+(\frac{1}{2})^2\times\frac{1}{6}+1^2\times\frac{1}{12}+2^2\times\frac{1}{4}=\frac{35}{24}$

例 9：设 $X\sim\pi(\lambda)$，求 $E(X^2)$

高等数学中级数的求和很关键！！！

解：$E(X^2)=\sum_k x_k^2 p_k=\sum_{k=0}^{\infty}k^2\cdot\frac{\lambda^k}{k!}e^{-\lambda}=\sum_{k=0}^{\infty}[k(k-1)+k]\frac{\lambda^k}{k!}e^{-\lambda}=\sum_{k=2}^{\infty}k(k-1)\frac{\lambda^k}{k!}e^{-\lambda}+\sum_{k=0}^{\infty}k\cdot\frac{\lambda^k}{k!}e^{-\lambda}$

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

例 10：设 $X\sim U[a,b]$，求 $E(X^2)$

解：由于 X 的概率密度为 于是 $f(x)=\begin{cases}\frac{1}{b-a}, a\le x\le b\\ 0, \quad 其它\end{cases}$

$$E(X^2)=\int_{-\infty}^{\infty}x^2 f(x)dx=\int_a^b x^2\cdot\frac{1}{b-a}dx=\frac{1}{b-a}\cdot\frac{x^3}{3}\Big|_a^b=\frac{b^3-a^3}{3(b-a)}=\frac{(b-a)(a^2+ab+b^2)}{3(b-a)}=\frac{a^2+ab+b^2}{3}$$

$X\sim U[2000,4000]$

例 11：国际市场上每年对我国某种商品的需求量为一个随机变量 X（单位：吨），且已知，　　并已知每售出一吨此种商品，可以为国家挣得外汇 3 万美元，但若售不出去，而屯售于仓库，每年需花费保养费每吨为一万美元，问应组织多少货源可使国家的平均收益达到最大？

解：设 a 为某年准备组织出口此种商品的数量（单位：吨）Y 为国家收益，于是 Y 是 X 的函数

$$Y=g(X)=\begin{cases}3a, X\ge a\\ 3X-(a-X), X\le a\end{cases}=\begin{cases}3a, X\ge a\\ 4X-a, X<a\end{cases}$$ 由于 $X\sim U[2000,4000]$，即其概率密度为

准备—实际需要=剩余

$$f(x)=\begin{cases}\frac{1}{2000}, 2000\le x\le 4000\\ 0, \quad 其它\end{cases}$$

于是国家的平均收益为

$$E(Y)=E[g(x)]=\int_{-\infty}^{+\infty}g(x)f(x)dx=\int_{2000}^{4000}g(x)\frac{1}{2000}dx$$

$$=\int_{2000}^{a}(4x-a)\frac{1}{2000}dx+\int_{a}^{4000}3a\frac{1}{2000}dx=\frac{1}{1000}[-a^2+7000a-4\times10^6]$$ 令

$$[E(Y)]'_a=\frac{1}{1000}[-2a+7000]=0$$

解得 a=3500（吨）

但 $[E(Y)]''_a=\frac{-2}{1000}<0$，故 E(Y)在 a=3500 时，E（Y）最大，即组织货源为 3500 吨时，可是国家的收益达到最大。

二 方差

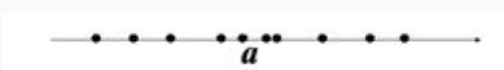
方差的概念

方差是各个数据与平均数之差的平方的平均数。在概率论和数理统计中，方差（英文 Variance）用来度量随机变量和其数学期望（即均值）之间的偏离程度。在许多实际问题中，研究随机变量和均值之间的偏离程度有着很重要的意义。

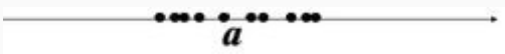
如下面的例子：

已知某零件的真实长度为 a，现用甲、乙两台仪器各测量 10 次，将测量结果 X 用坐标上的点表示如图：

甲仪器测量结果：



乙仪器测量结果：



两台仪器的测量结果的均值都是 a 。但是用上述结果评价一下两台仪器的优劣，很明显，我们会认为乙仪器的性能更好，因为乙仪器的测量结果集中在均值附近。

由此可见，研究随机变量与其均值的偏离程度是十分必要的。那么,用怎样的量去度量这个偏离程度呢?容易看到 E[|X-E[X]|]能度量随机变量与其均值 E(X)的偏离程度. 但由于上式带有绝对值,运算不方便，通常用量 E[(X-E[X])^2] 这一数字特征就是方差。

一般在计算式用下面公式进行计算

$X-E(X)\to\lvert X-E(X)\rvert\to[X-E(X)]^2\to E[X-E(X)]^2$
定义1：设X为随机变量，如果$E[X-E(X)]^2$存在，则称之为X的方差，记为 $D(X)$,即 $D(X)=E[X-E(X)]^2$

意义：D(X)表示 X 取值相对于平均值 E(X)的分散程度

方差的计算

1. 由方差定义直接计算

离散就求和
连续就积分

(1)若X为离散型随机变量，其分布律为$P\{X=x_i\}=p_i, i=1,2,\cdots$则

$$D(X)=E[X-E(X)]^2=\sum_i[x_i-E(X)]^2p_i$$

(2)若 X 为连续型随机变量，其概率密度为 $f(x)$，则 $D(X)=E[X-E(X)]^2=\int_{-\infty}^{+\infty}[x-E(X)]^2f(x)dx$

例1设$X\sim N(\mu,\sigma^2)$,求$D(X)$

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

$\left(令t=\frac{x-\mu}{\sigma},\quad x-\mu=\sigma t,\quad dx=\sigma dt\right)$

$$=\sigma^2\int_{-\infty}^{+\infty}t^2\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt=\frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}te^{-\frac{t^2}{2}}d\frac{t^2}{2}$$

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

分部积分法

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

2. 由下列重要公式计算

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

证：$D(X)=E[X-E(X)]^2=E\{X^2-2XE(X)+[E(X)]^2\}$

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

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

例 2. 设 $X\sim\pi(\lambda)$，求 $D(X)$

注意：记忆常见分布的数学期望和方差（最好都推导一遍）

解：前面已求得 $E(X)=\lambda, E(X^2)=\lambda^2+\lambda$ 于是 $D(X)=E(X^2)-[E(X)]^2=\lambda^2+\lambda-\lambda^2=\lambda$

例 3. 设 $X\sim U[a,b]$，求 $D(X)$

解：前面已求得 $E(X)=\frac{a+b}{2}$，$E(X^2)=\frac{a^2+ab+b^2}{3}$，于是

$$D(X)=E(X^2)-[E(X)]^2=\frac{a^2+ab+b^2}{3}-\frac{(a+b)^2}{4}=\frac{4a^2+4ab+4b^2-3a^2-6ab-3b^2}{12}$$

$$=\frac{a^2-2ab+b^2}{12}=\frac{(b-a)^2}{12}$$

方差的性质

性质1 若 c 为常数，则 $D(c)=0$　证：$D(c)=E[c-E(c)]^2=E[c-c]^2=E(0)=0$

性质2 若 c 为常数，X 为随机变量，则 $D(cX)=c^2D(X)$

证：$D(cX)=E[cX-E(cX)]^2=E\{c^2[X-E(X)]^2\}=c^2E[X-E(X)]^2=c^2D(X)$

性质3 如果 X 与 Y 相互独立，则 $D(X+Y)=D(X)+D(Y)$（注意：相加时期望没要求相互独立）

证：$D(X+Y)=E[X+Y-E(X+Y)]^2=E[(X-E(X))+(Y-E(Y))]^2$

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

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

$=D(X)+D(Y)$

其中 $E[X-E(X)][Y-E(Y)]=E[XY-XE(Y)-YE(X)+E(X)E(Y)]$

$=E(XY)-E(X)E(Y)-E(Y)E(X)+E(X)E(Y)$

$=E(XY)-E(X)E(Y)=0$，（X，Y 独立）

思考：如果二者独立
D(X-Y)=D(X)-D(Y)？
实际上 D(X-Y)=D(X)+D(Y)

推广：如果 n 个随机变量 $X_1,X_2,\cdots X_n$ 相互独立，则 $D\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}D(X_i)$

性质 4. 设 X 为随机变量，则 D(X)=0 的充分必要条件为 $P\{X=c\}=1$ 其中 c 为常数。

例 4. 设 X 为随机变量，E(X),D(X)存在，又设 $X^* = \frac{X - E(X)}{\sqrt{D(X)}}$， 证明 $E(X^*) = 0, D(X^*) = 1$

证：$E(X^*) = E\left[\frac{X - E(X)}{\sqrt{D(X)}}\right] = \frac{1}{\sqrt{D(X)}}[E(X) - E(X)] = 0$

$$D(X^*) = D\left[\frac{X - E(X)}{\sqrt{D(X)}}\right] = \frac{1}{D(X)}D[X - E(X)] = \frac{1}{D(X)}D(X) = 1$$

例 5. 设 X~B(n,p),求 E(X)，D(X)

解：设在贝努里试验中，事件 A 出现的概率为 p,将此贝努里试验独立重复进行几次，构成 n 重贝努里试验，令

$$X_i = \begin{cases} 1, & \text{第}i\text{次试验中}A\text{出现.} \\ 0, & \text{第}i\text{次试验中}A\text{不出现.} \end{cases} \quad i = 1,2,\cdots, n\text{且彼此独立}$$

显然 $X_i \sim B(1,p)$ 即其分布律为，

Xi	0	1
	p	q	p

i=1，2，…，n

$E(X_i) = p, E(X_i^2) = p$ 故 $D(X_i) = E(X_i^2) - [E(X_i)]^2 = p - p^2 = p(1-p) = pq$

另一方面，令 X 表示 n 重贝努里试验中事件 A 出现的次数，则 X~B(n,p) 且 $X = \sum_{i=1}^{n} X_i$ 于是

$$E(X) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} p = np \qquad D(X) = D\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} D(X_i) = \sum_{i=1}^{n} pq = npq$$

三 标准差

标准差（Standard Deviation），也称均方差（mean square error），是各数据偏离平均数的距离的平均数，它是离均差平方和平均后的方根，用 σ 表示。标准差是方差的算术平方根。标准差能反映一个数据集的离散程度。平均数相同的，标准差未必相同。

标准计算公式

假设有一组数值 X1,X2,X3,.....XN（皆为实数），其平均值为 μ，公式如图 1.

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

图 1

标准差也被称为标准偏差，或者实验标准差，公式如图 2。

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

图 2

简单来说，标准差是一组数据平均值分散程度的一种度量。一个较大的标准差，代表大部分数值和其平均值之间差异较大；一个较小的标准差，代表这些数值较接近平均值。

例如，两组数的集合 {0,5,9,14} 和 {5,6,8,9} 其平均值都是 7 ，但第二个集合具有较小的标准差。

标准差可以当作不确定性的一种测量。例如在物理科学中，做重复性测量时，测量数值集合的标准差代表这些测量的精确度。当要决定测量值是否符合预测值，测量值的标准差占有决定性重要角色：如果测量平均值与预测值相差太远（同时与标准差数值做比较），则认为测量值与预测值互相矛盾。这很容易理解，因为如果测量值都落在一定数值范围之外，可以合理推论预测值是否正确。

标准差应用于投资上，可作为量度回报稳定性的指标。标准差数值越大，代表回报远离过去平均数值，回报较不稳定故风险越高。相反，标准差数值越细，代表回报较为稳定，风险亦较小。

例如，A、B 两组各有 6 位学生参加同一次语文测验，A 组的分数为 95、85、75、65、55、45，B 组的分数为 73、72、71、69、68、67。这两组的平均数都是 70，但 A 组的标准差为 17.078 分，B 组的标准差为 2.16 分（此数据是在 R 统计软件中运行获得），说明 A 组学生之间的差距要比 B 组学生之间的差距大得多。

四 协方差及相关系数

协方差

1．协方差的概念

在概率论和统计学中，协方差用于衡量两个变量的总体误差。而方差是协方差的一种特殊情况，即当两个变量是相同的情况

期望值分别为 E(X) = μ 与 E(Y) = ν 的两个实数随机变量 X 与 Y 之间的协方差定义为：

COV(X，Y)=E[(X-E(X))(Y-E(Y))]

其中，E 是期望值。它也可以表示为：

直观上来看，协方差表示的是两个变量总体误差的方差，这与只表示一个变量误差的方差不同。

如果两个变量的变化趋势一致，也就是说如果其中一个大于自身的期望值，另外一个也大于自身的期望值，那么两个变量之间的协方差就是正值。

如果两个变量的变化趋势相反，即其中一个大于自身的期望值，另外一个却小于自身的期望值，那么两个变量之间的协方差就是负值。

如果 X 与 Y 是统计独立的，那么二者之间的协方差就是 0。

前面曾证明过：当 X，Y相互独立时，有

$E[X-E(X)][Y-E(Y)]=E(XY)-E(X)E(Y)=0$

定义 设X，Y为两个随机变量，如果 $E[X-E(X)][Y-E(Y)]$存在，则称它为

X与Y的协方差，记为$COV(X,Y)$，即：$COV(X,Y)=E[X-E(X)][Y-E(Y)]$

例1：设$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ 求 $COV(X,Y)$

解：由于 (X,Y)的概率密度为

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]}$$

于是 $cov(X,Y)=E(X-\mu_1)(Y-\mu_2)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-\mu_1)(y-\mu_2)f(x,y)dxdy$

$$=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-\mu_1)(y-\mu_2)e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]}dxdy$$

(令 $u=\frac{x-\mu_1}{\sigma_1},v=\frac{y-\mu_2}{\sigma_2}$)

$$=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\sigma_1\sigma_2 uve^{-\frac{1}{2(1-\rho^2)}\left[u^2-2\rho uv+v^2\right]}\sigma_1\sigma_2 dudv$$

$$=\frac{\sigma_1\sigma_2}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}vdv\int_{-\infty}^{+\infty}ue^{-\frac{1}{2(1-\rho^2)}\left[(u-\rho v)^2+(1-\rho^2)v^2\right]}du$$

$$=\frac{\sigma_1\sigma_2}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}ve^{-\frac{v^2}{2}}dv\int_{-\infty}^{+\infty}ue^{-\frac{1}{2(1-\rho^2)}\left[u-\rho v\right]^2}du$$

$$=\frac{\sigma_1\sigma_2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}ve^{-\frac{v^2}{2}}dv\int_{-\infty}^{+\infty}u\frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}}e^{-\frac{(u-\rho v)^2}{2(1-\rho^2)}}du\ .\ (N(\rho v,1-\rho^2))$$

$$=\frac{\rho\sigma_1\sigma_2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}v^2e^{-\frac{v^2}{2}}dv$$ (若 $\xi\sim N(0,1)$,则 $\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}v^2e^{-\frac{v^2}{2}}dv=E\xi^2=D\xi+(E\xi)^2=1$)

$=\rho\sigma_1\sigma_2$　　总之 $cov(X,Y)=\rho\sigma_1\sigma_2$

2. 协方差的性质

性质1.如果X，Y相互独立，则 $COV(X,Y)=0$

证：$COV(X,Y)=E[X-E(X)][Y-E(Y)]=E[XY-XE(Y)-YE(X)+E(X)E(Y)]$

$=E(XY)-E(X)\cdot E(Y)\underset{X,Y\text{独立}}{=}0$

性质2.如果a,b为常数，则 $COV(aX,bY)=abCOV(X,Y)$

证：$COV(aX,bY)=E[aX-E(aX)][bY-E(bY)]=abE[X-E(X)][Y-E(Y)]$
$=abCOV(a,b)$

性质3. $COV(X+Y,Z)=COV(X,Z)+COV(Y,Z)$

证：$COV(X+Y,Z)=E[X+Y-E(X+Y)][Z-E(Z)]=E[(X-EX)+(Y-EY)][Z-E(Z)]$
$=E[(X-EX)(Z-EZ)+(Y-EY)(Z-EZ)]=E[X-E(X)][Z-E(Z)]+E[Y-E(Y)][Z-E(Z)]$
$=COV(X,Z)+COV(Y,Z)$

3.一个重要公式　　$COV(X,Y)=E(XY)-E(X)E(Y)$

例 2：甲乙两人猜测箱中产品的数目，猜测结果分别记为 X 和 Y（单位：百个）已知（X，Y）的分布律和边缘分布律由下表给出：试求$COV(X,Y)$

X\Y	1	2	3	p_i
1	0.2	0.1	0.01	0.31
2	0.15	0.30	0.06	0.51
3	0.03	0.05	0.10	0.18
p_j	0.38	0.45	0.17	1

解：$E(X)=\sum_{i=1}^{3} ip_{i.}=1\times0.31+2\times0.51+3\times0.18=1.87$

$E(Y)=\sum_{j=1}^{3} jp_{.j}=1\times0.38+2\times0.45+3\times0.17=1.79$

$E(XY)=\sum_{i=1}^{3}\sum_{j=1}^{3} ijp_{ij}=1\times1\times0.20+1\times2\times0.10+1\times3\times0.01+2\times1\times0.15+2\times2\times0.30$

$+2\times3\times0.06+3\times1\times0.03+3\times2\times0.05+3\times3\times0.10=3.58$

于是得协方差 $COV(X,Y)=E(XY)-E(X)E(Y)=3.58-1.87\times1.79=0.2327$，在此例中，如果甲、乙两人猜测结果的单位选择为个，此时甲、乙猜测结果记为 X'，Y'，它们与原以百个为单位的猜测结果 X，Y 之间显有如下关系：$X'=100X$，$Y'=100Y$，此时由协方差的性质 2 可知，

$COV(X',Y')=COV(100X,100Y)=100^2COV(X,Y)=100^2\times0.2327=2327$

改进：$COV(X,Y)=E[X-E(X)][Y-E(Y)]\rightarrow\frac{COV(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=\frac{E[X-E(X)][Y-E(Y)]}{\sqrt{D(X)}\sqrt{D(Y)}}$

相关系数

1．相关系数的概念

相关系数又称线性相关系数.它是衡量变量之间线性相关程度的指标。样本相关系数用 r 表示,总体相关系数用 ρ 表示,相关系数的取值范围为[-1,1]。|r|值越大，误差 Q 越小，变量之间的线性相关程度越高；|r|值越接近 0，Q 越大，变量之间的线性相关程度越低。

定义 2.设 X，Y 为两个随机变量，且 $D(X)>0,D(Y)>0$,则称 $\rho_{XY}=\frac{COV(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$ 为X与Y的相关系数。

例 3：设$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma^2,\rho)$,求$\rho_{XY}$

解：由前面得到的结果可知 $E(X)=\mu_1,E(Y)=\mu_2,D(X)=\sigma_1^2,D(Y)=\sigma_2^2$，且

$COV(X,Y)=\rho\sigma_1\sigma_2$于是，$\rho_{XY}=\frac{COV(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=\frac{\rho\sigma_1\sigma_2}{\sigma_1\sigma_2}=\rho$

2． 相关系数的性质

（1）相关系数可正可负；

（2）相关系数的区间是[-1,1]，即｜ρxy｜≤1；

(3)具有对称性；即 X 与 Y 之间的相关系数（rXY）和 Y 与 X 之间的相关系数（rYX）；

(4)相关系数与原点和尺度无关；

（5）如果 X 与 Y 统计上独立，则它们之间的相关系数为零；但是 r=0 不等于说两个变量是独立的。即零相关并不一定意味着独立性；

（6）相关系数是线性关联或线性相依的一个度量，它不能用于描述非线性关系；

（7）虽然相关系数是两个变量之间的线性关联的一个度量，却不一定有因果关系的含义；

性质 1 如果 X，Y 相互独立，则 $\rho_{XY}=0$

性质 2 $|\rho_{XY}|\le 1$，即 $-1\le\rho_{XY}\le 1$

证：令 $X^*=\dfrac{X-E(X)}{\sqrt{D(X)}}$，$Y^*=\dfrac{Y-E(Y)}{\sqrt{D(Y)}}$

显然 $D(X^*)=1$，$D(Y^*)=1$ 且 $COV(X^*,Y^*)=E(X^*Y^*)-E(X^*)E(Y^*)=\dfrac{E[X-E(X)][Y-E(Y)]}{\sqrt{D(X)}\sqrt{D(Y)}}=\rho_{XY}$

（$E(X^*)=0, E(Y^*)=0$）

于是 $D(X^*\pm Y^*)=D(X^*)+D(Y^*)\pm 2COV(X^*,Y^*)$

$=1+1\pm 2\rho_{XY}=2\pm 2\rho_{XY}=2(1\pm\rho_{XY})$

因为方差 $D(X^*\pm Y^*)\ge 0$ 故 $1\pm\rho_{XY}\ge 0$，即 $|\rho_{XY}|\le 1$

X	-1	0	1
P	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

相关系数为 0，能否说二者无关了？NO

例 4：设 X 的分布律为

$Y=X^2$，求 ρ_{XY}

解：$E(X)=(-1)\times\frac{1}{3}+0\times\frac{1}{3}+1\times\frac{1}{3}=0$ $E(Y)=E(X^2)=(-1)^2\times\frac{1}{3}+0^2\times\frac{1}{3}+1^2\times\frac{1}{3}=\frac{2}{3}$

$E(X^2)=E(Y)=\frac{2}{3}$ $E(Y^2)=E(X^4)=(-1)^4\times\frac{1}{3}+0^4\times\frac{1}{3}+1^4\times\frac{1}{3}=\frac{2}{3}$

$E(XY)=E(X^3)=(-1)^3\times\frac{1}{3}+0^3\times\frac{1}{3}+1^3\times\frac{1}{3}=0$

于是 $D(X)=E(X^2)-[E(X)]^2=\frac{2}{3}-0^2=\frac{2}{3}>0$

$D(Y)=E(Y^2)-[E(Y)]^2=\frac{2}{3}-\left(\frac{2}{3}\right)^2=\frac{2}{9}>0$ 而

$COV(X,Y)=E(XY)-E(X)E(Y)=0-0\times\frac{2}{3}=0$

所以 $\rho_{xy}=\frac{COV(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=\frac{0}{\sqrt{\frac{2}{3}}\sqrt{\frac{2}{9}}}=0$

例5 设随机变量 $\hat{\theta}\sim U[0,2\pi]$ 又设 $X=\cos\hat{\theta}, Y=\cos\left(\hat{\theta}+a\right)$,(其中a为 1/2Pi) 试求 ρ_{XY}

解：由于 $\hat{\theta}\sim U[0,2\pi]$ 即 H 的概率密度为 $f(\theta)=\begin{cases}\frac{1}{2\pi} & 0\le\theta\le 2\pi\\ 0, & \text{其它}\end{cases}$

于是便得 $E(X)=E\left[\cos\hat{\theta}\right]=\int_{-\infty}^{+\infty}\cos f(\theta)d\theta=\int_0^{2\pi}\cos\theta\ \frac{1}{2\pi}d\theta=\frac{1}{2\pi}\sin\theta\Big|_0^{2\pi}=\frac{1}{2\pi}\sin(\theta+a)\Big|_0^{2\pi}=0$

$E(Y)=E\left[\cos\left(\hat{\theta}+a\right)\right]=\int_{-\infty}^{+\infty}\cos(\theta+a)f(\theta)d\theta=\frac{1}{2\pi}\int_0^{2\pi}\cos(\theta+a)d\theta$

$E(X^2)=E\left[\cos^2\hat{\theta}\right]=\int_{-\infty}^{+\infty}\cos^2\theta f(\theta)d\theta=\frac{1}{2\pi}\int_0^{2\pi}\frac{1+\cos 2\theta}{2}d\theta$

$=\frac{1}{2\pi}\left[\int_0^{2\pi}\frac{1}{2}d\theta+\frac{1}{2}\int_0^{2\pi}\cos 2\theta d\theta\right]=\frac{1}{2\pi}\left[\pi+\frac{1}{2}\sin 2\theta\Big|_0^{2\pi}\right]=\frac{1}{2}$

同样方法可得 $E(Y^2)=E\left[\cos^2(\theta+a)\right]=\frac{1}{2}$

$E(XY)=E\left[\cos\hat{\theta}\cos\left(\hat{\theta}+a\right)\right]=\int_{-\infty}^{+\infty}\cos\theta\cos(\theta+a)f(\theta)d\theta$

$=\frac{1}{2\pi}\int_0^{2\pi}\cos\theta\cos(\theta+a)d\theta=\frac{1}{4\pi}\int_0^{2\pi}\left[\cos(2\theta+a)+\cos a\right]d\theta$

$=\frac{1}{4\pi}\left[\int_0^{2\pi}\cos(2\theta+a)d\theta+\int_0^{2\pi}\cos a d\theta\right]=\frac{1}{4\pi}\left[0+2\pi\cos a\right]=\frac{1}{2}\cos a$

进而得 $D(X)=E(X^2)-[E(X)]^2=\frac{1}{2}-0^2=\frac{1}{2}$

$D(Y)=E(Y^2)-[E(Y)]^2=\frac{1}{2}-0^2=\frac{1}{2}$

$COV(X,Y)=E(X,Y)-E(X)E(Y)=\frac{1}{2}\cos a$

最后得 $\rho_{xy}=\frac{COV(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=\frac{\frac{1}{2}\cos a}{\sqrt{\frac{1}{2}}\cdot\sqrt{\frac{1}{2}}}=\cos a$

问题：相关系数到底说明什么问题？
似乎并不能完全反映两个变量的相关程度。
由此问题引出性质 3
相关系数实际上叫“线性相关系数”更准确

讨论如下：

（1） 当 $a=\frac{\pi}{2}$ 时，$\rho_{XY}=\cos\frac{\pi}{2}=0$，此时 $X=\cos\hat{\theta}, Y=\cos\left(\hat{\theta}+\frac{\pi}{2}\right)=-\sin\hat{\theta}$，从而有 $X^2+Y^2=1$

积变偶不变，符号看象限

（2） 当 $a=\theta$ 时，$\rho_{XY}=\cos\theta=1$，此时 $X=\cos\hat{\theta}, Y=\cos\hat{\theta}$，从而有 $Y=X$。

（3） 当$\alpha=\pi$时，$\rho_{XY}=\cos\pi=-1$，此时 $X=\cos\theta, Y=\cos(\theta+\pi)=-\cos\theta$

即有$Y=-X$。

性质 3 $\rho_{XY}=\pm1$的充分必要条件为$P\{Y=aX+b\}=1$，其中a，b为常数。

证：令$X^*=\dfrac{X-E(X)}{\sqrt{D(X)}}$，$Y^*=\dfrac{Y-E(Y)}{\sqrt{D(Y)}}$显然有$E(X^*)=0$，$E(Y^*)=0$

$D(X^*)=1, D(Y^*)=1$且$\rho_{XY}=E(X^*Y^*)$由于

$$1-\rho_{XY}^2=1-2\rho_{XY}\rho_{XY}+\rho_{XY}^2=E(Y^{*2})-2\rho_{XY}E(X^*Y^*)+\rho_{XY}^2E(X^{*2})$$

$$=E\left[Y^{*2}-2\rho_{XY}X^*Y^*+(\rho_{XY}X^*)^2\right]=E[Y^*-\rho_{XY}X^*]^2$$

$$=D(Y^*-\rho_{XY}X^*)\ (\because E(Y^*-\rho_{XY}X^*)=0)$$

可见$|\rho_{XY}|=1\Leftrightarrow D(Y^*-\rho_{XY}X^*)=0\Leftrightarrow P\{Y^*-\rho_{XY}X^*=C\}=0$

即 $P\left\{\dfrac{Y-E(Y)}{\sqrt{D(Y)}}-\rho_{XY}\dfrac{X-E(X)}{\sqrt{D(X)}}=C\right\}=0$，亦即

$$P\left\{Y=\frac{\sqrt{D(Y)}}{\sqrt{D(X)}}\rho_{XY}X+E(Y)-\frac{\sqrt{D(Y)}}{\sqrt{D(X)}}\rho_{XY}E(X)+C\sqrt{D(Y)}\right\}=1$$

即 $P\{Y=aX+b\}=1$。

定义3：如果$\rho_{XY}=0$，则称X与Y是线性不相关的，简称不相关。

而当$\rho_{XY}=$时，称X与Y是线性正相关，当$\rho_{XY}=-$时而称为负相关。

注意：X，Y独立$\Leftrightarrow$X，Y不相关

然而，当$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$时，X，Y独立 $\Leftrightarrow$ X，Y不相关。这是因为，由前面所得 结果可知，此时 $\rho_{XY}=\rho$且X，Y独立 $\Leftrightarrow\rho=0$，于是X，Y独立的重要条件是 $\rho_{XY}=0$，即X，Y不相关。