

# 第4章 随机变量的数字特征

## Chpt. 4 Digital Characteristics of Random Variables



# 为什么研究数字特征？

## 第二、三章：

引入随机变量把概率研究抽象化

引入分布函数使之可以借鉴已有的数学工具

[1]分布函数全面反映随机变量的性质，但在具体应用中，我们通常不需要了解得这么全面。**随机变量是否有一些特征，比较突出地反映其性质？**如取所有值的平均情况， $X$ 与均值的偏离情况。

例：研究生入学考试成绩

- ▣ 限定总分，由总分可以知道多科的平均分；
- ▣ 但是同时限定单科的最低分；
- ▣ 平均值高、偏离均值小的才是好的。

## 为什么研究数字特征？

[2] 当我们比较不同的随机变量时，分布函数或分布律并不可用，以分布律而言，它只说明变量取某值的可能性，是概率特性，从概率的角度我们不能说一个变量比另一个变量如何。

**Example 4.1** 设系统L由两个相互独立的子系统 $L_1, L_2$ 链接而成，连接的方式分别为（1）串联（2）并联（3）备用

$$\text{串联} \min\{X, Y\}, \quad f_1(z) = (\alpha + \beta)e^{-(\alpha + \beta)z}$$

$$\text{并联} \max\{X, Y\}, \quad f_2(z) = \alpha e^{-\alpha z} + \beta e^{-\beta z} - (\alpha + \beta)e^{-(\alpha + \beta)z}$$

$$\text{备用} X+Y, \quad f_3(z) = \frac{\alpha\beta}{\beta - \alpha} [e^{-\alpha z} - e^{-\beta z}]$$

直观上看，应该是系统寿命：备用 > 并联 > 串联；

但是，由于是随机的，一个具体的备用模式的寿命比其他还短



### Example 3.19

设系统  $L$  由两个相互独立的子系统  $L_1, L_2$  连接而成, 连接的方式分别为 (1) 串联; (2) 并联; (3) 备用 (当系统之一损坏时, 另一个开始工作)。设  $L_1, L_2$  的寿命分别为  $X, Y$ , 已知他们的概率密度分别服从指数分布

$$f_X(x) = \begin{cases} \alpha e^{-\alpha x} & x > 0 \\ 0 & x \leq 0 \end{cases} \quad f_Y(y) = \begin{cases} \beta e^{-\beta y} & y > 0 \\ 0 & y \leq 0 \end{cases}$$

其中  $\alpha > 0, \beta > 0$ 。

**求:** 三种情况下, 系统  $L$  寿命  $Z$  的概率分布。

**解:**

$$F_X(x) = \begin{cases} 1 - e^{-\alpha x} & x > 0 \\ 0 & x \leq 0 \end{cases} \quad F_Y(y) = \begin{cases} 1 - e^{-\beta y} & y > 0 \\ 0 & y \leq 0 \end{cases}$$



# 为什么研究数字特征？

[3] 根据经验，某些随机现象的随机变量服从某类分布，它们的一些参数可由某些数字特征确定。对这些随机现象，数字特征有更重要的意义。

E.g. 正态分布  $N(\mu, \sigma^2)$ ，二项分布  $B(n, p)$ ，泊松分布  $\pi(\lambda)$

## 主要的数字特征

- **数学期望**：描述一个随机变量的平均水平（即均值）
- **方差**：描述一个随机变量的相对于均值的离散程度
- **相关系数**：描述两个随机变量间线性关系的密切程度

## 4.1 随机变量的数学期望

### 分赌本问题

甲、乙二人各出同等数目的赌注，比方说1000元，然后进行博弈。每一局甲胜和乙胜都有同等概率，即0.5。二人约定：谁先胜满 $m$ 局（如在7盘棋中， $m=4$ ），谁就取走全部赌注2000元。

到某时为止，甲胜了 $a$ 局，而乙胜 $b$ 局， $a$ 、 $b$ 都比 $m$ 小，而二人因故要中止博弈，问这2000元赌注该如何分才算公平。

分析：

- [1] 如果 $a=b$ ，则二人态势一样，平分赌注不会有异议；
- [2] 若 $a$ 、 $b$ 不等，例如 $a>b$ ，则甲理应多分一些才公平。

但具体的比例该如何才算公平？



## 4.1 随机变量的数学期望

### 若干不同的解法

[1] 帕西奥利（1494年）

按 $a:b$ 的比例分配

[2] 塔泰格利亚（1556年）

按 $m+a-b:m-a+b$ 的比例分配。

同时，他又指出，此问题最好交由法官判决。

[3] 法雷斯泰尼（1603年）

按 $2m-1+a-b:2m-1-a+b$ 的比例分配

[4] 卡丹诺（1539年）

记 $u=m-a$ ， $v=m-b$ ，其中 $u$ 、 $v$ 是甲、乙最终取胜尚需的胜局数，若 $a>b$ ，则 $u<v$ 。

按 $v(v+1):u(u+1)$ 的比例分配赌注。

## [5] 巴斯卡 (1654年)

假这两个人继续赌下去, 记 $u=m-a$ ,  $v=m-b$

则至多不超过 $u+v-1$ 局, 就会最后见分晓。

甲、乙最后获胜都有一定的概率, 分别记为 $p$ 和 $q$  ( $p+q=1$ )。

当  $a>b$  时必然  $p>q$ 。算出 $p$ 和 $q$ , 巴斯卡主张赌注应按 $p:q$ 的比例去分配。

一般情况下:

比赛 $2m-1$ 盘, 先胜 $m$ 盘者获胜; 进行了 $a+b$ 局后终止比赛时, 只要没有哪位选手胜了 $m$ 盘, 后面理论上最多进行  $(2m-1) - (a+b) = u+v-1$ 盘比赛。





## [5] 巴斯卡（1654年）

假设  $a = 2$ ,  $b = 1$ ,

例如：设  $m = 4$  ( $2m - 1 = 7$ )，则理论上后4局还需要赌。

这4局的结果有16种可能，而甲获胜的可能为： $C_4^2 + C_4^3 + C_4^4 = 11$

故甲、乙最终获胜的概率分别为  $11 / 16$  和  $5 / 16$ 。

赌本应按11：5的比例分给甲和乙。

例如：设  $m = 5$  ( $2m - 1 = 9$ )，则理论上后6局还需要赌。

6局的结果有64种可能，而甲获胜的可能为：

$$C_6^3 + C_6^4 + C_6^5 + C_6^6 = 42$$

故甲、乙最终获胜的概率分别为  $42 / 64$  和  $22 / 64$ 。

赌本应接21：11的比例分给甲和乙。

[5] 巴斯卡（1654年）

为什么按最终获胜的概率之比去分配是公平的？

二人按各自的期望所得去分配应是合理的。各人的“期望所得”：

在最后4局未赌之前，甲有可能获得2000元，但机会只有 $11/16$ ，因而他的“期望所得”只有 $2000 \times (11/16)$ 元=1375元。

同样，乙的“期望所得”为 $2000 \times (5/16)$ 元=625元。

期望所得之比，与获胜概率之比是一回事，所以按最终获胜的概率之比去分配是公平的。

由此，引出了期望(Expectation)概念

## 4.1 随机变量的数学期望

**Example 4.2** 为评价甲的射击技术，随机观察甲的10次射击，统计各次击中环数 $x_k$ 和频数 $v_k$ 如下表，求他每次射击击中的平均环数. (其中 $N=\sum v_k=10$ ).

|                 |     |     |     |
|-----------------|-----|-----|-----|
| 击中环数 $x_k$      | 8   | 9   | 10  |
| 频数 $v_k$        | 2   | 5   | 3   |
| 频 率 $f_k=v_k/N$ | 0.2 | 0.5 | 0.3 |

**解** 这是一般求平均数  $\bar{x}$  问题，此时

$$\begin{aligned}\bar{x} &= (8 \times 2 + 9 \times 5 + 10 \times 3) / 10 \\ &= 8 \times \frac{2}{10} + 9 \times \frac{5}{10} + 10 \times \frac{3}{10} \\ &= 9\end{aligned}$$



## 4.1 随机变量的数学期望

写成一般形式就是

$$\bar{x} = \left( \sum_k x_k v_k \right) / N = \sum_k x_k v_k / N = \sum_k x_k f_k$$

若要全面考察甲的射击技术，仅凭这10次观察是不够的，因为频率  $v_k / N = f_k$  是与该次射击(试验)的结果有关的。

观察次数N不断增多，由于频率稳定于概率  $p_k$ ，和式  $\sum_k x_k f_k$  就趋于稳定

$$\sum_k x_k f_k \longrightarrow \sum_k x_k p_k$$

它是一个确定值，不依赖于具体试验，应该更能表示甲的射击水平。

## 4.1 随机变量的数学期望

### 4.1.1 数学期望定义

**[定义]** 设离散型随机变量  $X$  的分布列为

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_k & \cdots \\ p_1 & p_2 & \cdots & p_k & \cdots \end{pmatrix}$$

如果级数  $\sum x_k p_k$  绝对收敛, 即  $\sum_k |x_k| p_k < \infty$ , 就称此级数的和为  $X$  的数学期望 (mathematical expectation) 或均值 (mean), 记作  $E(X)$ 。



## 4.1 随机变量的数学期望

**Remark:** 条件  $\sum_k |x_k| p_k < \infty$  表示  $E(X) = \sum x_k p_k$  不受求和次序的影响, 从而  $E(X)$  是一个确定的值 (这是需要的, 因实际情况中事件出现可能无次序的)。当条件不满足时, 就称  $X$  的数学期望不存在。

**例** 设随机变量  $X$  取值为  $(-1)^k 2^k / k$  ( $k = 1, 2, \dots$ ), 取得各值的概率为  $P\{X = (-1)^k 2^k / k\} = \frac{1}{2^k}$

由  $\sum_{k=1}^{\infty} |x_k| p_k = \sum_{k=1}^{\infty} \frac{1}{k} = \infty$  (调和级数),  $\sum_{k=1}^{\infty} x_k p_k = \sum_{k=1}^{\infty} (-1)^k \frac{1}{k} < \infty$  (莱布尼兹级数), 仍说  $E(X)$  不存在

有限项的和不受次序变动的影响, 故有限个可能值的随机变量的数学期望总是存在的

**Example 4.3** 退化分布 $P\{X=a\}=1$ 的数学期望  $E(X)=a$ , 也即常数的数学期望就是它本身.

**Example 4.4** 二项分布 $X \sim B(n, p)$ 的数学期望

$$\begin{aligned} E(X) &= \sum_{k=0}^n k C_n^k p^k (1-p)^{n-k} = \sum_{k=0}^n k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^n \frac{(n-1)!}{(k-1)![(n-1)-(k-1)]!} p^{k-1} (1-p)^{(n-1)-(k-1)} \\ &= np \sum_{r=0}^{n-1} \frac{(n-1)!}{r!(n-1-r)!} p^r (1-p)^{(n-1)-r} \\ &= np(p + [1-p])^{n-1} \\ &= np \end{aligned}$$

**Remark 1:**  $n$  次试验中出现的平均数是 $np$ 。

**Remark 2:** 当 $n=1$ 时, 两点分布 (0—1分布)的数学期望为 $p$ 。



## 4.1 随机变量的数学期望

**Example 4.5** 泊松分布  $P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda}$  ( $k = 0, 1, 2, \dots$ ) 的数学期望

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

故泊松分布中参数 $\lambda$ 就是它的数学期望.





## 4.1 随机变量的数学期望

**Example 4.6** 几何分布  $P\{X = k\} = p(1-p)^{k-1} (k = 1, 2, \dots)$  的数学期望。

$$E(X) = \sum_{k=1}^{\infty} kp(1-p)^{k-1}$$

记  $q=1-p$

$$E(x) = p \frac{d}{dq} \left[ \sum_{k=1}^{\infty} q^k \right]$$

$$= p \frac{d}{dq} \left[ \frac{q}{1-q} \right]$$

$$= p \frac{(1-q) + q}{(1-q)^2}$$

$$= \frac{1}{p}$$



**Example 4.7** 在一个人数很多的单位中普查某疾病,  $N$ 个人验血, 可用两种方法:

- (1) 每个人化验一次, 共需化验 $N$ 次;
- (2)  $k$ 个人的血混合化验, 如果结果是不含该病菌, 说明这 $k$ 个人都无该病, 这样 $k$ 个人化验一次即可; 如果结果是含该病菌, 则该组每个人再分别化验一次,  $k$ 个人共化验  $(k+1)$  次.

试说明当 $p$ 较小时, 选取适当的 $k$ , 按第二种方法可以减少化验次数, 并说明 $k$ 取什么值时最适宜.

**解** 记用第二种方法化验时, 分 $N$ 人为 $(N/k)$  组。一组 $k$ 人所需化验次数为 $X$ .

第一种情况是 $k$ 个人化验一次, 则一组人化验 $X=1$ 次的概率为

$$P\{X = 1\} = (1 - p)^k$$

## 4.1 随机变量的数学期望

第二种情况是k个人化验 (k+1) 次, 其概率为

$$P\{X = k + 1\} = 1 - (1 - p)^k$$

|   |           |               |
|---|-----------|---------------|
| X | 1         | k+1           |
| p | $(1-p)^k$ | $1 - (1-p)^k$ |

$$\begin{aligned}\text{所以 } E(X) &= (1-p)^k + (k+1)[1 - (1-p)^k] \\ &= k + 1 - k(1-p)^k\end{aligned}$$

N个人分成N/k组所需的平均化验次数为

$$\frac{N}{k} E(X) = N \left[ 1 - (1-p)^k + \frac{1}{k} \right]$$



由此只要选择k使得下式成立，分组就比不分组节省次数。

$$N \left[ 1 - (1 - p)^k + \frac{1}{k} \right] < N$$

考虑N充分大，则使得  $L = 1 - (1 - p)^k + \frac{1}{k} < 1$ ，要  $k(1 - p)^k > 1$ 。

注意到p较小时，1-p 较大，容易取到 k 使  $k(1 - p)^k > 1$ 。

如果p=0.1

| k                     | 1   | 2    | 3      | 4      | 5       | 6      | 7      |
|-----------------------|-----|------|--------|--------|---------|--------|--------|
| $k(1 - p)^k$          | 0.9 | 1.62 | 2.187  | 2.6244 | 2.95425 | 3.1886 | 3.3480 |
| $1 - (1 - p)^k + 1/k$ | 1.1 | 0.69 | 0.6043 | 0.5939 | 0.60951 | 0.6352 |        |

可见取 k=4 最好。如果N=1000，第二种方法平均化验数为

$$N \left[ 1 - (1 - p)^k + \frac{1}{k} \right] \approx 594$$

## 连续型随机变量的数学期望

连续型随机变量 $X$ ，其密度函数为 $f(x)$ . 其数学期望的定义可借鉴离散型情形的定义及普通积分的导出过程来引入.

**STEP 1:** 先设 $X$ 只在有限区间 $[a, b]$ 上取值，将 $[a, b]$ 作分割：

$$a = x_0 < x_1 < \cdots < x_n = b$$

$X$ 落在各小段的概率

$$P\{x_k < X \leq x_{k+1}\} = \int_{x_k}^{x_{k+1}} f(x)dx \approx f(x_k)\Delta x_k$$

近似地视为落在一点上 $x_k$ 的概率

**STEP 2:** 与它相应的离散型随机变量的数学期望  $\sum x_k f(x_k)\Delta x_k$

**STEP 3:** 令 $n \rightarrow \infty$ ,  $\max\{\Delta x_k\} \rightarrow 0$ ，则和式的极限为积分.

$$\int_a^b xf(x)dx$$



## 连续型随机变量的数学期望

自然把它作为 $X$ 的数学期望. 如果 $X$ 在整个实轴 $(-\infty, \infty)$ 上取值, 让 $a \rightarrow \infty$ ,  $b \rightarrow \infty$ , 就得如下的一般定义.

**[定义]** 设 $X$ 为连续型随机变量, 有密度函数 $f(x)$ , 则当

$$\int_{-\infty}^{\infty} |x| f(x) dx < \infty$$

时, 称  $E(x) = \int_{-\infty}^{\infty} xf(x)dx$  为 $X$ 的数学期望.

如果  $\int_{-\infty}^{\infty} |x| f(x)dx = \infty$ , 就称 $X$ 的数学期望不存在.

**Example 4.8** 均匀分布的密度函数:  $a \leq x \leq b$  时, 它的数学期望

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



Example 4.9 正态分布  $X \sim N(\mu, \sigma^2)$  的数学期望.

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \sigma \int_{-\infty}^{\infty} \left\{ \left( \frac{x-\mu}{\sigma} \right) + \frac{\mu}{\sigma} \right\} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} d\left( \frac{x-\mu}{\sigma} \right) \\ &= \sigma \int_{-\infty}^{\infty} \left[ y + \frac{\mu}{\sigma} \right] \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \\ &= \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy + \sigma \int_{-\infty}^{\infty} y \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \\ &= \mu \end{aligned}$$

因此正态分布  $N(\mu, \sigma^2)$  中参数  $\mu$  表示它的均值, 这在密度函数的图形中已显示出来了.



### Example 4.10 指数分布的数学期望

设随机变量 $X$ 服从指数分布，其概率密度为  $f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$

其中 $\lambda > 0$ 是常数. 求:  $E(X)$

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} xf(x)dx \\ &= \int_0^{\infty} x\lambda e^{-\lambda x} dx \\ &= \int_0^{\infty} -x de^{-\lambda x} \\ &= -xe^{-\lambda x} \Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx \end{aligned} \quad \begin{aligned} &= \int_0^{\infty} e^{-\lambda x} dx \\ &= -\frac{1}{\lambda} e^{-\lambda x} \Big|_0^{\infty} \\ &= \frac{1}{\lambda} \end{aligned}$$

如果密度函数表示为  $f(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & x > 0 \\ 0 & x \leq 0 \end{cases}$   $E(x) = \theta$





### Example 4.11 顾客平均等待多长时间?

设顾客在某银行的窗口等待的服务的时间  $X$ (以分钟计)服从指数分布,其概率密度为

$$f(x) = \begin{cases} \frac{1}{5} e^{-x/5}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

试求顾客等待服务的平均时间?

解  $E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_0^{\infty} x \cdot \frac{1}{5} e^{-x/5} dx = 5(\text{分钟}).$

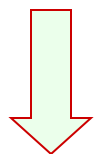
因此,顾客平均等待5分钟就可得到服务.



**Example 4.12** 设随机变量X服从 $\Gamma(\alpha, \beta)$ 分布, 求它的数学期望

解  $f(x) = \begin{cases} 0, & x \leq 0 \\ \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & x > 0 \end{cases}$   $\Gamma(a) = \int_0^{+\infty} u^{a-1} e^{-u} du$

$$EX = \int_0^{+\infty} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^\alpha e^{-x/\beta} dx$$



$$y = x / \beta$$

$$= \int_0^{+\infty} \frac{\beta y^\alpha}{\Gamma(\alpha)} e^{-y} dy$$

$$= \frac{\beta \Gamma(\alpha + 1)}{\Gamma(\alpha)}$$

$$= \alpha \beta$$

$$\Gamma(a+1) = \int_0^{+\infty} u^a e^{-u} du$$

$$= - \int_0^{+\infty} u^a d e^{-u}$$

$$= -u^a e^{-u} \Big|_0^{+\infty} + \int_0^{+\infty} e^{-u} du^a$$

$$= a \int_0^{+\infty} u^{a-1} e^{-u} du$$

$$= a \Gamma(a)$$



**Example 4.13** 设随机变量 $X$ 服从 $\beta(a, b)$ 分布, 求它的数学期望

解

$$f(x) = \begin{cases} \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1}, & 0 < x < 1 \\ 0, & \text{其他} \end{cases}$$

$$\begin{aligned} E(X) &= \int_0^1 x f(x) dx = \int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^a (1-x)^{b-1} dx \\ &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 x^a (1-x)^{b-1} dx \\ &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} = \frac{a}{a+b} \end{aligned}$$



## 4.1.2 随机变量函数的数学期望

我们知道 $X$ 的分布列为

|     |                         |
|-----|-------------------------|
| $X$ | $x_1 \cdots \cdots x_k$ |
| $p$ | $p_1 \cdots \cdots p_k$ |

可以知道  $X^2$  的分布列为

|       |                             |
|-------|-----------------------------|
| $X^2$ | $x_1^2 \cdots \cdots x_k^2$ |
| $p$   | $p_1 \cdots \cdots p_k$     |

故此 
$$E(X^2) = \sum x_k^2 p_k$$

函数 $g(x)$ ，随机变量 $Y=g(x)$ ，有  $E(g(X)) = \sum g(x_k) p_k$

对于连续随机变量 $X$ ，其概率密度函数为 $f(x)$ ，则

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

上式在计算某些随机变量的数学期望时很有用. 如果 $Y=g(X)$ ，我们不必求 $Y$ 的概率密度函数 $f_y(y)$ ，计算  $E(Y) = \int_{-\infty}^{\infty} y f_y(y) dy$ ，而是直接得到  $E(Y) = E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx$

设有联合概率密度函数  $f(x_1, \dots, x_n)$  ,  $Y = g(X_1, \dots, X_n)$  , 则

$$E(Y) = \iint_{R^n} g(x_1, \dots, x_n) f(x_1, \dots, x_n) dx_1 \cdots dx_n$$

特别地有

$$E(X_i) = \iint_{R^n} x_i f(x_1, \dots, x_n) dx_1 \cdots dx_n$$

---

**Example 4.14**  $X, Y$  在区域  $\Omega = \{(x, y) \mid x^2 + y^2 \leq 1\}$  上服从均匀分布  
求  $E(X)$ .

$$f(x, y) = \begin{cases} \frac{1}{\pi} & (x, y) \in \Omega \\ 0 & \text{others} \end{cases}$$

$$E(X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xf(x, y) dx dy = \iint_{x^2 + y^2 \leq 1} x \frac{1}{\pi} dx dy = 0$$

**Example 4.15** 设  $X \sim N(0,1), Y \sim N(0,1)$ ,  $X$  与  $Y$  相互独立, 求  $E(\sqrt{X^2 + Y^2})$

**解:** 我们没有必要去求  $g(X, Y) = \sqrt{X^2 + Y^2}$  的分布函数, 只要求得  $X, Y$  的联合分布函数  $f(x, y)$  即可。

注意到  $X, Y$  分别是标准正态分布, 又是独立的, 则联合分布为  $N(0,0,1,1)$ , 对应的分布密度为  $f(x, y)$

$$f(x, y) = f_X(x)f_Y(y) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x^2 + y^2)\right)$$

$$\begin{aligned} E(\sqrt{X^2 + Y^2}) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f(x, y) dx dy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \sqrt{x^2 + y^2} \frac{1}{2\pi} e^{-\frac{1}{2}(x^2 + y^2)} dx dy \end{aligned}$$

$$\begin{aligned} &= \int_0^{2\pi} d\theta \int_0^{+\infty} \frac{1}{2\pi} \cdot r^2 e^{-\frac{r^2}{2}} dr \\ &= \int_0^{+\infty} r^2 e^{-\frac{1}{2}r^2} dr \\ &= \frac{\sqrt{2\pi}}{2} \end{aligned}$$



**Example 4.16** 国际市场上每年对我国某商品的需求量 $X$ 服从  $[2000, 4000]$ 上的均匀分布.每售出一吨该商品可获利3万美元;但若积压于仓库, 每吨将损失1万美元. 问应组织多少货源才能使收益最大?

**解** 收益  $a_s$  与需求量 $X$ 有关, 也与组织的货源 $S$ 有关.

$a_s$  是随机变量, 故“收益最大”的含义是指“平均收益最大”.

抽象得到需求量  $X \sim U[S_1, S_2]$ ;

季内售出一吨获利  $b$  万美元, 积压一吨损失  $l$  万美元;

进货量为  $S$  ( $S_1 \leq S \leq S_2$ )

为使商店期望利润最大, 如何进货 $S$ ?

$$a_s(X) = \begin{cases} bX - (S - X)l & S_1 \leq X \leq S \\ bS & S \leq X \leq S_2 \end{cases}$$

$$a_s(X) = \begin{cases} (b+l)X - Sl & S_1 \leq X \leq S \\ bS & S \leq X \leq S_2 \end{cases}$$

X的概率密度是  $f(x) = \begin{cases} \frac{1}{S_2 - S_1} & S_1 \leq X \leq S_2 \\ 0 & \text{others} \end{cases}$

$$\begin{aligned} E(a_s(X)) &= \int_{-\infty}^{\infty} a_s(x)f(x)dx \\ &= \int_{S_1}^S \frac{1}{S_2 - S_1} [(b+l)x - Sl] dx + \int_S^{S_2} \frac{bS}{S_2 - S_1} dx \\ &= \frac{1}{S_2 - S_1} \left[ \frac{1}{2} (b+l)(S^2 - S_1^2) - Sl(S - S_1) \right] + \frac{bS(S_2 - S)}{S_2 - S_1} \\ &= \frac{1}{S_2 - S_1} \left[ \frac{1}{2} (b+l)(S^2 - S_1^2) - Sl(S - S_1) + bS(S_2 - S) \right] \\ &= \frac{1}{S_2 - S_1} \left[ -\frac{1}{2} (b+l)S^2 - \frac{1}{2} (b+l)S_1^2 + S(lS_1 + bS_2) \right] \end{aligned}$$





$E(a_s(X))$  是  $S$  的函数，为了使其达到最大， $S$  满足  $\frac{d}{ds} E(a_s(X)) = 0$

也就是：

$$-(b+l)S + (lS_1 + bS_2) = 0$$

$$S = \frac{lS_1}{b+l} + \frac{bS_2}{b+l}$$

容易知道：

$$(1) \quad S_1 \leq S \leq S_2$$

(2)  $S$  是  $S_1$  和  $S_2$  的加权和，如果售出一件商品的利润远远高于售不出而亏损的价值，即  $b \gg l$ ，则  $S \rightarrow S_2$ ，尽量多进货；反之，一件亏损额较大，则  $S \rightarrow S_1$ 。

上述例子中：当  $S=3500$  时， $E(\alpha_s)$  达到最大值 8250 万。

## Example 4.17

串联  $\min(X,Y)$ ,  $f_1(z) = (\alpha + \beta)e^{-(\alpha+\beta)z}$

并联  $\text{Max}(X,Y)$ ,  $f_2(z) = \alpha e^{-\alpha z} + \beta e^{-\beta z} - (\alpha + \beta)e^{-(\alpha+\beta)z}$

备用  $X+Y$ ,  $f_3(z) = \frac{\alpha\beta}{\beta - \alpha} [e^{-\alpha z} - e^{-\beta z}]$

比较三者的分布函数或概率密度没有意义，比如尽管说备用比串联好，但是这里是随机现象，而不是具体的两个元器件，很可能实际结果是第三种寿命比第一种还短。要从均值角度描述比较。

$$\begin{aligned} E(\text{Min}(X,Y)) &= \int_0^{\infty} z(\alpha + \beta)e^{-(\alpha+\beta)z} dz \\ &= -ze^{-(\alpha+\beta)z} \Big|_0^{\infty} + \int_0^{\infty} e^{-(\alpha+\beta)z} dz = -\frac{1}{\alpha + \beta} e^{-(\alpha+\beta)z} \Big|_0^{\infty} \\ &= \frac{1}{\alpha + \beta} \end{aligned}$$



## Example 4.17

$$\begin{aligned} E(\text{Max}(X, Y)) &= \int_0^{\infty} z \{ \alpha e^{-\alpha z} + \beta e^{-\beta z} - (\alpha + \beta) e^{-(\alpha + \beta)z} \} dz \\ &= \frac{1}{\alpha} + \frac{1}{\beta} - \frac{1}{\alpha + \beta} \end{aligned}$$

$$\begin{aligned} E(X + Y) &= \int_0^{\infty} z \frac{\alpha\beta}{\beta - \alpha} [e^{-\alpha z} - e^{-\beta z}] dz \\ &= \frac{\beta}{\beta - \alpha} \cdot \frac{1}{\alpha} - \frac{\alpha}{\beta - \alpha} \cdot \frac{1}{\beta} \\ &= \frac{1}{\alpha} + \frac{1}{\beta} \end{aligned}$$

$$M_3 > M_2 > M_1$$



### 4.1.3 数学期望的基本性质

(1) 设  $C$  为常数, 则有  $E(C) = C$

(2) 设  $X$  是一个随机变量,  $C$  为常数, 则有  $E(CX) = CE(X)$

(3) 设  $X_1, X_2, \dots, X_n$  是  $n$  个随机变量,  $c_1, c_2, \dots, c_n$  为实数, 则

$$E(c_1X_1 + c_2X_2 + \dots + c_nX_n) = c_1E(X_1) + c_2E(X_2) + \dots + c_nE(X_n)$$

(4) 设  $X, Y$  是相互独立的随机变量, 则有

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



Example 4.18 设 $X$ 服从二项分布  $B(n,p)$ , 求 $E(X)$ .

解 已算得 $E(X)=np$ , 现在利用性质2来计算 $E(X)$ .

设计一个伯努里试验,记

$$\xi_i = \begin{cases} 1, & \text{第}i\text{次试验事件}A\text{发生,} \\ 0, & \text{第}i\text{次试验事件}A\text{不发生,} \end{cases}$$

$P(A)=p$ , 此时 $\xi_i$ 服从0-1分布,  $E(\xi_i)=p$ ,

$$X = \sum_{i=1}^n \xi_i$$

$$E(X) = \sum_{i=1}^n E(\xi_i) = np$$



**Example 4.19** 设 $\xi$ 是服从超几何分布的随机变量,  $N$ 件产品中有 $M$ 件次品, 随机抽取 $n$ 件,  $\xi$ 为其中次品的数目

$$P(\xi = m) = C_M^m C_{N-M}^{n-m} / C_N^n \quad (m = 0, 1, 2, \dots, n)$$

**求**  $E(\xi)$ .

**解** 设计一个不放回抽样. 令 $\xi_i$ 为第 $i$ 次抽取时的废品数, 则

$$\xi = \sum_{i=1}^n \xi_i$$

每次取得次品的概率都是 $M/N$

$$P(\xi_i = 1) = M / N, i = 1, 2, \dots, n$$

$$E(X) = \sum_{i=1}^n E(\xi_i) = nM/N$$



## Example 4.20

一客车在有20位旅客从机场出发，沿途有 10 个车站可以下车。如到达一个车站没有旅客下车就不停车，以  $X$  表示停车次数。设每位旅客在各个车站下车是等可能的，并设旅客是否下车是相互独立的。

**求：** 平均停车次数

**解：** 引入随机变量  $X_i$

$$X_i = \begin{cases} 0, & \text{在第 } i \text{ 站没有人下车,} \\ 1, & \text{在第 } i \text{ 站有人下车,} \end{cases} \quad i = 1, 2, \dots, 10.$$

则

$$X = X_1 + X_2 + \dots + X_{10}.$$

$$P\{X_i = 0\} = \left(\frac{9}{10}\right)^{20} \quad (i = 1, 2, \dots, 10)$$

$$P\{X_i = 1\} = 1 - \left(\frac{9}{10}\right)^{20}$$

$$E(X_i) = 1 - \left(\frac{9}{10}\right)^{20}, \quad i = 1, 2, \dots$$

$$\begin{aligned} E(X) &= E(X_1 + X_2 + \dots + X_{10}) \\ &= E(X_1) + E(X_2) + \dots + E(X_{10}) \end{aligned}$$

$$= 10 \left[ 1 - \left(\frac{9}{10}\right)^{20} \right]$$

$$= 8.784(\text{次}).$$





某矿工在井下迷路，井下有三个门：

- 若选择第一个门，经过3小时可到达安全之处；
- 若选择第二个门，经过5小时回到原地；
- 若选择第三个门，经过7小时回到原地。

假设该矿工任意时刻都是随机选择这三个门，则他走到安全之处平均需要多长时间（单位：小时）？

- ☐ A 6
 ☐ B 10
 ☒ C 15
 ☐ D 51



## 条件期望 (conditional expectation)

与条件分布的定义类似，随机变量 $Y$ 的条件期望就是它在给定的附加条件之下的数学期望，以只有一个随机变量 $X$ 作为条件为例，在给定条件 $X = x$ 的条件下， $Y$ 的条件数学期望记为 $E(Y|X = x)$ ，在 $X$ 明确的前提下，可简记为 $E(Y|x)$ 。

如果已知 $(X, Y)$ 的联合密度函数，则 $E(Y|x)$ 可以定义为

$$E(Y|x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy$$

同理，我们可以定义离散情形下的条件期望如下：

$$E(Y|x_i) = \sum_{j=1}^{\infty} y \frac{p_{ij}}{p_{i\cdot}}$$

## 条件期望 (conditional expectation)

条件分布完全刻画了随机变量 $Y$ 和 $X$ 之间的相依关系，条件期望则反映了随着 $X$ 取值 $x$ 的变化， $Y$ 的平均变化如何，这常常是人们关心的主要内容：

- 随着身高 $X$ 的变化，具有身高 $x$ 的人的平均体重的变化情况如何；
- 随着受教育年数 $x$ 的变化，其平均收入的变化。

在统计学上，常把条件期望 $E(Y|x)$ 作为 $x$ 的函数，称为 $Y$ 对 $X$ 的“回归函数”，而回归分析，即关于回归函数的统计研究，构成统计学的重要分支。



## 条件期望 (conditional expectation)

**Example 3.13** 设  $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$  , 求条件密度

$$f_{Y|X}(y | x).$$

$$f_{Y|X}(y | x) = \frac{1}{\sqrt{2\pi} \sqrt{1-\rho^2} \sigma_2} \exp \left\{ -\frac{1}{2(1-\rho^2)\sigma_2^2} \left[ y - \left( \mu_2 + \frac{\sigma_2}{\sigma_1} \rho(x - \mu_1) \right) \right]^2 \right\}$$

它表明:

已知  $X = x$  条件下, 二维正态分布的对于  $Y$  的条件分布是正态分布

$$N \left( \mu_2 + \frac{\sigma_2}{\sigma_1} \rho(x - \mu_1), \left( \sqrt{1-\rho^2} \sigma_2 \right)^2 \right)$$

$$E(Y|x) = \mu_2 + \frac{\sigma_2}{\sigma_1} \rho(x - \mu_1)$$



## 条件期望 (conditional expectation)

从条件期望，可以求得通常的（无条件的）数学期望：

$$E(Y) = E(E(Y|X))$$

称为条件期望的平滑公式。

针对X      针对Y

离散型：  $E(Y) = E(E(Y|X)) = \sum_{i=1}^{\infty} E(Y|x_i)p_i.$

连续型：  $E(Y) = E(E(Y|X)) = \int_{-\infty}^{\infty} E(Y|x)f_X(x)dx$



**例：**某矿工在井下迷路，井下有三个门：

- 若选择第一个门，经过3小时可到达安全之处；
- 若选择第二个门，经过5小时回到原地；
- 若选择第三个门，经过7小时回到原地。

假设该矿工任意时刻都是随机选择这三个门，则他走到安全之处平均需要多长时间（单位：小时）？

解：令 $X=i$ 表示矿工选择第 $i$ 个门， $i=1,2,3$

$Y$ 表示矿工到达安全之处需要的时间（h）

则 $E(Y|X=1)=3, E(Y|X=2)=5+E(Y), E(Y|X=3)=7+E(Y)$

根据条件期望平滑公式：

$$E(Y) = \frac{1}{3} [E(Y|X=1) + E(Y|X=2) + E(Y|X=3)]$$

解得 $E(Y)=15$

## 4.2 随机变量的方差 (Variation)

### 4.2.1 定义

**Example 4.21** 比较两个随机试验，已知分布为：

甲X:

|   |     |     |     |     |     |     |     |
|---|-----|-----|-----|-----|-----|-----|-----|
| X | 4   | 6   | 8   | 10  | 12  | 14  | 16  |
| p | 1/7 | 1/7 | 1/7 | 1/7 | 1/7 | 1/7 | 1/7 |

乙Y:

|   |     |     |     |     |     |     |     |
|---|-----|-----|-----|-----|-----|-----|-----|
| Y | 4   | 5   | 7   | 10  | 13  | 15  | 16  |
| p | 1/7 | 1/7 | 1/7 | 1/7 | 1/7 | 1/7 | 1/7 |

二人的平均分： $E(X) = E(Y) = 10$

从均值无法分辨优孰劣.还要考虑他们取值的离散程度：如同研究生入学考试，既要考总成绩，又要约定任一单科成绩.

## 4.2.1 Introduction/Definition

### 几种可能的比较方式：

(1) 最大值－最小值＝**极差**

$$X: 16 - 4 = 12, Y: 16 - 4 = 12$$

不反映分散程度，其他值被忽略。

(2) **平均差**  $E(X - E(X)) = 0$

对一切随机变量均成立， $X$ 的离差正负相消。

(3) 必须相加不能抵消。

$$E |X - E(X)| \Rightarrow E(X - EX)^2$$

**[定义]**  $X$ 为随机变量，若存在  $E(X - \mu)^2$ ，就称它是随机变量  
 $X$ 的**方差**(variance)，记作 $D(X)$ 或 $\text{Var}(X)$ 。其中  $\mu = E(X)$ 。



**Important.** 方差可以看作是随机变量函数 $(X - E(X))^2$ 的数学期望，根据数学期望的定义，可写出方差的计算公式

$$\begin{aligned} \text{Var}(X) &= E(X - E(X))^2 \\ &= \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx \\ &= \int_{-\infty}^{\infty} \left\{ x^2 - 2xE(X) + [E(X)]^2 \right\} f(x) dx \\ &= \int_{-\infty}^{\infty} x^2 f(x) dx - 2E(X) \int_{-\infty}^{\infty} xf(x) dx + [E(X)]^2 \\ &= E(X^2) - [E(X)]^2 \end{aligned}$$

**Remark1.** 注意到 $D(X)$ 的量纲与 $X$ 不同

为了统一量纲，有时用 $\sqrt{D(X)}$ ，称为 $X$ 的标准差  
(standard deviation)或均方差，记为 $\sigma(X)$ 。

**Remark2.** 前面例子中，区分两个成绩的优异

$$E(X) = E(Y) = 10$$

$$D(X) = 16, D(Y) = 20$$

如此，我们知道甲比乙要好一些



## 4.2.2 Some Examples

### Example 4.22 二项分布的方差

$$P\{X = k\} = C_n^k p^k (1-p)^{n-k} \quad \text{分布律 } X \sim B(n, p)$$

$$E(X^2) = \sum_{k=0}^n k^2 P\{X = k\} \quad E(X) = np$$

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

$$E(X^2) = n(n-1)p^2 + np$$

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

$$= n(n-1)p^2 + np - (np)^2$$

$$= np(1-p)$$



**Example 4.23** 泊松分布  $X \sim \pi(\lambda)$  的方差.

$$\begin{aligned} EX^2 &= \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \sum_{k=1}^{\infty} k \frac{\lambda^k}{(k-1)!} e^{-\lambda} \\ &= \sum_{k=1}^{\infty} (k-1) \frac{\lambda^k}{(k-1)!} e^{-\lambda} + \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} e^{-\lambda} \\ &= e^{-\lambda} \sum_{j=0}^{\infty} j \frac{\lambda^{j+1}}{j!} + \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \end{aligned}$$



**Example 4.23** 计算泊松分布 $\pi(\lambda)$ 的方差.

$$\begin{aligned} &= e^{-\lambda} \sum_{j=1}^{\infty} \frac{\lambda^{j+1}}{(j-1)!} + \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \\ &= \lambda^2 e^{-\lambda} \sum_{j=1}^{\infty} \frac{\lambda^{j-1}}{(j-1)!} + \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \\ &= \lambda^2 + \lambda \end{aligned}$$

所以

$$\text{Var}(X) = EX^2 - (EX)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$



### Example 4.24 均匀分布的方差

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

$$E(X) = \frac{a+b}{2}$$

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

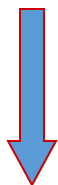
方差  $D(X) = E(X^2) - [E(X)]^2 = \frac{1}{12} (b-a)^2$



**Example 4.25** 设X服从正态分布  $N(\mu, \sigma^2)$  , 求D(X).

解  $E(X) = \mu$

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



$$t = \frac{x - \mu}{\sigma}$$

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

可见正态分布中参数 $\sigma^2$ 就是它的方差,  $\sigma$ 就是标准差.



**Example 4.26** 指数分布的方差: 设随机变量  $X$  服从指数分布,

其概率密度为  $f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$

其中  $\lambda > 0$  是常数. **求:**  $D(X)$

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} x^2 f(x) dx \\ &= \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx \\ &= -\int_0^{\infty} x^2 de^{-\lambda x} \\ &= -x^2 e^{-\lambda x} \Big|_0^{\infty} + \int_0^{\infty} 2xe^{-\lambda x} dx \\ &= \frac{2}{\lambda} \int_0^{\infty} x(\lambda e^{-\lambda x}) dx \end{aligned}$$

$$\begin{aligned} &= \frac{2}{\lambda} E(X) \\ &= \frac{2}{\lambda^2} \\ D(x) &= E(X^2) - (EX)^2 \\ &= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} \\ &= \frac{1}{\lambda^2} \end{aligned}$$





## 常见分布及其期望和方差列表

| 分布名称  | 数学期望 $E(X)$                       | 方差 $D(X)$                             |
|-------|-----------------------------------|---------------------------------------|
| 0-1分布 | $p$                               | $pq$                                  |
| 二项分布  | $np$                              | $npq$                                 |
| 泊松分布  | $\lambda$                         | $\lambda$                             |
| 均匀分布  | $\frac{a+b}{2}$                   | $\frac{(b-a)^2}{12}$                  |
| 正态分布  | $\mu$                             | $\sigma^2$                            |
| 指数分布  | $\frac{1}{\lambda}$ (或 $\theta$ ) | $\frac{1}{\lambda^2}$ (或 $\theta^2$ ) |



设随机变量 $X$ 服从 $\Gamma(\alpha, \beta)$ 分布, 求它的方差

$$f(x) = \begin{cases} 0, & x \leq 0 \\ \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}, & x > 0 \end{cases} \quad \Gamma(a) = \int_0^{+\infty} u^{a-1} e^{-u} du$$

解  $E(X) = \alpha\beta$

$$\begin{aligned} E(X^2) &= \int_0^{\infty} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha+1} e^{-\frac{x}{\beta}} dx = \frac{1}{\Gamma(\alpha)} \beta^2 \int_0^{\infty} \left(\frac{x}{\beta}\right)^{\alpha+1} e^{-\frac{x}{\beta}} d\frac{x}{\beta} \\ &= \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)} \beta^2 = \alpha(\alpha+1)\beta^2 \end{aligned}$$

$$D(X) = E(X^2) - E(X)^2 = \alpha\beta^2$$



设随机变量 $X$ 服从 $\beta(a, b)$ 分布, 求它的方差

$$f(x) = \begin{cases} \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1}, & 0 < x < 1 \\ 0, & \text{其他} \end{cases}$$

解  $E(X) = \frac{a}{a+b}$

$$E(X^2) = \int_0^1 \frac{1}{B(a, b)} x^{a+1} (1-x)^{b-1} dx = \frac{1}{B(a, b)} B(a+2, b)$$

$$= \frac{\Gamma(a+b) \Gamma(a+2) \Gamma(b)}{\Gamma(a) \Gamma(b) \Gamma(a+b+2)} = \frac{a(a+1)}{(a+b)(a+b+1)}$$

$$D(X) = E(X^2) - E(X)^2 = \frac{ab}{(a+b)^2(a+b+1)}$$



## 4.2.3 方差的性质

$$(1) \quad D(C)=0; \quad (C \text{ 为常数})$$

$$(2) \quad D(aX + b) = a^2 D(X);$$

此式说明：

随机变量平移 $X+b$ 不影响方差；

随机变量的比例（尺度）变换 $aX$ ，会比例放大方差。

$$(3) \quad X_1, \dots, X_n \text{ 为随机变量, } \underline{\text{如果 } X_1, \dots, X_n \text{ 相互独立,}}$$

$$c_1, \dots, c_n \text{ 为常数, } D(c_1 X_1 + \dots + c_n X_n) = \sum_{j=1}^n c_i^2 D(X_i)$$



### 4.2.3 方差的性质

(4)  $D(X)=0$  的充分必要条件是X以概率1取得常数C, 即

$$P(X = C) = 1$$

证明:

如果 $P(X=C)=1$

那么 $\mu = E(X) = C$

$$\begin{aligned} D(X) &= E(X - \mu)^2 \\ &= \int_{-\infty}^{+\infty} (x - C)^2 f(x) dx \\ &= 0 \end{aligned}$$



## 4.2.3 方差的性质

证明： 如果 $D(X)=0$ ,

$$P\{|X - \mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2} \quad (\text{Chebyshev不等式})$$

$$P\{|X - \mu| \geq \frac{1}{n}\} \leq 0$$

$$P\{|X - \mu| \geq \frac{1}{n}\} = 0$$

$$P\{|X - \mu| < \frac{1}{n}\} = 1$$



$$P\{|X - \mu| = 0\} = 1$$



切比雪夫(Chebyshev)不等式 若随机变量的方差存在, 则

对任意给定的正数 $\varepsilon$ , 恒有  $P\{|X - \mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}$

**Proof:**

$$\begin{aligned} P\{|X - \mu| \geq \varepsilon\} &= \int_{|x-\mu| \geq \varepsilon} f(x) dx \\ &\leq \int_{|x-\mu| \geq \varepsilon} \frac{(x - \mu)^2}{\varepsilon^2} f(x) dx \\ &\leq \int_{-\infty}^{\infty} \frac{(x - \mu)^2}{\varepsilon^2} f(x) dx = \frac{\sigma^2}{\varepsilon^2} \end{aligned}$$

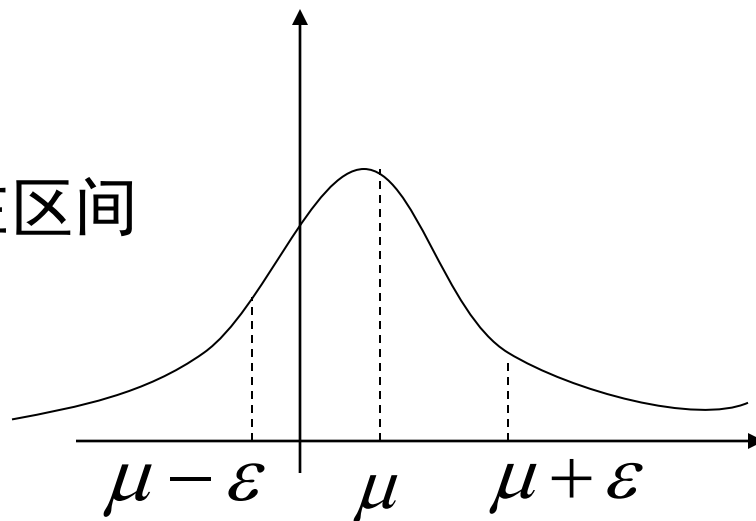
上面的式子等价于

$$P\{|X - \mu| < \varepsilon\} = 1 - P\{|X - \mu| \geq \varepsilon\} \geq 1 - \frac{\sigma^2}{\varepsilon^2}$$

# 切比雪夫不等式意义

- 证明方法上
- 结论

事实上，该式断言随机变量 $X$ 落在区间  
 $(-\infty, E(X) - \varepsilon)$ 和 $(E(X) + \varepsilon, \infty)$   
内的概率小于等于  $D(X) / \varepsilon^2$



或者说， $X$ 落在区间  $(E(X) - \varepsilon, E(X) + \varepsilon)$  之外的概率小于  $D(X) / \varepsilon^2$ ，之内的概率大于  $1 - D(X) / \varepsilon^2$ ，从而只用数学期望和方差就可对上述概率进行估计。



## 切比雪夫不等式意义.

这一定律可用来估计分布未知情况下事件的概率

$$\{|X - \mu| \geq \varepsilon\} \text{ 或 } \{|X - \mu| \leq \varepsilon\}$$

例如  $P\{|X - u| < \sigma\} \geq 0$

$$P\{|X - u| < 2\sigma\} \geq 0.75$$

$$P\{|X - u| < 3\sigma\} \geq 0.8888$$

$$P\{|X - u| < 4\sigma\} \geq 0.9375$$



## 切贝雪夫不等式意义.

以正态分布  $X \sim N(u, \sigma^2)$  为例, 已知其均值和方差  $\mu, \sigma^2$ , 我们有 $3\sigma$ 原则:

$$P\{|X - u| < \sigma\} = 2\Phi(1) - 1 \approx 0.6826 \geq 0$$

$$P\{|X - u| < 2\sigma\} = 2\Phi(2) - 1 \approx 0.9544 \geq 0.75$$

$$P\{|X - u| < 3\sigma\} = 2\Phi(3) - 1 \approx 0.9973 \geq 0.8888$$

$$P\{|X - u| < 4\sigma\} = 2\Phi(4) - 1 \approx 1 \geq 0.9375$$

可与Chebyshev不等式比较;

可见Chebyshev不等式的估计比较粗略。



**Example 4.27** 设随机变量 $X$ 具有数学期望 $E(X)=\mu$ ，方差 $D(X)=\sigma^2$ ，则其标准化变量  $X^* = \frac{X - \mu}{\sigma}$  ( $\sigma > 0$ ) 的均值为0，方差为1。

$$E(X^*) = E\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma}[E(X) - \mu] = 0$$

$$\begin{aligned} E(X^{*2}) &= E\left(\frac{X - \mu}{\sigma}\right)^2 \\ &= \frac{1}{\sigma^2} E[(X - \mu)^2] \\ &= 1 \end{aligned}$$

$$D(X) = E(X^{*2}) - [E(X^*)]^2 = 1$$



**Example 4.28** 设  $c$  是常数, 则  $D(X) \leq E(X-c)^2$

**证明:**

$$\begin{aligned} E(X-c)^2 - D(X) &= E(X-c)^2 - E(X-\mu)^2 \\ &= E[(X-c)^2 - (X-\mu)^2] \\ &= E[(\mu-c)(2X-\mu-c)] \\ &= (\mu-c)E(2X-\mu-c) \\ &= (\mu-c)^2 \geq 0 \end{aligned}$$

**此结果说明:** 均值  $\mu$  是最能描述一个变量的平均分布的。任何一个其它值  $c$ , 随机变量  $X$  偏离  $c$  的情况都比偏离  $\mu$  要糟。



**Example 4.29** 设相互独立同分布随机变量  $X_1, \dots, X_n$

$$E(X_i) = \mu, \quad D(X_i) = \sigma^2 \quad (i = 1, 2, \dots, n) \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

求  $E(\bar{X}), D(\bar{X})$

$$E(\bar{X}) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \mu$$

$$\begin{aligned} D(\bar{X}) &= D\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n^2} D\left(\sum_{i=1}^n X_i\right) \\ &= \frac{1}{n^2} \sum_{i=1}^n D(X_i) = \frac{\sigma^2}{n} \end{aligned}$$

**Remark** 这说明在独立同分布时，作为各  $X_i$  的算术平均  $\bar{X}$ ，它的数学期望与各  $X_i$  的数学期望相同，但方差只有  $X_i$  的  $1/n$  倍。这一事实在数理统计中有重要意义。

设随机变量  $X \sim N(0,1)$ , 则  $E(X^2 + 1) = ( \quad )$

A 0

B 1

C 2

D 3



## 条件方差\*

与条件期望的定义一样，我们可以定义在 $X = x$ 的条件之下 $Y$ 的条件方差：

$$D(Y|x) = E[(Y - E(Y|x))^2|x]$$

即 $D(Y|x)$ 是 $Y$ 和它的条件期望之差的平方的条件期望值。

与普通方差一样，条件方差也有

$$D(Y|x) = E(Y^2|x) - (E(Y|x))^2$$

等式两边关于 $X$ 求期望得

$$E[D(Y|x)] = E(Y^2) - E[(E(Y|x))^2]$$

## 条件方差\*

$$E[D(Y|x)] = E(Y^2) - E[(E(Y|x))^2]$$

另一方面

$$\begin{aligned} D[E(Y|x)] &= E[E(Y|x)^2] - [E(E(Y|x))]^2 \\ &= E[E(Y|x)^2] - [E(Y)]^2 \end{aligned}$$

两式相加得：

$$E[D(Y|x)] + D[E(Y|x)] = E(Y^2) - [E(Y)]^2 = D(Y)$$

因此，我们得到如下公式：

$$D(Y) = E[D(Y|x)] + D[E(Y|x)] \quad \text{条件方差公式}$$

关于x求期望、方差





**例：**设对任意时间 $t$ ，在 $(0, t)$ 内到达火车站的人数服从泊松分布，均值为 $\lambda t$ ，若火车在 $(0, T)$ 内随机到达（即到达时间服从 $(0, T)$ 上的均匀分布），且与旅客到达火车站的时间独立。

求：火车到达时，上火车的旅客人数的期望与方差。

分析：火车在 $t$ 时刻到站后，在 $t$ 时刻之前到达的旅客都能上车，这时车上的人数应与 $(0, t)$ 内到站的旅客人数相等。

解：对任意的 $t \geq 0$ ，令 $N(t)$ 表示 $t$ 以前到达车站的人数， $Y$ 表示火车到站的时间， $N(Y)$ 表示火车上的人数。给定 $Y$ 的条件下有

$$\{N(Y)|Y = t\} = \{N(t)|Y = t\}$$

因此  $E[N(Y)|Y = t] = E[N(t)|Y = t] = E[N(t)] = \lambda t$

两边关于 $Y$ 取期望： $E[E[N(Y)|Y = t]] = E(\lambda t) = \lambda E(t) = \frac{\lambda T}{2}$

故 $E[N(Y)] = \frac{\lambda T}{2}$

为了计算方差，我们利用条件方差公式

$$D(N(Y)) = E[D(N(Y)|Y = t)] + D[E(N(Y)|Y = t)] \quad (1)$$

其中第二项已经计算过了

对于第一项  $D(N(Y)|Y = t) = D(N(t)|Y = t)$

根据 $N(t)$ 与 $Y$ 独立， $D(N(t)|Y = t) = D(N(t)) = \lambda t$

带入（1）式可得

$$D(N(Y)) = E(\lambda t) + D(\lambda t)$$

$$= \frac{\lambda T}{2} + \frac{\lambda^2 T^2}{12}$$

## 练习：计算下面分布的期望和方差

**1. 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**  $P\{X=k\}=P\{\text{前}k-1\text{次试验中有}r-1\text{次成功，有}k-r\text{次不成功，且第}k\text{次成功}\}$

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

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

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



**2. 对数正态分布** 设  $X \sim N(\mu, \sigma^2)$ , 那么随机变量  $Y = e^X$  就称为参数为  $\mu, \sigma^2$  的对数正态随机变量 (如果  $\ln(Y)$  服从正态分布, 则随机变量  $Y$  服从对数正态分布)。

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma y} \exp\left\{-\frac{(\ln(y) - \mu)^2}{2\sigma^2}\right\}, y > 0$$



3. 超几何分布 设 $\xi$ 是服从超几何分布的随机变量，N件产品中有M件次品，随机抽取n件， $\xi$ 为其中次品的数目

$$P(\xi = m) = C_M^m C_{N-M}^{n-m} / C_N^n (m = 0, 1, 2, \dots, n)$$



# 中位数

**定义：** 设连续型随机变量 $X$ 的分布函数是 $F(X)$ ，则满足条件 $P(X \leq m) = F(m) = \frac{1}{2}$ 的数 $m$ 称为 $X$ 或者分布 $F$ 的中位数。

$$P(X \leq m) = P(X < m) = P(X > m) = P(X \geq m) = 1/2$$

有时，中位数比期望更能说明问题，如评估平均收入，中位数不易受特大值和特小值影响。

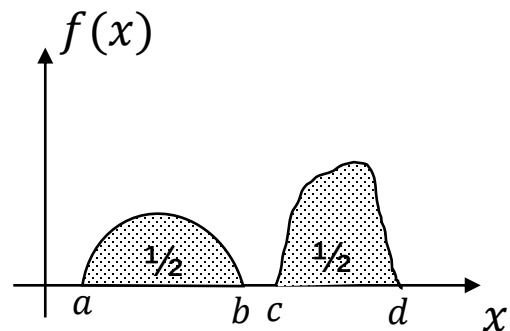
中位数总是存在的。



# 中位数

期望具有良好的数学性质，中位数没有。

中位数不唯一。



离散情况的中位数？

|     |     |     |     |
|-----|-----|-----|-----|
| $X$ | 1   | 2   | 3   |
| $P$ | 2/7 | 4/7 | 1/7 |



## 4.3 协方差与相关系数

### 4.3.1 协方差

单个随机变量的期望与方差分别反映了随机变量取值的平均水平和随机变量取值相对于均值的分散程度.

对于两个随机变量 $X$ 、 $Y$ ，他们之间的关系如何描述？

前面知道如果 $X, Y$ 独立，则  $E(XY) = E(X)E(Y)$

$$\begin{aligned} & E[(X - E(X))(Y - E(Y))] \\ &= E[XY - XE(Y) - YE(X) + E(X)E(Y)] \\ &= E(XY) - E(X)E(Y) = 0 \end{aligned}$$

而当 $X, Y$ 不独立， $E[(X - E(X))(Y - E(Y))]$ 是否也可以为0？





## 4.3 协方差与相关系数

### 4.3.1 协方差

**[定义]** 设  $(X, Y)$  是二维随机变量, 如果  $E(X), E(Y)$  都存在, 且  $E(X - E(X))(Y - E(Y))$  也存在, 则称之为  $X$  与  $Y$  的协方差, 记为  $\text{Cov}(X, Y)$

### 协方差的性质

$$(1) \text{Cov}(X, X) = D(X)$$

$$(2) \text{Cov}(X, Y) = \text{Cov}(Y, X).$$

### 4.3.1 协方差

$$\begin{aligned}(3) \quad Cov(X, Y) &= E[XY - XE(Y) - E(X)Y + E(X)E(Y)] \\ &= E(XY) - E(X)E(Y)\end{aligned}$$

$$(4) \quad D(X + Y) = D(X) + D(Y) + 2Cov(X, Y)$$

$$\begin{aligned}D(X + Y) &= E[(X + Y) - E(X + Y)]^2 \\ &= E(X - E(X))^2 + E(Y - E(Y))^2 + 2E(X - E(X))(Y - E(Y)) \\ &= D(X) + D(Y) + 2Cov(X, Y)\end{aligned}$$

## 性质:

(1)  $Cov(X, C) = 0$ ; ( $C$ 为任意常数)

(2)  $Cov(a_1X_1 + a_2X_2, Y) = a_1Cov(X_1, Y) + a_2Cov(X_2, Y)$

$$\begin{aligned} Cov(a_1X_1 + a_2X_2, Y) &= E(a_1X_1 + a_2X_2 - E(a_1X_1 + a_2X_2))(Y - EY) \\ &= E(a_1(X_1 - EX_1) + a_2(X_2 - EX_2))(Y - EY) \\ &= E(a_1(X_1 - EX_1)(Y - EY) + a_2(X_2 - EX_2)(Y - EY)) \\ &= a_1Cov(X_1, Y) + a_2Cov(X_2, Y) \end{aligned}$$

(3)  $Cov(aX, bY) = abCov(X, Y)$ , ( $a, b$ 为常数).

关于每个参数是线性的, 关于两个参数是双线性的。

## 4.3.2 相关系数

协方差虽在某种意义上表示了两个随机变量间的关系，但  $\text{Cov}(X, Y)$  的取值大小与  $X, Y$  的量纲有关。我们考虑消除量纲，同时把这种关系归一化。

**定义** 称下式为  $X, Y$  的相关系数 (correlation coefficient)。

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

Remark:  $\rho_{XY}$  是无量纲的，它事实上是  $X, Y$  的标准化随机变量的协方差

$$\begin{aligned}\rho_{XY} &= \frac{\text{Cov}(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}} = E \left[ \frac{(X - E(X))}{\sqrt{D(X)}} \frac{(Y - E(Y))}{\sqrt{D(Y)}} \right] \\ &= \text{Cov}(X^*, Y^*)\end{aligned}$$

# 相关系数的意义：随机变量X和Y的线性相关性

假设以X的线性函数  $aX + b$  逼近Y，以均方误差

$$e = E[(Y - (aX + b))^2]$$

表示逼近的程度，e越小说明逼近越好，反之越差。

那么最好的逼近是否存在？

如存在是什么样的？

$$e = E[Y - (aX + b)]^2$$

$$= E[Y^2 - 2aXY - 2bY + a^2X^2 + b^2 + 2abX]$$

$$= E(Y^2) - 2aE(XY) - 2bE(Y) + a^2E(X^2) + 2abE(X) + b^2$$



为求得最佳的 $a, b$ 应该满足偏导为零：

$$0 = \frac{\partial e}{\partial a} = -2E(XY) + 2aE(X^2) + 2bE(X)$$

$$0 = \frac{\partial e}{\partial b} = -2E(Y) + 2aE(X) + 2b$$

$$\begin{cases} aE(X^2) + bE(X) = E(XY) \\ aE(X) + b = E(Y) \end{cases}$$

解出：

$$a^* = \frac{E(XY) - E(X)E(Y)}{E(X^2) - [E(X)]^2} = \frac{Cov(X, Y)}{D(X)}$$

$$b^* = \frac{E(X^2)E(Y) - E(X)E(XY)}{E(X^2) - [E(X)]^2}$$



## 相关系数的意义:

$$b^* = \frac{\{E(X^2) - [E(X)]^2\}E(Y) + [E(X)]^2 E(Y) - E(X)E(XY)}{E(X^2) - [E(X)]^2}$$

$$= E(Y) - E(X) \frac{\text{Cov}(X, Y)}{D(X)}$$

将  $a^*, b^*$  代入得

$$\min e = E(Y - (a^* X + b^*))^2$$

$$= E\left(Y - \frac{\text{Cov}(X, Y)}{D(X)} X - E(Y) + E(X) \frac{\text{Cov}(X, Y)}{D(X)}\right)^2$$

$$= E\left(Y - E(Y) - \frac{\text{Cov}(X, Y)}{D(X)} [X - E(X)]\right)^2$$



$$\begin{aligned}
\min(e) &= E \left[ [Y - E(Y)]^2 - 2 \frac{\text{Cov}(X, Y)}{D(X)} [Y - E(Y)][X - E(X)] \right. \\
&\quad \left. + \left[ \frac{\text{Cov}(X, Y)}{D(X)} \right]^2 [X - E(X)]^2 \right] \\
&= D(Y) - 2 \frac{\text{Cov}(X, Y)}{D(X)} \text{Cov}(X, Y) + \left[ \frac{\text{Cov}(X, Y)}{D(X)} \right]^2 D(X) \\
&= D(Y) - \frac{[\text{Cov}(X, Y)]^2}{D(X)} \\
&= D(Y) \left( 1 - \frac{[\text{Cov}(X, Y)]^2}{D(X)D(Y)} \right) \\
&= D(Y)[1 - \rho_{XY}^2]
\end{aligned}$$

以  $aX + b$  逼近  $Y$ ，最好的逼近与  $Y$  的均方差是  $D(Y)[1 - \rho_{XY}^2]$



**[定理]:** (1)  $|\rho_{XY}| \leq 1$

(2)  $|\rho_{XY}| = 1$  的充分必要条件是, 存在常数  $a, b$  使得

$$P\{Y = aX + b\} = 1$$

**证明:** (1) 由  $e$  的非负性得到  $|\rho_{XY}| \leq 1$

引理 
$$\rho_{XY}^2 = \frac{[Cov(X, Y)]^2}{D(X)D(Y)} \leq 1$$

$$\Rightarrow [Cov(X, Y)]^2 \leq D(X)D(Y) \quad \text{施瓦茨 (Schwarz) 不等式}$$

(2) 如果  $|\rho_{XY}| = 1$ , 则对于上述的  $a^*, b^*$  得到

$$E[(Y - a^*X - b^*)^2] = D(Y)(1 - \rho_{XY}^2) = 0$$



由  $D(X) = E(X^2) - (EX)^2$  得到

$$D(Y - a^*X - b^*) + [E(Y - a^*X - b^*)]^2 = E(Y - a^*X - b^*)^2 = 0$$

上式中两项全是正数，故此必须

$$\begin{cases} D(Y - a^*X - b^*) = 0 \\ E(Y - a^*X - b^*) = 0 \end{cases}$$

第一式子得到

$$P\{Y - a^*X - b^* = C\} = 1$$

进一步由第二式得到  $C=0$ ;

$$P\{Y = a^*X + b^*\} = 1$$



反之，如果存在  $a^*, b^*$  使得  $P\{Y = a^* X + b^*\} = 1$ ，则得到

$$P\{Y - a^* X - b^* = 0\} = 1$$

$$P\{[Y - a^* X - b^*]^2 = 0\} = 1$$

$$E[(Y - a^* X - b^*)^2] = 0$$

而

$$E[Y - a^* X - b^*]^2 = 0 \geq \min e = [1 - \rho_{XY}^2] D(Y)$$

必然  $|\rho_{XY}| = 1$



**Remark:**  $e$  是  $|\rho_{XY}|$  的严格单调递减函数。

$|\rho_{XY}|$  越大,  $e$  越小, 说明  $X, Y$  的线性相关性越大。

特别地  $|\rho_{XY}| = 1$ , 则  $X, Y$  以概率1线性相关。

$|\rho_{XY}|$  表征了  $X, Y$  的线性相关程度, 故称为**相关系数**。

**定义** 如果  $\rho_{XY} = 0$ , 称 $X, Y$ 不相关

**Remark**

(1)  $X, Y$  独立必然有  $\rho_{XY} = 0$ , 则  $X, Y$  不相关;

(2)  $\rho_{XY} = 0$  未必有  $X, Y$  独立。



**Example 4.30** 设随机变量 $\theta$ 服从均匀分布 $U(0, 2\pi)$ ,  
 $X=\cos\theta$ ,  $Y=\sin\theta$ , 显然 $X^2+Y^2=1$ , 故 $X$ 与 $Y$ 不独立. 但

$$EX = \int_0^{2\pi} \cos\varphi \frac{1}{2\pi} d\varphi = 0$$

$$EY = \int_0^{2\pi} \sin\varphi \frac{1}{2\pi} d\varphi = 0$$

$$\begin{aligned} E(XY) &= E(\cos\theta \sin\theta) \\ &= \int_0^{2\pi} \cos\varphi \sin\varphi \frac{1}{2\pi} d\varphi = 0 \end{aligned}$$

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0$$

即 $X$ 与 $Y$ 不相关.

**[定理]** 对随机变量 $X$ 和 $Y$ , 下列事实等价:

(1)  $Cov(X, Y) = 0$

(2)  $X$ 与 $Y$ 不相关;

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

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

对于正态分布情形, 独立与不相关是一致的。



**Example 4.31** 设 $X, Y$ 服从二元正态分布  $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$   
试求  $\rho_{XY}$  .

**解：** 前面知 $X, Y$ 分别服从正态分布  $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$

$$\begin{aligned} \text{Cov}(X, Y) &= E[(X - E(X))(Y - E(Y))] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_1)(y - \mu_2) f(x, y) dx dy \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}\sigma_1\sigma_2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_1)(y - \mu_2) \\ &\quad \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\} dx dy \end{aligned}$$



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



$$t = \frac{1}{\sqrt{1-\rho^2}} \left( \frac{y-\mu_2}{\sigma_2} - \rho\frac{x-\mu_1}{\sigma_1} \right) \quad z = \frac{x-\mu_1}{\sigma_1}$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}\sigma_1\sigma_2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\sigma_1 z)(\sigma_2\sqrt{1-\rho^2}t + \rho\sigma_2 z) \exp\left[-\frac{1}{2}z^2 - \frac{1}{2}t^2\right] (\sigma_1\sqrt{1-\rho^2}\sigma_2) dz dt$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\sigma_1\sigma_2\sqrt{1-\rho^2}zt + \rho\sigma_1\sigma_2z^2) \exp\left[-\frac{1}{2}(z^2 + t^2)\right] dz dt$$





$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sigma_1 \sigma_2 \sqrt{1-\rho^2} zt \exp\left[-\frac{1}{2}(z^2 + t^2)\right] dzdt +$$

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho \sigma_1 \sigma_2 z^2 \exp\left[-\frac{1}{2}(z^2 + t^2)\right] dzdt$$

$$= \frac{\sigma_1 \sigma_2 \sqrt{1-\rho^2}}{2\pi} \int_{-\infty}^{\infty} z e^{-\frac{1}{2}z^2} dz \int_{-\infty}^{\infty} t e^{-\frac{1}{2}t^2} dt +$$

$$\frac{\rho \sigma_1 \sigma_2}{2\pi} \int_{-\infty}^{\infty} z^2 e^{-\frac{1}{2}z^2} dz \int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2} dt$$

$$= 0 + \frac{\rho \sigma_1 \sigma_2}{2\pi} \sqrt{2\pi} \sqrt{2\pi}$$

$$= \rho \sigma_1 \sigma_2$$

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{D(X)} \sqrt{D(Y)}} = \rho$$



**Remark 1:** 二维正态随机变量的分布函数完全由于参数  $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$  决定;

**Remark 2:** 如果  $(X, Y)$  服从正态分布, 则  $X, Y$  独立与  $X, Y$  不相关是等价的。条件是  $\rho = 0$ 。

设二维随机变量 $(X, Y)$ 服从二维正态分布，则下列条件中不是 $X, Y$ 相互独立的充分必要条件的是（ ）

- ☐ A  $\rho = 0$
- ☐ B  $E(XY) = E(X)E(Y)$
- ☒ C  $E(X) = E(Y) = 0$
- ☐ D  $P(XY) = P(X)P(Y)$

1. 设随机变量 $X, Y$ 具有如下数字特征:  $E(X) = 1, E(Y) = 2, E(X^2) = 2, E(Y^2) = 8, \rho_{XY} = 0.5$ , 若令 $U = 2X + 3Y$ , 则 $U$ 和 $X$ 的协方差为  
[填空1]

2. 令 $X \sim U\left(-\frac{1}{2}, \frac{1}{2}\right), Y = \cos X$ , 则 $\rho_{XY} =$  [填空2]

正常使用填空题需3.0以上版本雨课堂



**Example 4.31** 已知随机向量 $(X, Y)$ 的联合密度函数如下

$$f(x, y) = \begin{cases} \frac{8}{3}, & 0 < x - y < 0.5, 0 < x, y < 1 \\ 0, & \text{其他} \end{cases}$$

求 $X, Y$ 的相关系数。



**Example 4.32** 用1万元投资甲、乙两种证券，分别投资 $x_1, x_2$ 万元，且 $x_1 + x_2 = 1$ ，于是 $(x_1, x_2)$ 就形成了一种投资组合。设 $X$ 是投资甲证券的收益率， $Y$ 是投资乙证券的收益率，它们都是随机变量，若已知 $X$ 和 $Y$ 的平均收益分别是 $\mu_1, \mu_2$ ，方差（代表风险）分别是 $\sigma_1^2, \sigma_2^2$ ， $X$ 和 $Y$ 之间的相关系数是 $\rho$ 。试求该投资组合的平均收益与风险（方差），并求使得投资风险最小的 $x_1$ 是多少？



## 4.4 矩(moment)与协方差阵

**定义：** 随机变量 $X, Y$

$E(X^k)$  ( $k = 1, 2, \dots$ ) 称为 $X$ 的 $k$ 阶原点矩，简称 $k$ 阶矩。

$E[X - E(X)]^k$  ( $k = 1, 2, \dots$ ) 称为 $X$ 的 $k$ 阶中心矩。

$E(X^k Y^l)$  ( $k, l = 1, 2, \dots$ ) 称为 $X, Y$ 的 $k+l$ 阶混合矩。

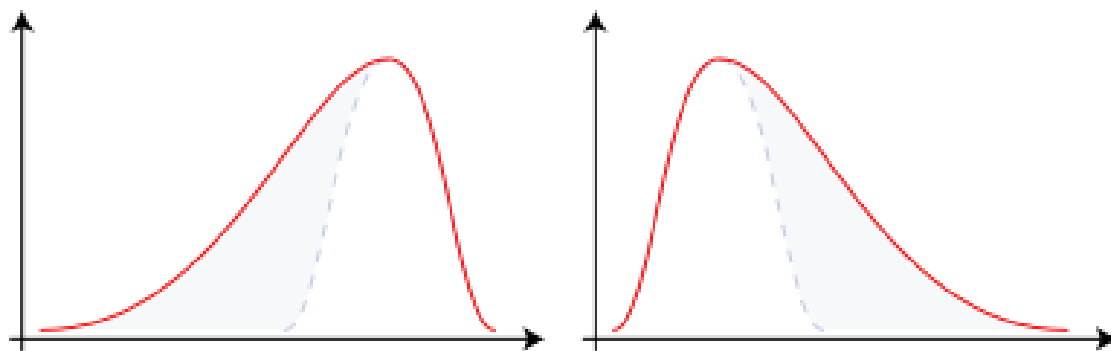
$E[X - E(X)]^k [Y - E(Y)]^l$  ( $k, l = 1, 2, \dots$ )

称为 $X, Y$ 的 $k+l$ 阶混合中心矩。

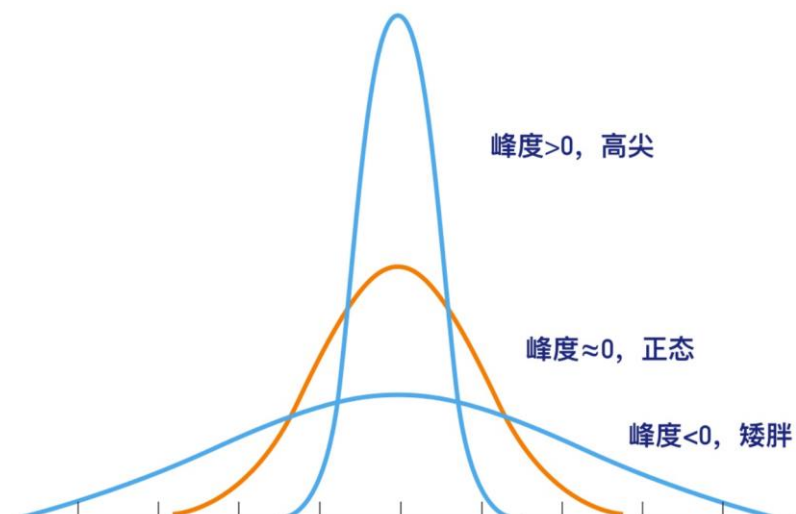


# 偏度和峰度

偏度 (Skewness)



峰度 (Kurtosis)





**Example 4.33** 设 $X$ 服从正态分布 $N(0, \sigma^2)$ 的随机变量,  $EX=0$ , 求 $X$ 的 $n$ 阶原点矩 (中心矩)

$$\begin{aligned} E(X^n) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^n e^{-x^2/2\sigma^2} dx \\ &= \begin{cases} 0, & n = 2k + 1, \\ 1 \times 3 \times \cdots \times (n-1) \sigma^n, & n = 2k. \end{cases} \end{aligned}$$

**Example 4.34** 如果 $\xi$ 服从参数为 $\lambda$ 的指数分布, 那么 对于 $k \geq 1$ ,

$$E\xi^k = \int_0^{+\infty} x^k \lambda e^{-\lambda x} dx = \frac{k}{\lambda} E\xi^{k-1}$$

根据递推关系得  $E\xi^k = \frac{k!}{\lambda^k}$

即指数分布的任意阶矩存在.

对于n个随机变量  $X_1, X_2, \dots, X_n$  , 可以定义:

$$C_{ij} = \text{Cov}(X_i, X_j) \quad i, j = 1, 2, \dots, n$$

进一步定义

$$C = \begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{pmatrix}$$

称  $C$  为  $X_1, X_2, \dots, X_n$  的协方差矩阵。



# n维正态分布

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot$$

$$\exp \left\{ \frac{-1}{2(1-\rho^2)} \left[ \frac{(x_1 - \mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right] \right\}$$

将上式中exp里面的部分写成矩阵形式

$$\mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

协方差矩阵为

$$\mathbf{C} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$



# n维正态分布

协方差矩阵  $\mathbf{C} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$

行列式  $|\mathbf{C}| = \sigma_1^2 \sigma_2^2 (1 - \rho^2)$

$\mathbf{C}$  的逆矩阵  $\mathbf{C}^{-1} = \frac{1}{|\mathbf{C}|} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}$

$$(\mathbf{X} - \boldsymbol{\mu})^T \mathbf{C}^{-1} (\mathbf{X} - \boldsymbol{\mu})$$

$$= \frac{1}{|\mathbf{C}|} (x_1 - \mu_1 \quad x_2 - \mu_2) \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}$$

## n维正态分布

$$\begin{aligned} & (\mathbf{X} - \boldsymbol{\mu})^T \mathbf{C}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \\ &= \frac{1}{|\mathbf{C}|} (x_1 - \mu_1 \quad x_2 - \mu_2) \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix} \\ &= \frac{1}{1 - \rho^2} \left[ \frac{(x_1 - \mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right] \end{aligned}$$

所以二维正态分布的概率密度函数可以写为

$$f(x_1, x_2) = \frac{1}{(2\pi)^{2/2} |\mathbf{C}|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{X} - \boldsymbol{\mu})^T \mathbf{C}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \right\}$$

# n维正态分布

n维正态分布的概率密度 $(X_1, X_2, \dots, X_n)$ 函数定义为

$$f(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^{n/2} |\mathbf{C}|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{X} - \boldsymbol{\mu})^T \mathbf{C}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \right\}$$

其中 $\mathbf{C}$ 是 $(X_1, X_2, \dots, X_n)$ 的协方差矩阵。

## n维正态随机变量的重要性质

1) n维正态随机变量 $(X_1, X_2, \dots, X_n)$ 的每一个分量 $X_i, i = 1, 2, \dots, n$ 都是正态随机变量; 反之若 $X_i, i = 1, 2, \dots, n$ 都是正态随机变量, 且相互独立, 则 $(X_1, X_2, \dots, X_n)$ 是n维正态随机变量。

# n维正态分布

## n维正态随机变量的重要性质

2) n维随机变量 $(X_1, X_2, \dots, X_n)$ 服从n维正态分布的充要条件是 $X_1, X_2, \dots, X_n$  的任意的线性组合

$$l_1X_1 + l_2X_2 + \dots + l_nX_n$$

服从一维正态分布, 其中 $l_1, l_2, \dots, l_n$  不全为0。

3) 若 $(X_1, X_2, \dots, X_n)$ 服从n维正态分布, 设 $Y_1, Y_2, \dots, Y_k$  是 $X_1, X_2, \dots, X_n$ 的线性函数, 则 $(Y_1, Y_2, \dots, Y_k)$  也服从多维正态分布。

这一性质被称为**正态变量的线性变换不变性**。



# n维正态分布

## n维正态随机变量的重要性质

4) 设 $(X_1, X_2, \dots, X_n)$ 服从n维正态分布, 则“ $X_1, X_2, \dots, X_n$ 相互独立”和“ $X_1, X_2, \dots, X_n$ 两两不相关”是等价的

**[定理]** 设 $X \sim N(\mu, \sigma^2)$ ,  $B$  是 $n$  维的可逆矩阵, 设 $Y = BX$ , 则

$$Y \sim N(B\mu, B\sigma^2 B^T)$$





谢谢！

