概率统计与随机过程

# 第三章：随机变量的数字特征

赵俊舟

junzhou.zhao@xjtu.edu.cn

2025 年 3 月 28 日

# 目录

# 目录

# 目录

# 数学期望的来源：分赌本问题

**例 (分赌本问题，17 世纪中叶)**

甲乙两赌徒赌技不相上下，各出赌注 50 法郎，每局无平局。事先约定谁先赢三局则得全部赌注 100 法郎。当甲赢了两局，乙赢了一局时，赌博终止，现问这 100 法郎如何分才公平？

- 这个问题是 17 世纪中叶一个很有名的问题，标志着概率论的诞生。
- 平均分对甲不公平，全给甲对乙不公平。
- 合理的分法是：按一定比例，甲多分些，乙少分些。问题是如何确定这个合理比例？
- 1645 年，法国数学家帕斯卡（1623–1662）提出如下解法：
- 设想再赌下去，则甲最终所得 $X$ 为一随机变量，其可能取值为 0 或 100。

# 数学期望的来源：分赌本问题

- 再赌两局必可结束，其结果无外乎甲甲、甲乙、乙甲、乙乙四种情况。
- 这四种情况中有三种可使甲赢得 100 法郎，只有一种情况下（乙乙）甲赢得 0 法郎。
- 因为赌技不相上下，所以上面四种情况是等可能的，那么甲赢得 100 法郎的概率为 3/4，赢得 0 法郎的概率为 1/4。
- 即 $X$ 的分布律为：

| $x_i$ | 0 | 100 |
|---|---|---|
| $p_i$ | 0.25 | 0.75 |

- 帕斯卡认为，甲的**期望所得**应为：

$$0 \times 0.25 + 100 \times 0.75 = 75$$

- 故应给甲 75 法郎，给乙 25 法郎。

# 数学期望的概念

- 每个随机变量都有一个概率分布，例如分布律、概率密度或分布函数。
- 概率分布全面描述了随机变量的统计规律，由概率分布可以计算出有关随机变量的事件的概率。
- 用概率分布去刻画随机变量的优点是：全面、详细、完整
- 但也存在很明显的缺点：复杂、重点不突出、不能宏观地描述随机变量的统计特性
- 所以，怎样才能宏观地描述随机变量的统计特性?
- 简单明了、特征鲜明、直观实用

# 数学期望的概念

**例 (评估射击水平)**

甲、乙两射手进行打靶训练，每人各打 100 发子弹，成绩如下：

甲成绩

| 环数 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|
| 中靶数 | 17 | 28 | 45 | 10 |

乙成绩

| 环数 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|
| 中靶数 | 32 | 28 | 22 | 18 |

问怎样评估两人的射击水平？

分别计算两人的**总环数**：

- 甲： $7 \times 17 + 8 \times 28 + 9 \times 45 + 10 \times 10 = 848$
- 乙： $7 \times 32 + 8 \times 28 + 9 \times 22 + 10 \times 18 = 826$

# 数学期望的概念

## 例 (评估射击水平)

甲、乙两射手进行打靶训练，每人各打 100 发子弹，成绩如下：

甲成绩

| 环数 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|
| 中靶数 | 17 | 28 | 45 | 10 |

乙成绩

| 环数 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|
| 中靶数 | 32 | 28 | 22 | 18 |

问怎样评估两人的射击水平？

分别计算两人的**平均环数**：

- 甲： $7\times\frac{17}{100}+8\times\frac{28}{100}+9\times\frac{45}{100}+10\times\frac{10}{100}=8.48$
- 乙： $7\times\frac{32}{100}+8\times\frac{28}{100}+9\times\frac{22}{100}+10\times\frac{18}{100}=8.26$

# 数学期望的概念

- 用“频率近似概率”的思想，得到甲、乙两射手击中环数的分布律如下：

甲分布律

| 环数 $x_i$ | 7 | 8 | 9 | 10 |
|---|---|---|---|---|
| 概率 $p_i$ | 0.17 | 0.28 | 0.45 | 0.10 |

乙分布律

| 环数 $y_j$ | 7 | 8 | 9 | 10 |
|---|---|---|---|---|
| 概率 $p_j$ | 0.32 | 0.28 | 0.22 | 0.18 |

- 则计算出两人的平均环数分别为：

$$\sum_{i=1}^{n} x_i p_i = 7 \times 0.17 + 8 \times 0.28 + 9 \times 0.45 + 10 \times 0.10 = 8.48$$

$$\sum_{j=1}^{n} y_j p_j = 7 \times 0.32 + 8 \times 0.28 + 9 \times 0.22 + 10 \times 0.18 = 8.26$$

# 数学期望的概念

平均值的概念广泛存在:

- 某课程考试的平均成绩;
- 电子产品的平均无故障时间;
- 某地区的日平均气温和日平均降水量;
- 某地区水稻的平均亩产量;
- 某地区的家庭平均年收入;
- 某国家或地区人的平均寿命;

# 离散型随机变量的数学期望

## 定义 (离散型随机变量的数学期望)

设离散型随机变量 $X$ 的分布律为

$$P(X=x_i)=p_i,\quad i=1,2,\ldots$$

若级数 $\sum_{i=1}^{\infty} x_i p_i$ 绝对收敛到

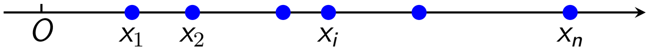
$$\mathbb{E}\left(X\right)\triangleq\sum_{i=1}^{\infty} x_i p_i$$

则称 $\mathbb{E}\left(X\right)$ 为随机变量 $X$ 的**数学期望**。

- 数学期望也称为期望、均值或加权平均。
- $\sum_{i=1}^{\infty} x_i p_i$ 绝对收敛保证了 $\sum_{i=1}^{\infty} x_i p_i$ 与求和次序无关。
- 若级数 $\sum_{i=1}^{\infty} x_i p_i$ 不绝对收敛，则称数学期望 $\mathbb{E}\left(X\right)$ 不存在。

# 离散型随机变量数学期望的物理意义

- 散布在直线上的 $n$ 个质点，坐标为 $x_i$，质量为 $p_i$。




- 则该系统对原点 $O$ 的力矩为 $\sum_{i=1}^{n} x_i p_i$。
- 该系统的质心坐标为

$$\frac{\sum_{i=1}^{n} x_i p_i}{\sum_{i=1}^{n} p_i}$$

- 当总质量 $\sum_{i=1}^{n} p_i = 1$ 时，质心坐标为

$$\mathbb{E}(X) = \sum_{i=1}^{n} x_i p_i$$

# 举例：离散型随机变量的数学期望

**例 (遇到红灯的平均数量)**

从学校去火车站要经过 3 个路口，在每个路口遇到红绿的事件是独立的，其概率均为 0.5，记 $X$ 表示途中遇到的红灯数，求数学期望 $\mathbb{E}(X)$。

$X \sim B(3, 0.5)$，其分布律为

| $x_i$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $p_i$ | 1/8 | 3/8 | 3/8 | 1/8 |

所以

$$\mathbb{E}(X) = \sum_{i=0}^{3} x_i p_i = 1.5$$

即从学校乘车去火车站平均要遇到 1.5 次红灯。

# 离散型随机变量的数学期望

## 例 (二项分布的数学期望)

设 $X \sim B(n,p)$，求 $\mathbb{E}(X)$。

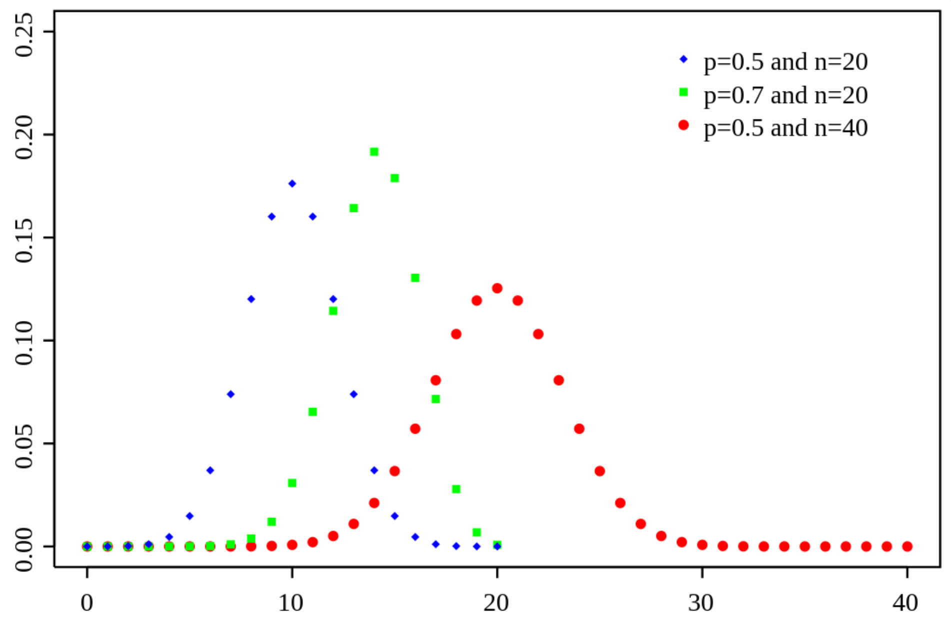
二项分布的分布律为

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}, \quad k=0,\ldots,n$$

利用 $k\binom{n}{k}=n\binom{n-1}{k-1}, k=1,\ldots,n$，可得期望为

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k=0}^{n} kP(X=k)=\sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} \\ &= np\sum_{k=1}^{n}\binom{n-1}{k-1}p^{k-1}(1-p)^{n-k} \\ &= np\end{aligned}$$

# 离散型随机变量的数学期望



# 离散型随机变量的数学期望

**例 (泊松分布的数学期望)**

设 $X \sim P(\lambda)$，求 $\mathbb{E}(X)$。

泊松分布的分布律为

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}, \quad k=0,1,\ldots$$

数学期望为

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

# 连续型随机变量的数学期望

## 定义 (连续型随机变量的数学期望)

设连续型随机变量 $X$ 的概率密度为 $f(x)$。若 $\int_{-\infty}^{+\infty} xf(x)\,\mathrm{d}x$ 绝对收敛到

$$\mathbb{E}(X) \triangleq \int_{-\infty}^{+\infty} xf(x)\,\mathrm{d}x$$

则称 $\mathbb{E}(X)$ 为随机变量 $X$ 的**数学期望**。

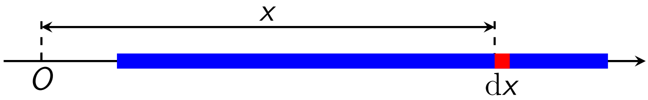
## 例 (均匀分布的数学期望)

设 $X \sim U(a, b)$，求 $X$ 的数学期望。

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} xf(x)\,\mathrm{d}x = \frac{1}{b-a}\int_a^b x\,\mathrm{d}x = \frac{a+b}{2}$$

# 连续型随机变量数学期望的物理意义

- 考虑 $x$ 轴上一根线密度为 $f(x)$ 的细棒。




- 用微元法求得长为 $\mathrm{d}x$ 的细棒对原点 $O$ 的力矩为 $xf(x)\,\mathrm{d}x$。
- 则细棒的质心坐标为

$$\frac{\int_{-\infty}^{+\infty} xf(x)\,\mathrm{d}x}{\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x}$$

- 当总质量 $\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1$ 时，质心坐标为

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} xf(x)\,\mathrm{d}x$$

# 连续型随机变量的数学期望

**例 (指数分布的数学期望)**

设 $X \sim \exp(\lambda)$，求 $\mathbb{E}(X)$。

$X$ 的概率密度为

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

数学期望为

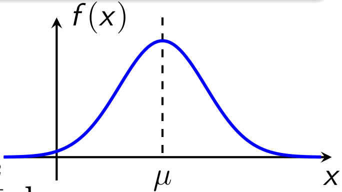
$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f(x)\,\mathrm{d}x = \int_0^{\infty} x\lambda e^{-\lambda x}\,\mathrm{d}x$$

$$= -\int_0^{\infty} x\,\mathrm{d}e^{-\lambda x} = \int_0^{\infty} e^{-\lambda x}\,\mathrm{d}x = \frac{1}{\lambda}$$

# 连续型随机变量的数学期望

## 例 (正态分布的数学期望)

设 $X \sim N(\mu, \sigma^2)$，求 $\mathbb{E}(X)$。




$$
\begin{aligned}
\mathbb{E}(X) &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty} (x-\mu+\mu) e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,\mathrm{d}x \\
&\overset{t=x-\mu}{=} \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty} t e^{-\frac{t^2}{2\sigma^2}}\,\mathrm{d}t + \frac{\mu}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,\mathrm{d}x \\
&= 0 + \mu \\
&= \mu
\end{aligned}
$$

# 目录

# 随机变量函数的数学期望

实际背景问题:

- 设风速 $V \sim U(0, a)$，飞机机翼受到的压力为 $W = kV^2$，其中 $k$ 是正常数，问机翼受到的平均压力为多大？
- 在机械加工中，工件的直径 $D \sim N(\mu, \sigma^2)$，问工件的平均截面积是多大？
- 在一个电路中, 流过电阻 $R$ 的电流 $i$ 是一个概率密度为 $f(i)$ 的随机变量，问电阻 $R$ 消耗的平均功率是多大？

# 随机变量函数的数学期望

**问题：求随机变量函数的数学期望**

已知随机变量 $X$ 的概率密度为 $f(x)$，$y=g(x)$ 为连续函数，求数学期望 $\mathbb{E}\left(g(X)\right)$？

一般思路：

- $X \sim f(x), y = g(x) \quad \Rightarrow \quad Y \sim f_Y(y)$
- $\mathbb{E}\left(g(X)\right) = \mathbb{E}\left(Y\right) = \int_{-\infty}^{+\infty} y f_Y(y)\,\mathrm{d}y$

难点：

- 求 $Y$ 的概率密度比较难；
- 整个计算过程比较复杂。

# 随机变量函数的数学期望

- 若 $g(x)$ 单调且连续可导，则其反函数 $x = h(y)$ 存在且连续可导，且有 $f_Y(y) = f(h(y))|h'(y)|$。
- 不妨设 $g(x)$ 单调增，此时 $f_Y(y) = f(h(y))h'(y)$，故期望为

$$
\begin{aligned}
\mathbb{E}(g(X)) &= \mathbb{E}(Y) = \int_{-\infty}^{+\infty} y f_Y(y)\,\mathrm{d}y \\
&= \int_{-\infty}^{+\infty} y f(h(y))h'(y)\,\mathrm{d}y \overset{x=h(y)}{=} \int_{-\infty}^{+\infty} g(x)f(x)\,\mathrm{d}x
\end{aligned}
$$

- 有意思的结果：

$$
\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f(x)\,\mathrm{d}x \qquad \mathbb{E}(g(X)) = \int_{-\infty}^{+\infty} g(x) f(x)\,\mathrm{d}x
$$

- 这个结论对函数 $g(x)$ 非单调时也成立。

# 随机变量函数的数学期望

## 定理

设 $g(x)$ 为连续函数，则

- 设 $X$ 为离散型随机变量，其分布律为
$$P(X = x_i) = p_i, \quad i = 1, 2, \ldots$$
若 $\sum_{i=1}^{\infty} g(x_i)p_i$ 绝对收敛，则
$$\mathbb{E}\left(g(X)\right) = \sum_{i=1}^{\infty} g(x_i)p_i$$
- 设 $X$ 为连续型随机变量，其概率密度为 $f(x)$，若 $\int_{-\infty}^{+\infty} g(x)f(x)\,\mathrm{d}x$ 绝对收敛，则
$$\mathbb{E}\left(g(X)\right) = \int_{-\infty}^{+\infty} g(x)f(x)\,\mathrm{d}x$$

# 随机变量函数的数学期望

**例**

设风速 $V \sim U(0, a)$，飞机机翼受到的压力为 $W = kV^2$，其中 $k$ 是正常数，问机翼受到的平均压力为多大？

$V$ 的概率密度为

$$f(v) = \begin{cases} 1/a, & 0 < v < a, \\ 0, & \text{其他} \end{cases}$$

故机翼受到的平均压力为

$$\mathbb{E}(W) = \mathbb{E}\left(kv^2\right) = \int_{-\infty}^{+\infty} kv^2 f(v)\,\mathrm{d}v = \frac{k}{a}\int_0^a v^2\,\mathrm{d}v = \frac{1}{3}ka^2$$

# 随机变量函数的数学期望

## 例

设离散型随机变量 $X$ 的分布律为

| $x_i$ | $-1$ | $0$ | $1$ | $2$ |
|---|---|---|---|---|
| $p_i$ | $0.1$ | $0.3$ | $0.4$ | $0.2$ |

求 $\mathbb{E}\left(X^2\right), \mathbb{E}\left(3X+1\right)$。

$$\mathbb{E}\left(X^2\right) = (-1)^2 \times 0.1 + 0 \times 0.3 + 1^2 \times 0.4 + 2^2 \times 0.2 = 1.3$$
$$\mathbb{E}\left(3X+1\right) = -2 \times 0.1 + 1 \times 0.3 + 4 \times 0.4 + 7 \times 0.2 = 3.1$$

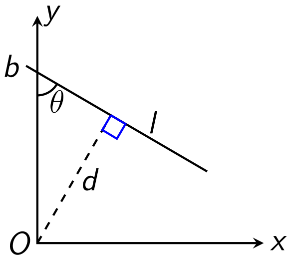
# 随机变量函数的数学期望

**例**

过平面上点 $(0, b)$ 任作一条直线 $l$，求由坐标原点 $O$ 到直线 $l$ 的平均距离。

设 $l$ 与 $y$ 轴的夹角为 $\theta$，则 $\theta \sim U(0, \pi)$。于是

$$d = |b| \sin\theta$$




所以

$$\begin{aligned}
\mathbb{E}(d) &= \mathbb{E}(|b| \sin\theta) \\
&= |b| \int_0^{\pi} \sin\theta \cdot \frac{1}{\pi} \, \mathrm{d}\theta \\
&= \frac{|b|}{\pi} \int_0^{\pi} \sin\theta \, \mathrm{d}\theta = \frac{2|b|}{\pi}
\end{aligned}$$

# 二维随机变量函数的数学期望

## 定理

设 $g(x,y)$ 为二元函数，则

- 设离散型随机变量 $(X,Y)$ 的联合分布律为
$$P(X=x_i, Y=y_j)=p_{ij} \qquad i,j=1,2,\ldots$$
若 $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} g(x_i,y_j)p_{ij}$ 绝对收敛，则
$$\mathbb{E}\left(g(X,Y)\right)=\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} g(x_i,y_j)p_{ij}$$
- 设连续型随机变量 $(X,Y)$ 的联合概率密度为 $f(x,y)$，若 $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f(x,y)\,\mathrm{d}x\,\mathrm{d}y$ 绝对收敛，则
$$\mathbb{E}\left(g(X,Y)\right)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

# 二维随机变量函数的数学期望

**例**

设 $(X,Y)$ 服从半圆域 $G: x^2+y^2\le 1, y\ge 0$ 上的均匀分布，计算 $X, Y$ 以及 $X^2Y$ 的数学期望。

$(X,Y)$ 的联合概率密度为 $f(x,y)=\begin{cases}\frac{2}{\pi}, & x^2+y^2\le 1, y\ge 0,\\ 0, & \text{其他}\end{cases}$

$$\mathbb{E}(X)=\iint_G x f(x,y)\,\mathrm{d}x\,\mathrm{d}y=\frac{2}{\pi}\int_{-1}^{1}x\,\mathrm{d}x\int_0^{\sqrt{1-x^2}}\mathrm{d}y=0$$

$$\mathbb{E}(Y)=\iint_G y f(x,y)\,\mathrm{d}x\,\mathrm{d}y=\frac{2}{\pi}\int_{-1}^{1}\mathrm{d}x\int_0^{\sqrt{1-x^2}}y\,\mathrm{d}y=\frac{4}{3\pi}$$

$$\mathbb{E}(X^2Y)=\iint_G x^2 y f(x,y)\,\mathrm{d}x\,\mathrm{d}y=\frac{2}{\pi}\int_{-1}^{1}x^2\,\mathrm{d}x\int_0^{\sqrt{1-x^2}}y\,\mathrm{d}y=\frac{4}{15\pi}$$

# 二维随机变量函数的数学期望

**例**

设 $(X, Y)$ 服从单位圆域 $G: x^2 + y^2 \le 1$ 上的均匀分布，计算数学期望 $\mathbb{E}(XY)$。

由前面定理得

$$\mathbb{E}(XY) = \iint_G xyf(x,y)\,\mathrm{d}x\,\mathrm{d}y = \iint_G xy\frac{1}{\pi}\,\mathrm{d}x\,\mathrm{d}y$$

令 $x = \rho\cos\theta, y = \rho\sin\theta, \theta \in [0, 2\pi], \rho \in [0, 1]$，所以

$$\mathbb{E}(XY) = \frac{1}{\pi}\int_0^{2\pi}\int_0^1 \rho^2\cos\theta\sin\theta\rho\,\mathrm{d}\rho\,\mathrm{d}\theta$$

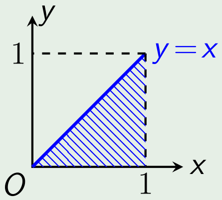
$$= \frac{1}{2\pi}\int_0^1 \rho^3\,\mathrm{d}\rho\int_0^{2\pi}\sin(2\theta)\,\mathrm{d}\theta = 0$$

# 二维随机变量函数的数学期望

**例**

设随机变量 $(X, Y)$ 的密度函数为

$$f(x,y)=\begin{cases}15xy^2, & 0\le y\le x\le 1,\\ 0, & \text{其他}\end{cases}$$




计算数学期望 $\mathbb{E}(X), \mathbb{E}(Y), \mathbb{E}(XY)$。

$$\mathbb{E}(X)=\iint_G xf(x,y)\,\mathrm{d}x\,\mathrm{d}y=15\int_0^1 x^2\,\mathrm{d}x\int_0^x y^2\,\mathrm{d}y=\frac{5}{6}$$

$$\mathbb{E}(Y)=\iint_G yf(x,y)\,\mathrm{d}x\,\mathrm{d}y=15\int_0^1 x\,\mathrm{d}x\int_0^x y^3\,\mathrm{d}y=\frac{5}{8}$$

$$\mathbb{E}(XY)=\iint_G xyf(x,y)\,\mathrm{d}x\,\mathrm{d}y=15\int_0^1 x^2\,\mathrm{d}x\int_0^x y^3\,\mathrm{d}y=\frac{15}{28}$$

# 二维随机变量函数的数学期望

**例**

某产品出售一件可获利 $m$ 元，而积压一件则需要支付 $n$ 元保管费。该产品的需求量为 $Y$ 件，服从参数为 $\lambda$ 的指数分布。要使销售该产品所获利润最大，需要生产该产品多少件？

需要生产 $x$ 件产品，销售量为 $Y$ 时利润为 $Q$，则

$$Q(Y) = \begin{cases} mY - n(x - Y), & Y < x \\ mx, & Y \geq x \end{cases}$$

即 $Q$ 是随机变量 $Y$ 的函数，而 $Y$ 服从参数为 $\lambda$ 的指数分布，其概率密度为

$$f_Y(y) = \begin{cases} \lambda e^{-\lambda y}, & y > 0 \\ 0, & y \leq 0 \end{cases}$$

# 二维随机变量函数的数学期望

故 $Q$ 的数学期望为

$$\begin{aligned}\mathbb{E}(Q) &= \int_{-\infty}^{+\infty} Q(y) f_Y(y)\,\mathrm{d}y \\ &= \int_0^x (my - n(x-y))\lambda e^{-\lambda y}\,\mathrm{d}y + \int_x^{+\infty} mx\lambda e^{-\lambda y}\,\mathrm{d}y \\ &= \frac{m+n}{\lambda} - \frac{m+n}{\lambda}e^{-\lambda x} - nx\end{aligned}$$

令

$$\frac{\mathrm{d}\mathbb{E}(Q)}{\mathrm{d}x} = (m+n)e^{-\lambda x} - n = 0$$

得 $x^* = -\frac{1}{\lambda}\ln\frac{n}{m+n}$。

又因为 $\frac{\mathrm{d}^2\mathbb{E}(Q)}{\mathrm{d}x^2} = -\lambda(m+n)e^{-\lambda x} \leq 0$，所以此时的极值 $x^*$ 也是最大值。

# 目录

# 数学期望的性质

- 观察数学期望的计算公式，有什么数学特点？
- 对于离散型随机变量

$$\mathbb{E}\left(X\right) \triangleq \sum_{i=1}^{\infty} x_i p_i$$

- 对于连续型随机变量

$$\mathbb{E}\left(X\right) \triangleq \int_{-\infty}^{+\infty} xf\left(x\right) \mathrm{d}x$$

- 在分布律或概率密度给定的情况下，数学期望本质是一种线性运算。

# 数学期望的性质

## 定理（数学期望的性质）

1. 若 $a \le X \le b$，则 $a \le \mathbb{E}(X) \le b$（$a, b$ 均为常数）。
2. 若 $c$ 为常数，则 $\mathbb{E}(cX) = c\mathbb{E}(X)$
3. $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$
4. 若 $X, Y$ 相互独立，则 $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$

## 推论

- 若 $X = c$ 为常数，则 $\mathbb{E}(X) = c$
- 若 $c_1, \ldots, c_n$ 均为常数，则 $\mathbb{E}\left(\sum_{i=1}^n c_i X_i\right) = \sum_{i=1}^n c_i \mathbb{E}(X_i)$
- 若 $X_1, \ldots, X_n$ 相互独立，则 $\mathbb{E}(X_1 \cdots X_n) = \mathbb{E}(X_1) \cdots \mathbb{E}(X_n)$

# 数学期望的性质

**性质四证明.**

设 $(X, Y) \sim f(x, y)$，$X \sim f_X(x)$，$Y \sim f_Y(y)$。若 $X, Y$ 相互独立，则 $f(x, y) = f_X(x) f_Y(y)$。所以

$$
\begin{aligned}
\mathbb{E}(XY) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf(x, y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf_X(x) f_Y(y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{-\infty}^{+\infty} xf_X(x)\,\mathrm{d}x \int_{-\infty}^{+\infty} yf_Y(y)\,\mathrm{d}y \\
&= \mathbb{E}(X)\,\mathbb{E}(Y)
\end{aligned}
$$

$\square$

# 数学期望的性质

## 例 (电梯停的次数的期望)

一电梯载有 12 位乘客自一楼至十一楼，如到达某楼层没有乘客下电梯，则该层不停。以 $X$ 表示电梯停的次数，求 $\mathbb{E}(X)$。（假设每个乘客在任一层下电梯是等可能的且各乘客是否下电梯相互独立）。

- 令 $X_i = \begin{cases} 1, & \text{第 } i \text{ 层停} \\ 0, & \text{第 } i \text{ 层不停} \end{cases}$ $i = 2, \ldots, 11$，则 $X = \sum_{i=2}^{11} X_i$
- 电梯在某层不停，等价于这 12 个人都不去该层，而是另外 9 层，所以有 $P(X_i = 0) = (9/10)^{12}, \forall i$。
- 于是期望为

$$\mathbb{E}(X) = \mathbb{E}\left(\sum_{i=2}^{11} X_i\right) = \sum_{i=2}^{11} \mathbb{E}(X_i) = 10P(X_i = 1) = 7.18$$

# 数学期望的性质

## 例 (能打开房门的期望)

旅游团的 $N$ 个游客出酒店时都将自己房间的钥匙交给了导游。回到酒店后，每人从导游处任取一把钥匙去开自己房间的门。试问平均有多少人能开打房门。

- 令 $X_i = \begin{cases} 1, & \text{第 } i \text{ 人能打开房门} \\ 0, & \text{第 } i \text{ 人不能打开房门} \end{cases}, \quad i = 1, \ldots, N$
- 任意一个人能打开房门的概率为 $P(X_i = 1) = 1/N$
- 于是

$$\mathbb{E}\left(\sum_{i=1}^{N} X_i\right) = \sum_{i=1}^{N} \mathbb{E}(X_i) = 1$$

# 数学期望的性质

**例 (超几何分布的期望)**

设 $N$ 件产品中有 $M$ 件次品，在该批产品中任意取 $n$ 件，记 $X$ 表示取出的次品个数，求 $\mathbb{E}(X)$。

- $X$ 的分布律为

$$p_k = P(X = k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}, \quad k = 0, \ldots, n$$

- 称 $X$ 服从超几何分布。
- 期望为

$$\mathbb{E}(X) = \sum_{k=0}^{n} k p_k = \sum_{k=0}^{n} k \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$$

# 数学期望的性质

- 二项分布有性质

$$\sum_{i=0}^{k} C_n^i C_m^{k-i} = C_{m+n}^k$$

和

$$kC_n^k = nC_{n-1}^{k-1}, \quad k = 1, 2, \ldots, n$$

- 所以

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k=0}^{n} kp_k = \sum_{k=0}^{n} k\frac{C_M^k C_{N-M}^{n-k}}{C_N^n} = \sum_{k=1}^{n} M\frac{C_{M-1}^{k-1} C_{N-M}^{n-k}}{C_N^n} \\ &= M\frac{C_{N-1}^{n-1}}{C_N^n} = \frac{MN}{N}\frac{C_{N-1}^{n-1}}{C_N^n} = \frac{Mn}{N}\frac{C_N^n}{C_N^n} = \frac{Mn}{N}\end{aligned}$$

# 数学期望的性质：利用随机变量的分解求解

- 令 $X_i = \begin{cases} 1, & \text{第 } i \text{ 件取出的是次品} \\ 0, & \text{第 } i \text{ 件取出的是正品} \end{cases}$，$i = 1, \ldots, n$，则

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

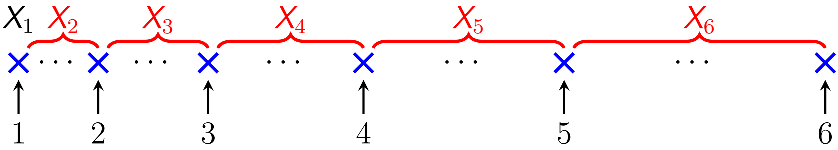
- 由排队抽签问题的结论知 $P(X_i = 1) = M/N$。
- 所以

$$\mathbb{E}(X) = \mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{E}(X_i) = \sum_{i=1}^{n} P(X_i = 1) = \frac{Mn}{N}$$

# 数学期望的性质

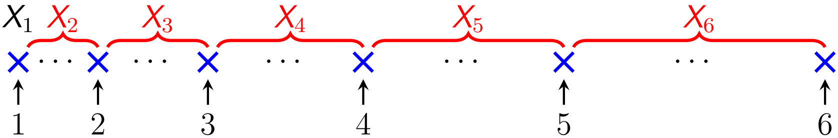
**例**

掷一颗骰子直到所有点数全部出现，求所需投掷次数 $X$ 的期望。




- $X_1$： 出现第 1 个点需要的的投掷次数（$X_1 = 1$）；
- $X_2$： 出现第 1 个点后等待第 2 个不同点需要的投掷次数；
- $X_3$： 出现第 2 个点后等待第 3 个不同点需要的投掷次数；
- 同理定义 $X_4, X_5, X_6$

# 数学期望的性质




- 用 $p_i$ 表示已经出现了 $i-1$ 个不同点，当前投掷出现第 $i$ 个点的概率，则

$$p_i = 1 - \frac{i-1}{6}, \quad i = 1, \ldots, 6$$

- 得到

$$P(X_i = k) = (1-p_i)^{k-1} p_i, \quad k = 1, 2, \ldots$$

- 从而得到 $\mathbb{E}(X_i) = 1/p_i$。
- 所以 $\mathbb{E}(X) = \sum_{i=1}^{6} \mathbb{E}(X_i) = \sum_{i=1}^{6} \frac{1}{p_i} = 14.7$

# 数学期望的性质

**例**

一公司经营某种原料，该原料的市场需求 $X \sim U(300, 500)$（单位：吨）。每出售一吨原料可获利 1 千元，若积压一吨，则要损失 0.5 千元。问公司应该组织多少货源可使收益最大？

设组织货源 $a$ 吨，$a \in [300, 500]$ 时，可获利 $Y$ 千元，则

$$Y = g(X) = \begin{cases} a, & a \le X \le 500, \\ X - 0.5(a - X), & 300 \le X < a \end{cases}$$

于是公司平均获利为

$$\mathbb{E}(Y) = \mathbb{E}(g(X)) = \int_{-\infty}^{+\infty} g(x) f(x)\,\mathrm{d}x = \frac{1}{800}(-3a^2 + 2600a - 90000)$$

得 $a^* \approx 433.3$

# 数学期望的性质

**例**

设 $X \sim B(n, p)$ 试由期望性质求 $\mathbb{E}(X)$。

- 注意到 $X$ 是 $n$ 次重复独立伯努利试验中成功的试验的次数，且每次试验事件发生的概率为 $p$。
- 令 $X_i = \begin{cases} 1, & \text{第 } i \text{ 次试验成功} \\ 0, & \text{第 } i \text{ 次试验失败} \end{cases} \quad i = 1, \dots, n$
- 则 $P(X_i = 1) = p$，得到 $\mathbb{E}(X_i) = p$。
- 由 $X = \sum_{i=1}^{n} X_i$，得到

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

# 目录

# 目录

# 方差的概念

**例**

甲、乙两射手击中环数分别为 $X, Y$，其分布律为

甲分布律

| 环数 $x_i$ | 7 | 8 | 9 | 10 |
|---|---|---|---|---|
| 概率 $p_i$ | 0.17 | 0.28 | 0.45 | 0.10 |

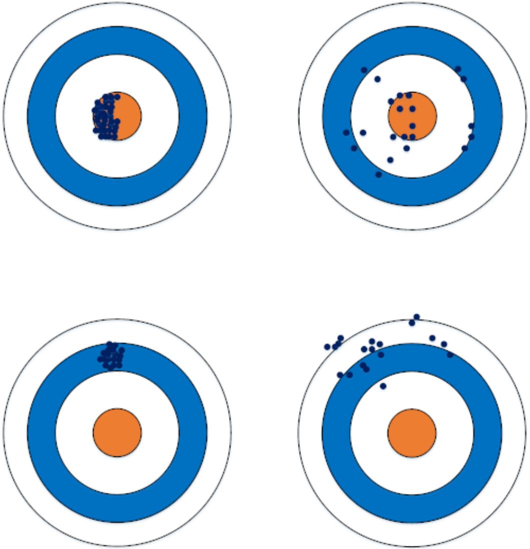
乙分布律

| 环数 $y_j$ | 7 | 8 | 9 | 10 |
|---|---|---|---|---|
| 概率 $p_j$ | 0.32 | 0.28 | 0.22 | 0.18 |

问怎样评估两人的射击水平？

- 甲、乙两射手击中环数的数学期望分别为
$$\mathbb{E}(X) = \sum_i x_i p_i = 8.48 \qquad \mathbb{E}(Y) = \sum_j y_j p_j = 8.26$$
- 所以，从平均水平来说甲的射击水平略高。
- 哪位射手的稳定性好？怎样用数字特征描述稳定性？

# 方差的概念



期望可以区分上下的分布差异，但不能区分左右的分布差异。

# 方差的概念

**例**

甲、乙两种品牌的手表其走时误差（秒）分别为

甲分布律

| 误差 $x_i$ | $-1$ | $0$ | $1$ |
|---|---|---|---|
| 概率 $p_i$ | $0.1$ | $0.8$ | $0.1$ |

乙分布律

| 误差 $y_j$ | $-2$ | $-1$ | $0$ | $1$ | $2$ |
|---|---|---|---|---|---|
| 概率 $p_j$ | $0.2$ | $0.1$ | $0.4$ | $0.1$ | $0.2$ |

- 易知 $\mathbb{E}(X) = \mathbb{E}(Y) = 0$
- 能认为两种手表质量一样吗？否！甲品牌手表的质量较好。
- 怎样用数字特征描述离散性？

# 方差的概念

设 $X$ 为随机变量，数学期望为 $\mathbb{E}(X)$

- 考虑偏差： $X - \mathbb{E}(X)$ $\Leftarrow$ 正负偏差会抵消
- 考虑绝对偏差： $|X - \mathbb{E}(X)|$ $\Leftarrow$ 数学处理不方便
- 考虑平方偏差： $(X - \mathbb{E}(X))^2$ $\Leftarrow$ 随机变量非数字特征
- 考虑平均平方偏差： $\mathbb{E}\{(X - \mathbb{E}(X))^2\}$ $\Leftarrow$ 方差

# 方差的概念

**定义 (方差和标准差)**

设 $X$ 为随机变量，若 $\mathbb{E}\left([X-\mathbb{E}(X)]^2\right)$ 存在，则称 $\mathbb{E}\left([X-\mathbb{E}(X)]^2\right)$ 为随机变量 $X$ 的**方差**，记为 $D(X)$ 或 $\mathrm{var}(X)$，即

$$D(X)=\mathrm{var}(X)\triangleq\mathbb{E}\left([X-\mathbb{E}(X)]^2\right)$$

称 $\sqrt{D(X)}$ 为 $X$ 的**标准差**，记为 $\sigma(X)$。

- 令 $g(x)=(x-\mathbb{E}(X))^2$，则 $D(X)=\mathbb{E}(g(X))$
- 离散型随机变量：$D(X)=\sum_{i=1}^{\infty}[x_i-\mathbb{E}(X)]^2p_i$
- 连续型随机变量：$D(X)=\int_{-\infty}^{+\infty}[x-\mathbb{E}(X)]^2f(x)\,\mathrm{d}x$
- 常用计算公式：$D(X)=\mathbb{E}\left(X^2\right)-\mathbb{E}(X)^2$

# 方差的概念

## 例 (射击水平评价)

甲、乙两射手击中环数分别为 $X, Y$，其分布律为

甲分布律

| 环数 $x_i$ | 7 | 8 | 9 | 10 |
|---|---|---|---|---|
| 概率 $p_i$ | 0.17 | 0.28 | 0.45 | 0.10 |

乙分布律

| 环数 $y_j$ | 7 | 8 | 9 | 10 |
|---|---|---|---|---|
| 概率 $p_j$ | 0.32 | 0.28 | 0.22 | 0.18 |

计算 $D(X), D(Y)$。

已算得 $\mathbb{E}(X) = 8.48, \mathbb{E}(Y) = 8.26$，所以

$$D(X) = \sum_i (x_i - 8.48)^2 p_i = 0.7896$$

$$D(Y) = \sum_j (y_j - 8.26)^2 p_j = 1.1924$$

甲的平均水平比较高且更稳定。

# 方差的概念

## 例 (均匀分布的方差)

设 $X \sim U(a, b)$，计算 $D(X)$。

均匀分布随机变量 $X$ 的概率密度为

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

所以

$$\mathbb{E}\left(X^2\right) = \int_a^b x^2 \frac{1}{b-a} \mathrm{d}x = \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + ab + b^2}{3}$$

所以

$$D(X) = \mathbb{E}\left(X^2\right) - \mathbb{E}\left(X\right)^2 = \frac{a^2 + ab + b^2}{3} - (\frac{a+b}{2})^2 = \frac{(a-b)^2}{12}$$

# 方差的概念

## 例 (泊松分布的方差)

设 $X$ 服从参数为 $\lambda > 0$ 的泊松分布，即 $X \sim P(\lambda)$，计算 $D(X)$

利用公式 $D(X) = \mathbb{E}\left(X^2\right) - \mathbb{E}(X)^2$ 计算，已算得 $\mathbb{E}(X) = \lambda$，又

$$\begin{aligned}
\mathbb{E}\left(X^2\right) &= \mathbb{E}(X(X-1)+X) \\
&= \sum_{k=0}^{\infty} k(k-1)\frac{\lambda^k}{k!}e^{-\lambda} + \lambda \\
&= \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!}e^{-\lambda} + \lambda \\
&= \lambda^2 + \lambda
\end{aligned}$$

所以 $D(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$，表明泊松分布的方差也是参数 $\lambda$。

# 方差的概念

**例 (正态分布的方差)**

设 $X \sim N(\mu, \sigma^2)$，求 $D(X)$。

因为 $\mathbb{E}(X) = \mu$，故

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

$$\overset{t=\frac{x-\mu}{\sigma}}{=} \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t^2 e^{-\frac{t^2}{2}} \mathrm{d}t$$

$$= \sigma^2$$

# 方差的概念

**例 (指数分布的方差)**

设 $X$ 服从参数为 $\lambda > 0$ 的指数分布，求 $D(X)$。

指数分布随机变量 $X$ 的概率密度为

$$f(X) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \le 0 \end{cases}$$

已算得 $\mathbb{E}(X) = 1/\lambda$，又

$$\mathbb{E}(X^2) = \int_0^{+\infty} x^2 \lambda e^{-\lambda x} \, \mathrm{d}x = \frac{2}{\lambda^2}$$

所以

$$D(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{1}{\lambda^2}$$

# 目录

# 方差的性质

1. 若 $X = c$ 为常数，则 $D(X) = 0$
2. 若 $c$ 为常数，则 $D(cX) = c^2 D(X)$

**证明.**

$$
\begin{aligned}
D(cX) &= \mathbb{E}\left\{[cX - \mathbb{E}(cX)]^2\right\} \\
&= \mathbb{E}\left\{[cX - c\mathbb{E}(X)]^2\right\} \\
&= \mathbb{E}\left\{c^2[X - \mathbb{E}(X)]^2\right\} \\
&= c^2\mathbb{E}\left\{[X - \mathbb{E}(X)]^2\right\} \\
&= c^2 D(X)
\end{aligned}
$$

□

# 方差的性质

3. $D(X \pm Y) = D(X) + D(Y) \pm 2\mathbb{E}\{[X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]\}$
特别地，当 $X, Y$ 相互独立时，有 $D(X \pm Y) = D(X) + D(Y)$

**证明.**

$$[(X \pm Y) - \mathbb{E}(X \pm Y)]^2 = [(X - \mathbb{E}(X)) \pm (Y - \mathbb{E}(Y))]^2$$
$$= [X - \mathbb{E}(X)]^2 \pm 2[X - \mathbb{E}(X)][Y - \mathbb{E}(Y)] + [Y - \mathbb{E}(Y)]^2$$

- 两边同时求期望，根据方差的定义及期望的性质，得
$$D(X \pm Y) = D(X) + D(Y) \pm 2\mathbb{E}\{[X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]\}$$
- 当 $X, Y$ 相互独立时，$X - \mathbb{E}(X)$ 与 $Y - \mathbb{E}(Y)$ 也相互独立，由期望的性质得
$$\mathbb{E}\{[X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]\} = \mathbb{E}(X - \mathbb{E}(X))\,\mathbb{E}(Y - \mathbb{E}(Y)) = 0$$
- 所以 $D(X \pm Y) = D(X) + D(Y)$. □

# 方差的性质

4. $D(X)=0$ 的充要条件是 $P(X=\mathbb{E}(X))=1$

**引理 (Markov 不等式)**

对随机变量 $X$，若 $\mathbb{E}(|X|^r)<\infty, r>0$，则对任意 $\epsilon>0$，有

$$P(|X|\geq\epsilon)\leq\frac{\mathbb{E}(|X|^r)}{\epsilon^r}$$

# 方差的性质

**证明.**

设 $D(X)=0$，在 Markov 不等式中取 $r=2$，并用 $X-\mathbb{E}(X)$ 替换 $X$，得

$$P(|X-\mathbb{E}(X)|\geq\epsilon)\leq\frac{D(X)}{\epsilon^2}=0$$

因为 $\{|X-\mathbb{E}(X)|\neq 0\}=\bigcup_{n=1}^{\infty}\{|X-\mathbb{E}(X)|\geq\frac{1}{n}\}$，由概率的连续性得

$$P(|X-\mathbb{E}(X)|\neq 0)=\lim_{n\to\infty}P(|X-\mathbb{E}(X)|\geq\frac{1}{n})=0$$

即

$$P(|X-\mathbb{E}(X)|=0)=1$$

从而 $P(X=\mathbb{E}(X))=1$。 $\square$

# 方差的性质

5. 若 $X_1, \ldots, X_n$ 是相互独立的随机变量，则

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

**例**

设 $X \sim B(n, p)$，计算 $D(X)$

二项分布来自 $n$ 重伯努利试验，故有 $X = \sum_{i=1}^{n} X_i$，其中 $X_i = 1$ 表示第 $i$ 次试验时事件发生，否则 $X_i = 0$。所以

$$D(X) = D(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} D(X_i) = np(1-p)$$

# 目录

# 目录

# 如何刻画随机变量之间的关系？

- 设随机变量 $(X, Y) \sim f(x, y), X \sim f_X(x), Y \sim f_Y(y)$
- 则 $X, Y$ 相互独立 $\Leftrightarrow f(x, y) = f_X(x)f_Y(y)$
- 问题：若 $X, Y$ 不相互独立，如何刻画它们之间的关系？
- 若 $X, Y$ 相互独立，则
$$\mathbb{E}\{[X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]\} = 0$$
反之，若
$$\mathbb{E}\{[X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]\} \neq 0$$
则 $X, Y$ 必不相互独立。

# 协方差

## 定义（协方差，Covariance）

设随机变量 $X, Y$ 的方差都存在，称

$$\mathrm{Cov}(X, Y) \triangleq \mathbb{E}\left\{[X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]\right\}$$

为 $X, Y$ 的**协方差**。

易知 $D(X) = \mathbb{E}\left([X - \mathbb{E}(X)]^2\right) = \mathrm{Cov}(X, X)$

## 性质

1. **独立性**：若 $X, Y$ 相互独立，则 $\mathrm{Cov}(X, Y) = 0$。
2. **柯西-施瓦茨不等式**：若 $\mathbb{E}\left(X^2\right), \mathbb{E}\left(Y^2\right)$ 存在，则有 $\mathbb{E}(XY)^2 \le \mathbb{E}\left(X^2\right)\mathbb{E}\left(Y^2\right)$。特别地，$\mathrm{Cov}(X, Y)^2 \le D(X)D(Y)$

# 柯西-施瓦茨不等式证明

- 考虑如下关于实变量 $t$ 的函数
$$q(t)=\mathbb{E}\left([tX+Y]^2\right)=t^2\mathbb{E}\left(X^2\right)+2t\mathbb{E}\left(XY\right)+\mathbb{E}\left(Y^2\right)$$
- 对于任意实数 $t$，由于 $(tX+Y)^2$ 是取非负值的随机变量，所以其期望亦非负，即对于一切 $t$，有 $q(t)\geq 0$。
- 故二次方程 $q(t)=0$ 的判别式满足条件
$$4\mathbb{E}\left(XY\right)^2-4\mathbb{E}\left(X^2\right)\mathbb{E}\left(Y^2\right)\leq 0$$
- 所以
$$\mathbb{E}\left(XY\right)^2\leq\mathbb{E}\left(X^2\right)\mathbb{E}\left(Y^2\right)$$
该不等式也称为柯西-施瓦茨不等式。
- 令 $U=X-\mathbb{E}\left(X\right), V=Y-\mathbb{E}\left(Y\right)$，对 $U,V$ 应用上式得
$$\mathrm{Cov}(X,Y)^2\leq D(X)D(Y)$$

# 协方差的性质

3. $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
4. $\mathrm{Cov}(X, X) = D(X)$
5. $\mathrm{Cov}(aX, bY) = ab\mathrm{Cov}(X, Y)$
6. $\mathrm{Cov}(X_1 + X_2, Y_1 + Y_2) = \mathrm{Cov}(X_1, Y_1) + \mathrm{Cov}(X_1, Y_2) + \mathrm{Cov}(X_2, Y_1) + \mathrm{Cov}(X_2, Y_2)$
7. $D(X \pm Y) = D(X) + D(Y) \pm 2\mathrm{Cov}(X, Y)$
8. $\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$

# 协方差及随机变量之间的关系

- 随机变量 $X, Y$ 相互独立 $\Rightarrow \mathrm{Cov}(X, Y) = 0$
- 若 $\mathrm{Cov}(X, Y) \neq 0$ $\Rightarrow X, Y$ 不独立 $\Rightarrow X, Y$ 之间存在某种关系

**问题**

1. 协方差不为 0 时，随机变量之间的关系是什么关系？
2. 这种关系的强弱能否直接用 $\mathrm{Cov}(X, Y)$ 的大小来衡量？

# 协方差及随机变量之间的关系

- 对于问题（2），$\forall k \in (-\infty, +\infty)$，由协方差的性质，有
$$\mathrm{Cov}(kX, kY) = k^2\mathrm{Cov}(X, Y)$$

- 故用 $\mathrm{Cov}(X, Y)$ 衡量随机变量 $X, Y$ 之间的关系不合适。

- 考虑对随机变量进行“单位化”，令
$$X' = \frac{X - \mathbb{E}(X)}{\sqrt{D(X)}}, \quad Y' = \frac{Y - \mathbb{E}(Y)}{\sqrt{D(Y)}}$$

- 易知 $\mathbb{E}(X') = \mathbb{E}(Y') = 0, D(X') = D(Y') = 1$。

# 相关系数

## 定义（相关系数，Correlation Coefficient）

称

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

为随机变量 $X, Y$ 的**相关系数**。

## 相关系数的性质

1. 若 $X, Y$ 相互独立，则 $\rho_{XY} = 0$。
2. $|\rho_{XY}| \leq 1$。
3. $|\rho_{XY}| = 1$ 的充要条件是存在常数 $a, b$ 使 $P(Y = a + bX) = 1$。

性质三回答了问题（1）：相关系数描述了 $X, Y$ 之间的线性关系。

# 性质三的推导

- 用随机变量 $\hat{Y} = a + bX$ 近似 $Y$，考虑均方误差

$$\epsilon \triangleq \mathbb{E}\left\{(Y - \hat{Y})^2\right\} = \mathbb{E}\left\{[Y - (a + bX)]^2\right\}$$

- 令

$$\frac{\partial \epsilon}{\partial a} = 2a + 2b\mathbb{E}(X) - 2\mathbb{E}(Y) = 0$$
$$\frac{\partial \epsilon}{\partial b} = 2b\mathbb{E}\left(X^2\right) - 2\mathbb{E}(XY) + 2a\mathbb{E}(X) = 0$$

- 解得 $b_0 = \mathrm{Cov}(X, Y)/D(X), a_0 = \mathbb{E}(Y) - b_0\mathbb{E}(X)$。
- 即有

$$\begin{aligned}\min_{a,b} \epsilon &= \mathbb{E}\left\{[Y - (a_0 + b_0X)]^2\right\} \\ &= \mathbb{E}\left\{[Y - (\mathbb{E}(Y) - b_0\mathbb{E}(X) + b_0X)]^2\right\}\end{aligned}$$

# 性质三的推导

$$
\begin{aligned}
\min_{a,b} \epsilon &= \mathbb{E}\left\{[(Y-\mathbb{E}(Y))-b_0(X-\mathbb{E}(X))]^2\right\} \\
&= \mathbb{E}\left\{(Y-\mathbb{E}(Y))^2\right\}+b_0^2\mathbb{E}\left\{(X-\mathbb{E}(X))^2\right\} \\
&\quad -2b_0\mathbb{E}\left\{(X-\mathbb{E}(X))(Y-\mathbb{E}(Y))\right\} \\
&= D(Y)+b_0^2D(X)-2b_0\mathrm{Cov}(X,Y) \\
&\overset{b_0=\frac{\mathrm{Cov}(X,Y)}{D(X)}}{=} D(Y)+b_0\mathrm{Cov}(X,Y)-2b_0\mathrm{Cov}(X,Y) \\
&= D(Y)-b_0\mathrm{Cov}(X,Y) \\
&= D(Y)-\frac{\mathrm{Cov}(X,Y)^2}{D(X)} \\
&= D(Y)\left(1-\frac{\mathrm{Cov}(X,Y)^2}{D(X)D(Y)}\right)=D(Y)(1-\rho_{XY}^2)
\end{aligned}
$$

# 性质三的推导

- 所以得到 $\min_{a,b} \epsilon = D(Y)(1-\rho_{XY}^2)$。
- $|\rho_{XY}|$ 较小 $\Rightarrow \epsilon$ 较大 $\Rightarrow X, Y$ 之间线性关系较弱；
- $|\rho_{XY}|$ 较大 $\Rightarrow \epsilon$ 较小 $\Rightarrow X, Y$ 之间线性关系较强；
- 当 $|\rho_{XY}| = 1$ 时，误差 $\epsilon = 0$，说明 $X, Y$ 之间是线性关系。


**独立 vs 不相关**

- 独立 $\Rightarrow$ 不相关
- 不相关 $\nRightarrow$ 独立
- 因为相关系数仅仅描述随机变量之间线性关系的强弱。

# 相关系数

**例 (不相关 $\nRightarrow$ 独立)**

设 $X, Y$ 服从单位圆域 $G: x^2 + y^2 \le 1$ 上的均匀分布，讨论 $X, Y$ 的独立性与相关性。

- 前面曾求得 $\mathbb{E}(XY) = \mathbb{E}(X) = \mathbb{E}(Y) = 0$
- 所以 $\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0$，故 $X, Y$ 不相关
- 先前求得 $X, Y$ 的边缘概率密度分别为

$$f_X(x) = \begin{cases} \frac{2}{\pi}\sqrt{1 - x^2}, & |x| < 1 \\ 0, & |x| \ge 1 \end{cases} \quad f_Y(y) = \begin{cases} \frac{2}{\pi}\sqrt{1 - y^2}, & |y| < 1 \\ 0, & |y| \ge 1 \end{cases}$$

- 因为 $f(x, y) = \frac{1}{\pi} \ne f_X(x) f_Y(y), \forall (x, y) \in G$，故 $X, Y$ 不独立

# 相关系数

## 例

设随机变量 $X$ 服从区间 $(-\theta,\theta)$ 上的均匀分布，其中常数 $\theta>0$。若 $Y=\cos X$，求 $\rho_{XY}$。

因为 $X\sim U(-\theta,\theta)$，则 $\mathbb{E}(X)=0$，所以

$$\begin{aligned}\operatorname{Cov}(X,Y)&=\mathbb{E}(XY)-\mathbb{E}(X)\,\mathbb{E}(Y)\\&=\mathbb{E}(X\cos X)\\&=\int_{-\theta}^{\theta}x\cos x\frac{1}{2\theta}\,\mathrm{d}x\\&=0\end{aligned}$$

故有 $\rho_{XY}=0$。

# 目录

# 矩

## 定义 (矩, Moment)

设 $X, Y$ 为随机变量且各阶矩都存在，$k > 0, l > 0$。

1. 称 $\mathbb{E}\left(X^k\right)$ 为 $X$ 的 $k$ 阶原点矩（$k$ 阶矩），记 $\alpha_k$。
2. 称 $\mathbb{E}\left\{[X - \mathbb{E}(X)]^k\right\}$ 为 $X$ 的 $k$ 阶中心矩，记 $\mu_k$。
3. 称 $\mathbb{E}\left(X^k Y^l\right)$ 为 $X, Y$ 的 $k + l$ 阶混合原点矩，记为 $\alpha_{kl}$。
4. 称 $\mathbb{E}\left\{[X - \mathbb{E}(X)]^k [Y - \mathbb{E}(Y)]^l\right\}$ 为 $X, Y$ 的 $k + l$ 阶混合中心矩，记为 $\mu_{kl}$。

- $\mathbb{E}(X)$ 为 1 阶原点矩
- $D(X)$ 为 2 阶中心矩
- $\mathrm{Cov}(X, Y)$ 为 2 阶混合中心矩
- $\rho_{XY} = \dfrac{\mu_{11}}{\sqrt{\mu_{20}}\sqrt{\mu_{02}}}$
- 矩来自物理学中力矩的概念
- 高阶矩存在则低阶矩必存在

# 目录

# 协方差矩阵

## 定义 (协方差矩阵)

设 $n$ 维随机向量 $\mathbf{X}=(X_1,\ldots,X_n)$ 的二阶混合中心矩

$$v_{ij} \triangleq \mathbb{E}\{[X_i-\mathbb{E}(X_i)][X_j-\mathbb{E}(X_j)]\}=\mathrm{Cov}(X_i,X_j),\ i,j=1,\ldots,n$$

都存在，则称矩阵

$$\mathbf{V}=\begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nn} \end{bmatrix}$$

为 $n$ 维随机向量 $\mathbf{X}=(X_1,\ldots,X_n)$ 的协方差矩阵。

# 协方差矩阵的性质

1. $\mathbf{V}^\top = \mathbf{V}$，即协方差矩阵 $\mathbf{V}$ 为对称阵；
2. 协方差矩阵 $\mathbf{V}$ 为非负定阵。

**性质二证明.**

设 $\mathbf{x} = [x_1, \ldots, x_n]^\top \in \mathbb{R}^n$ 为任意 $n$ 维实值非零向量，则

$$\mathbf{x}^\top \mathbf{V} \mathbf{x} = \sum_{i=1}^n \sum_{j=1}^n v_{ij} x_i x_j = \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}\left\{[X_i - \mathbb{E}(X_i)][X_j - \mathbb{E}(X_j)]\right\} x_i x_j$$

$$= \mathbb{E}\left\{\sum_{i=1}^n \sum_{j=1}^n (X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j)) x_i x_j\right\}$$

$$= \mathbb{E}\left\{\left[\sum_{i=1}^n x_i (X_i - \mathbb{E}(X_i))\right]^2\right\} \geq 0$$

# 二维正态随机向量的协方差矩阵

设 $(X,Y)\sim N(\mu_1,\sigma_1^2;\mu_2,\sigma_2^2;\rho)$，联合概率密度为

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

$X$ 和 $Y$ 的协方差为

$$\operatorname{Cov}(X,Y)=\iint(x-\mu_1)(y-\mu_2)f(x,y)\,\mathrm{d}x\,\mathrm{d}y$$
$$\overset{u=\frac{x-\mu_1}{\sigma_1},v=\frac{y-\mu_2}{\sigma_2}}{=}\frac{\sigma_1\sigma_2}{2\pi\sqrt{1-\rho^2}}\iint uv\exp\left\{-\frac{u^2-2\rho uv+v^2}{2(1-\rho^2)}\right\}\mathrm{d}u\,\mathrm{d}v$$

# 二维正态随机向量的协方差矩阵

$$
\begin{aligned}
&=\frac{\sigma_1\sigma_2}{2\pi\sqrt{1-\rho^2}}\int v\exp\left\{-\frac{v^2}{2(1-\rho^2)}\right\}\mathrm{d}v\int u\exp\left\{-\frac{u^2-2\rho uv}{2(1-\rho^2)}\right\}\mathrm{d}u\\
&=\frac{\sigma_1\sigma_2}{2\pi\sqrt{1-\rho^2}}\int v\exp\left\{-\frac{v^2}{2(1-\rho^2)}\right\}\mathrm{d}v\\
&\qquad\int u\exp\left\{-\frac{(u-\rho v)^2-\rho^2v^2}{2(1-\rho^2)}\right\}\mathrm{d}u\\
&\overset{t=\frac{u-\rho v}{\sqrt{1-\rho^2}}}{=}\frac{\sigma_1\sigma_2}{2\pi\sqrt{1-\rho^2}}\int v\exp\left\{-\frac{v^2}{2(1-\rho^2)}\right\}\mathrm{d}v\\
&\qquad\int\sqrt{1-\rho^2}(t\sqrt{1-\rho^2}+\rho v)\exp\left\{-\frac{1}{2}(t^2-\frac{\rho^2v^2}{1-\rho^2})\right\}\mathrm{d}t
\end{aligned}
$$

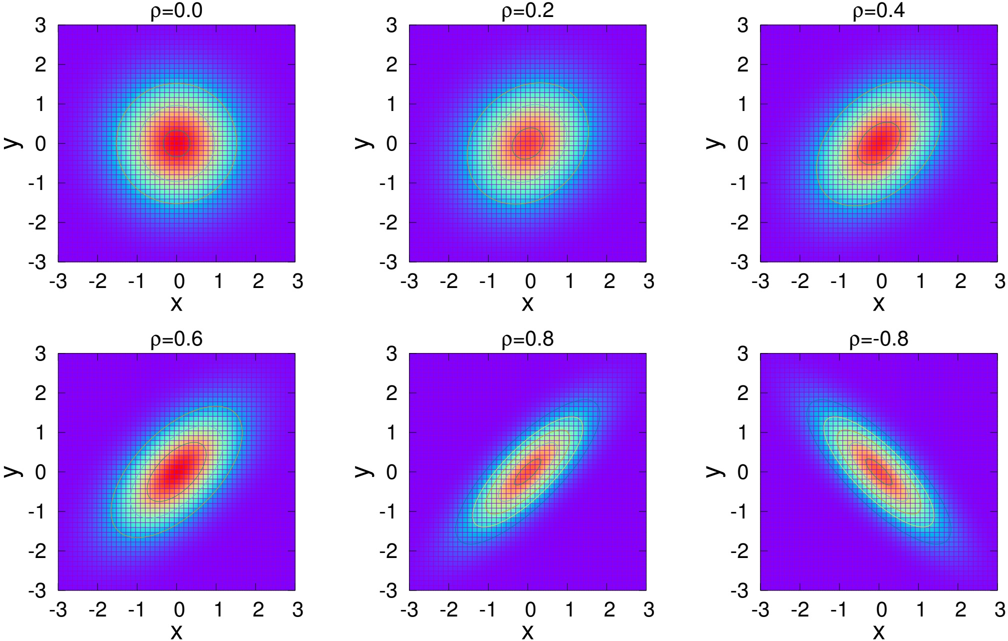
# 二维正态随机向量的协方差矩阵

$$
\begin{aligned}
&= \frac{\rho\sigma_1\sigma_2}{2\pi} \int v^2 e^{-\frac{v^2}{2}} \, \mathrm{d}v \int e^{-\frac{t^2}{2}} \, \mathrm{d}t \\
&= \frac{\rho\sigma_1\sigma_2}{2\pi} \cdot \sqrt{2\pi} \cdot \sqrt{2\pi} \\
&= \rho\sigma_1\sigma_2
\end{aligned}
$$

所以相关系数

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

# 二维正态分布的协方差矩阵



# 二维正态随机向量的协方差矩阵

**定理**

设二维随机变量 $(X, Y) \sim N(\mu_1, \sigma_1^2; \mu_2, \sigma_2^2; \rho)$，则

1. $X, Y$ 的相关系数 $\rho_{XY} = \rho$。
2. $X, Y$ 的协方差矩阵为
$$\mathbf{V} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$
3. $X, Y$ 相互独立 $\Leftrightarrow$ 互不相关。

# 二维正态随机向量概率密度的矩阵表示法

设 $(X, Y) \sim N(\mu_1, \sigma_1^2; \mu_2, \sigma_2^2; \rho)$，联合概率密度为

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

协方差矩阵为

$$\mathbf{V} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

可以看到 $|\mathbf{V}| = \sigma_1^2\sigma_2^2(1-\rho^2)$

# 二维正态随机向量概率密度的矩阵表示法

设 $(X, Y) \sim N(\mu_1, \sigma_1^2; \mu_2, \sigma_2^2; \rho)$，联合概率密度为

$$f(x,y) = \frac{1}{2\pi|\mathbf{V}|^{1/2}} \exp\left\{ \frac{-1}{2(1-\rho^2)} \left[ \frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right] \right\}$$

可以利用伴随矩阵求 $\mathbf{V}$ 的逆阵

$$\mathbf{V}^{-1} = \frac{1}{|\mathbf{V}|}\mathbf{V}^* = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}$$

二次型部分等于

$$\frac{-1}{2|\mathbf{V}|}[x-\mu_1, y-\mu_2] \begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix} \begin{bmatrix} x-\mu_1 \\ y-\mu_2 \end{bmatrix}$$

$$= -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{V}^{-1}(\mathbf{x}-\boldsymbol{\mu})$$

# 正态随机向量联合概率密度的矩阵表示法

**定理（二维正态随机向量联合概率密度的矩阵表示）**

设 $(X, Y) \sim N(\mu_1, \sigma_1^2; \mu_2, \sigma_2^2; \rho)$，则其概率密度为

$$f(\mathbf{x}) = \frac{1}{2\pi|\mathbf{V}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\top}\mathbf{V}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

其中 $\mathbf{x} = [x, y]^{\top}, \boldsymbol{\mu} = [\mu_1, \mu_2]^{\top}$。

**定理（$n$ 维正态随机向量联合概率密度的矩阵表示）**

$n$ 维正态随机变量 $\mathbf{X} = (X_1, \ldots, X_n)$ 的概率密度为

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\mathbf{V}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\top}\mathbf{V}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

其中 $\mathbf{x} \in \mathbb{R}^n, \boldsymbol{\mu} \in \mathbb{R}^n$。

# $n$ 维正态分布的性质

设 $\mathbf{X}=(X_1,\ldots,X_n)\sim N(\boldsymbol{\mu},\mathbf{V})$，则

1. $\mu_i=\mathbb{E}\left(X_i\right), i=1,\ldots,n$
2. $\mathbf{V}=[v_{ij}]_{n\times n}$ 是 $(X_1,\ldots,X_n)$ 的协方差矩阵，且
$$D(X_i)=v_{ii},\quad i=1,\ldots,n$$
$$\mathrm{Cov}(X_i,X_j)=v_{ij},\quad i,j=1,\ldots,n$$
3. $X_i\sim N(\mu_i,v_{ii}), i=1,\ldots,n$
4. $X_1,\ldots,X_n$ 相互独立 $\Leftrightarrow X_1,\ldots,X_n$ 两两不相关 $\Leftrightarrow \mathbf{V}=\mathrm{diag}(v_{11},\ldots,v_{nn})$
5. 若 $X_1,\ldots,X_n$ 相互独立，且 $X_i\sim N(\mu_i,\sigma_i^2)$，则
$$(X_1,\ldots,X_n)\sim N(\boldsymbol{\mu},\mathbf{V})$$
其中 $\boldsymbol{\mu}=(\mu_1,\ldots,\mu_n),\mathbf{V}=\mathrm{diag}(\sigma_1^2,\ldots,\sigma_n^2)$。

# $n$ 维正态分布的性质

6. $(X_1,\ldots,X_n)\sim N(\boldsymbol{\mu},\mathbf{V})\Leftrightarrow X_1,\ldots,X_n$ 的任一非零线性组合 $l_1X_1+\cdots+l_nX_n$ 服从一维正态分布。

7. 正态随机向量的线性变换不变性：若 $\mathbf{X}=(X_1,\ldots,X_n)\sim N(\boldsymbol{\mu},\mathbf{V})$，令

$$
\begin{aligned}
Y_1 &= a_{11}X_1+a_{12}X_2+\cdots+a_{1n}X_n\\
Y_2 &= a_{21}X_1+a_{22}X_2+\cdots+a_{2n}X_n\\
&\vdots\\
Y_m &= a_{m1}X_1+a_{m2}X_2+\cdots+a_{mn}X_n
\end{aligned}
$$

则 $\mathbf{Y}=(Y_1,\ldots,Y_m)$ 仍服从多维正态分布。

# 小结

1. 数学期望
2. 方差
3. 协方差与相关系数、矩