

# 概率统计

依据孔祥仁的概率统计课程整理。 [孔祥仁概率统计课程视频链接](#)

教材：概率论与数理统计 浙大第五版

斜体为个人批注，仅供参考。

## 第一章 概率论的基本概念

### 1. 随机试验

#### 1.1 名词

确定性现象：结果呈现确定性的现象。

随机现象：在个别实验中呈现不确定性，在大量重复实验中表现出统计规律性的现象。

#### 1.2 随机试验

随机试验：对随机现象的实现或者对其的观察，记为  $E$ 。

特点：

1. 相同条件可重复
2. 试验结果明确可知且结果不止一个
3. 试验前不能确定哪个结果会出现

### 2. 样本空间与随机事件

#### 2.1 样本空间

定义：将  $E$  的所有可能结果组成的集合称为  $E$  的样本空间，记为  $S$ 。

样本点：  $S$  的元素。

#### 2.2 随机事件

定义：称  $E$  的样本空间 ( $S$ ) 的子集为  $E$  的随机事件。

事件发生：在一次实验中，该子集的一个样本点出现。

基本事件：由一个样本点组成的单点集。不可以再划分的事件。

必然事件：  $S$  本身。

不可能事件：  $\emptyset$ 。

### 3. 事件间的关系及运算

1.  $A \subset B$ ：包含关系。  $A$  包含于  $B$  ( $B$  包含  $A$ )。  $A$  发生  $\Rightarrow B$  发生。
2. 和事件 (并事件)：  $A$  与  $B$  至少发生一个，记作  $A \cup B$  或  $A + B$ 。
3. 积事件 (交事件)：  $A$  与  $B$  同时发生，记作  $A \cap B$  或  $AB$ 。
4. 差事件：  $A$  发生且  $B$  不发生，记作  $A - B$ 。
5. 互斥事件 (互不相容事件)：  $A$  与  $B$  不能同时发生，  $A \cap B = \emptyset$ 。
6. 逆事件 (对立事件)：  $A$  与  $B$  有且只有一个发生，  $A \cap B = \emptyset$  且  $A \cup B = S$

## 4. 事件的运算律

1. 交换律:  $A \cup B = B \cup A$ ;  $AB = BA$
2. 结合律:  $A \cup (B \cup C) = (A \cup B) \cup C$ ;  
 $A \cap (B \cap C) = (A \cap B) \cap C$ ;  
括号里外开口相同。
3. 分配律:  $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ ;  
 $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ ;  
括号里外开口不同。
4. 德摩根律:  $\overline{A \cup B} = \bar{A} \cap \bar{B}$ ;  
 $\overline{A \cap B} = \bar{A} \cup \bar{B}$ ;  
“长 (短) 杠变短 (长) 杠, 开口换方向”。

## 5. 频率与概率

### 5.1 频率

定义: 事件发生的频数与试验总数之间的比值

基本性质:

1.  $0 \leq f_n(A) \leq 1$
2.  $f_n(S) = 1$
3. 若  $A_1, A_2, \dots, A_k$  为两两不相容事件, 则  
 $f_n(A_1 \cup A_2 \cup \dots \cup A_k) = f_n(A_1) + f_n(A_2) + \dots + f_n(A_k)$

### 5.2 概率

含义: 用于衡量事件A发生的可能性的, 用P来表示

基本性质:

1. 非负性: 任一事件A,  $P(A) \geq 0$
2. 规范性: 必然事件  $S \Rightarrow P(S) = 1$ , 反之不成立
3. 可列可加性: 若  $A_1, A_2, \dots$  为两两不相容事件,  $P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots$

重要性质:

1.  $P(\emptyset) = 0$ : 不可能事件  $\Rightarrow$  概率为0, 反之不成立。
2. 有限可加性: 若  $A_1, A_2, \dots, A_n$  为两两不相容事件, 则  
 $P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_k)$
3. 若  $A \subset B$ , 则  $P(B) \geq P(A)$
4. 任一事件A,  $P(A) \leq 1$
5. 任一事件A,  $P(\bar{A}) = 1 - P(A)$
6. 对于任意两个事件A, B, 有  $P(A \cup B) = P(A) + P(B) - P(AB)$ 。这是一个一定成立的性质。
  - 任意三个事件A, B, C, 有  
 $P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$
  - 任意四个事件A, B, C, D, 有  
 $P(A) + P(B) + P(C) + P(D) - P(AB) - P(AC) - P(AD) - P(BC) - P(BD) - P(CD) +$

$$P(ABC) + P(ABD) + P(ACD) + P(BCD) - P(ABCD)$$

7.

补充一个性质:  $P(A - B) = P(A) - P(AB)$

## 6. 古典概型

### 6.1 特点

1. 有限性:  $S$  包含的样本点是有限个
2. 等可能性: 样本点 (基本事件) 发生的可能性相同

### 6.2 计算方法

事件  $A$  包含了  $k$  个基本事件,  $S$  有  $n$  个样本点。

$$P(A) = \frac{k}{n}$$

## 7. 条件概率

### 7.1 定义

设  $A, B$  为两个事件, 且  $P(A) > 0$ , 称  $P(B|A)$  为条件概率

$$P(B|A) = \frac{P(AB)}{P(A)}$$

### 7.2 特点

1. 非负性: 对于任一事件  $B$ ,  $P(B|A) \geq 0$
2. 规范性: 必然事件  $S \Rightarrow P(S|A) = 1$
3. 可列可加性: 设  $B_1, B_2, B_3, \dots$  是两两互不相容的事件, 那么  

$$P(B_1 \cup B_2 \cup B_3 | A) = P(B_1|A) + P(B_2|A) + P(B_3|A)$$
4. 补充:  $P(B_1 \cup B_2 | A) = P(B_1|A) + P(B_2|A) - P(B_1 B_2 | A)$

## 8. 乘法定理

前提:  $P(A) > 0$

乘法公式:

$$P(AB) = P(A)P(B|A)$$

引申:

$$P(ABC) = P(A)P(B|A)P(C|AB)$$

## 9. 全概率公式

### 9.1 样本空间的划分

$B_1, B_2, \dots, B_n$  是  $E$  的一组事件, 若满足

1.  $B_i B_j = \emptyset, i \neq j, i, j = 1, 2, \dots, n$
2.  $B_1 \cup B_2 \cup \dots \cup B_n = S$

则  $B_1, B_2, \dots, B_n$  为  $S$  的一个划分 (或完备事件组)。

对立事件是一个特殊的划分。

## 9.2 全概率公式

设 $E$ 的样本空间为 $S$ ,  $A$ 是 $E$ 的事件,  $B_1, B_2, \dots, B_n$ 为 $S$ 的划分且 $P(B_i) > 0$ , 则

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

## 10. 贝叶斯公式

设 $E$ 的样本空间为 $S$ ,  $A$ 是 $E$ 的事件,  $B_1, B_2, \dots, B_n$ 为 $S$ 的划分且 $P(A) > 0, P(B_i) > 0$ , 则

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^n P(B_j)P(A|B_j)}$$

这就是贝叶斯公式。

它的推导过程是, 首先写出 $P(B_i|A)$ 的条件概率公式, 对分子用乘法公式改写, 对分母用全概率公式改写。

## 11. 独立性

设试验 $E$ 的事件 $A, B$ , 若 $P(A) > 0$ , 就可以定义 $P(B|A)$ , 一般情况下 $P(B|A) \neq P(B)$ , 即事件 $A$ 发生与否会对 $B$ 发生的概率产生影响; 有的时候 $P(B|A) = P(B)$ , 即事件 $A$ 发生与否不会对 $B$ 发生的概率产生影响

### 11.1 定义

设 $A, B$ 为两个事件, 若满足 $P(AB) = P(A)P(B)$ , 那么我们称 $A, B$ 相互独立。

设 $A, B, C$ 三个事件, 当且仅当满足以下四个条件 $A, B, C$ 相互独立:

$$\left\{ \begin{array}{l} P(AB) = P(A)P(B) \Rightarrow A, B \text{ 相互独立} \\ P(AC) = P(A)P(C) \Rightarrow A, C \text{ 相互独立} \\ P(BC) = P(B)P(C) \Rightarrow B, C \text{ 相互独立} \\ P(ABC) = P(A)P(B)P(C) \end{array} \right\} \left. \begin{array}{l} A, B, C \\ \text{两两独立} \end{array} \right\} A, B, C \text{ 相互独立}$$

## 第二章 随机变量及其分布

### 1. 随机变量

#### 1.1 定义

随机试验 $E$ 的样本空间 $S = \{e\}$ ,  $X = X(e)$ 是定义在 $S$ 上的实值单值函数, 则称 $X = X(e)$ 是随机变量。

#### 1.2 注意

1. 随机变量用大写字母表示
2. 实数用小写字母表示
3. 某些试验结果本身就是一个数, 可以将实验结果本身作为随机变量。

## 2. 离散型随机变量及分布

### 2.1 离散型随机变量

取值是有限多个或可列无穷多个的随机变量。

### 2.2 分布律的性质

离散型随机变量的分布律： $X$ 的所有取值； $X$ 每个取值各自概率；写成图表或统一的表达式

1.  $p_k \geq 0$
2.  $\sum_{k=1}^{\infty} p_k = 1$

分布律满足上面2个性质；满足上面2个性质就是某个分布律。

### 2.3 重要分布

#### 2.3.1 0 - 1分布（两点分布）

设 $X$ 只可能取0, 1两个值，它的分布律满足

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

称 $X$ 服从以 $p$ 为参数的0 - 1分布， $p$ 代表 $X = 1$ 发生的概率。

#### 2.3.2 伯努利试验及二项分布

如果 $E$ 只有两个结果： $A$ 和 $\bar{A}$ ，则称为伯努利试验。

若将该试验**重复独立**地进行 $n$ 次，则称为 $n$ 重伯努利试验。

- 重复： $P(A)$ 不变， $P(\bar{A})$ 也不变
- 独立：每次试验互不影响

设 $X$ 表示 $n$ 重伯努利试验中 $A$ 发生的次数， $X = 0, 1, 2, \dots, n$ 。假定 $P(A) = p$ ， $P(\bar{A}) = 1 - p = q$ ，

$$P(X = k) = C_n^k p^k q^{n-k}$$

称 $X$ 服从参数为 $n, p$ 的二项分布，记作 $X \sim b(n, p)$ ，其中 $k = 0, 1, 2, \dots, n$

之所以叫二项分布是因为它和二项式定理有关。

0 - 1分布是特殊的二项分布，即 $n = 1$ 时的二项分布。

#### 2.3.3 泊松分布

泊松分布：设随机变量 $X = 0, 1, 2, \dots$ ，每个取值的概率满足

$$P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda}$$

其中 $k = 0, 1, 2, \dots$ ，要求 $\lambda$ 为大于0的常数，可以称 $X$ 服从一个参数为 $\lambda$ 的泊松分布，记作 $X \sim \Pi(\lambda)$ 。

泊松定理：设 $\lambda$ 为大于0的常数， $n$ 为任意正整数。又设 $np = \lambda$ ，则对任一固定的非负整数 $k$ ，有下式成立

$$\lim_{n \rightarrow \infty} C_n^k p^k q^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}$$

即，将伯努利试验做很多次时，二项分布的分布律 = 泊松分布的分布律。

一般来说，当 $n \geq 20, p \leq 0.05$ 时，可以利用泊松分布近似二项分布。

### 2.3.4 几何分布与超几何分布

几何分布的数学模型：伯努利试验“达到目的”的概率为 $p$ ，试验第 $k$ 次才成功的概率为

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

几何分布的定义： $X$ 的分布律满足上述等式，就是几何分布，记作 $X \sim G(p)$ 。

超几何分布的数学模型：从有限 $N$ 个物品（其中有 $D$ 个特殊物品）中抽出 $n$ 个物品，包含了特定物品 $k$ 个（ $k \leq \min\{D, n\}$ ）的概率为

$$P\{X = k\} = \frac{C_D^k C_{N-D}^{n-k}}{C_N^n}$$

超几何分布的定义： $X$ 的分布律满足上述等式，就是超几何分布，记作 $X \sim H(N, D, n)$ 。

## 3. 随机变量的分布函数

### 3.1 定义

设 $X$ 为随机变量， $x$ 是任意实数， $F(x) = P\{X \leq x\}$ ， $-\infty < x < +\infty$ ，为 $X$ 的分布函数。

另外： $P\{x_1 < X \leq x_2\} = P\{X \leq x_2\} - P\{X \leq x_1\} = F(x_2) - F(x_1)$ 。

分布函数的性质：

1.  $F(x)$ 为不减函数
2.  $0 \leq F(x) \leq 1$ ,  $\lim_{x \rightarrow -\infty} F(x) = 0$ ,  $\lim_{x \rightarrow \infty} F(x) = 1$
3.  $F(x)$ 是一个右连续函数， $F(x+0) = F(x)$ ， $+0$ 表示往右一点点。

## 4. 连续型随机变量及概率密度

### 4.1 定义

随机变量 $X$ 的分布函数为 $F(x)$ ， $F(x)$ 由一个非负的可积的函数 $f(x)$ 积分得来，

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

称 $X$ 是连续性随机变量， $f(x)$ 是 $X$ 的概率密度函数。

注意：连续型随机变量 $X$ 的分布函数是连续函数， $f(x)$ 不一定连续。

可以认为概率密度对应离散型随机变量的分布律。

### 4.2 概率密度的性质

1.  $f(x) \geq 0$ .
2.  $\int_{-\infty}^{+\infty} f(x) = 1$ .

满足上述两点，可说明 $f(x)$ 是某一个随机变量 $X$ 的概率密度函数。

3. 任意实数 $x_1, x_2$  ( $x_1 \leq x_2$ ) ,  $P(x_1 < X \leq x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)dx$ .
4. 如果 $f(x)$ 在 $x = x_0$ 处连续，那么 $F'(x_0) = f(x_0)$ .
5.  $P\{X = a\} = 0$ .
6. 对于连续型随机变量，  
 $P\{x_1 \leq X \leq x_2\} = F(x_2) - F(x_1) + P\{X = x_1\} = P(x_1 < X \leq x_2)$ ,  
 $P\{x_1 < X < x_2\}$ 和 $P\{x_1 \leq X < x_2\}$ 同理。即，端点值不影响区域上的概率。

### 4.3 概率分布

随机变量的概率分布包括分布函数和分布律/概率密度。求解概率分布时，如果题目给出了其中一方，求另一方即可；如果都没有，就要求双方。

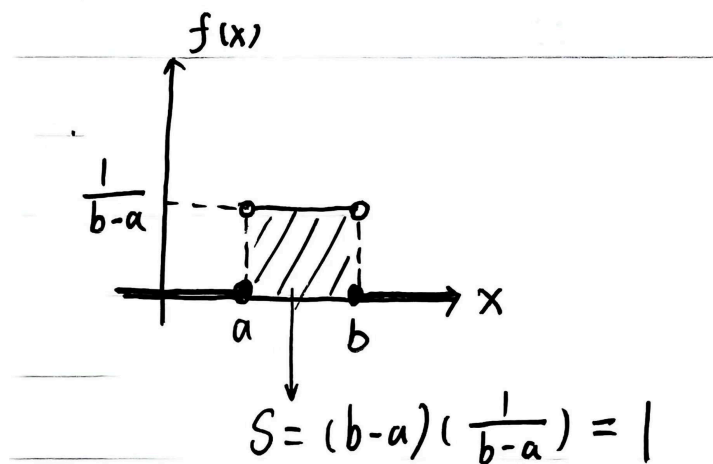
### 4.4 重要的连续性随机变量

#### 4.4.1 服从均匀分布的随机变量

$X$ 的概率密度为

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

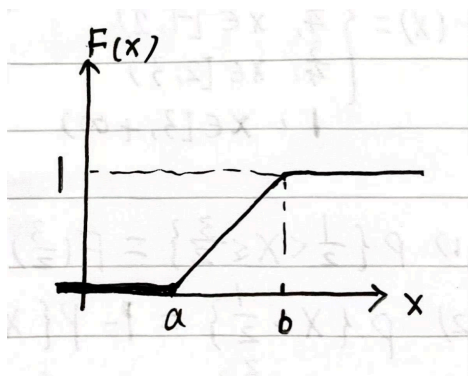
称 $X$ 在区间 $(a, b)$ 服从均匀分布，记作 $X \sim u(a, b)$ 。



分布函数为

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

$\frac{x-a}{b-a}$ 可由矩形图像面积直接得出。



#### 4.4.2 服从指数分布的随机变量

$X$ 的概率密度为

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

$\lambda$ 为大于零的常数，称 $X$ 服从指数分布，记作 $X \sim E(\lambda)$ 。

注意： $\lambda$ 与泊松分布相区分。

分布函数为

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0 \\ 0, & \text{others} \end{cases}$$

无记忆性：对于任意两个数  $s > 0, t > 0$ ,

$$P\{X > t\} = P\{X > s + t | X > s\}$$

无记忆性的例子：一个没有明显衰老的人的寿命  $X$ .  $P\{X > 10\} = P\{X > 5 + 10 | X > 5\}$ , 一个刚出生的人活十年的概率和一个五岁的人再活十年的概率相同。若不然,  $P\{X > 10\} \neq P\{X > 80 + 10 | X > 80\}$ .

#### 4.4.3 服从正态分布的随机变量

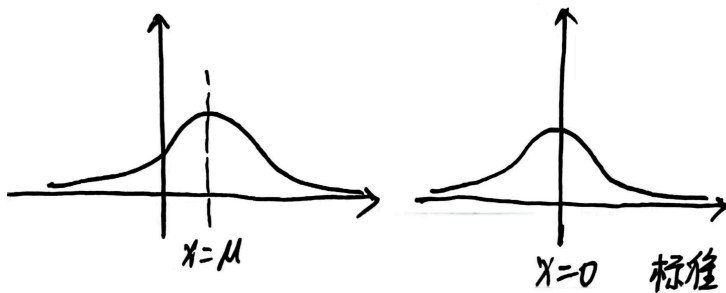
$X$  的概率密度为

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

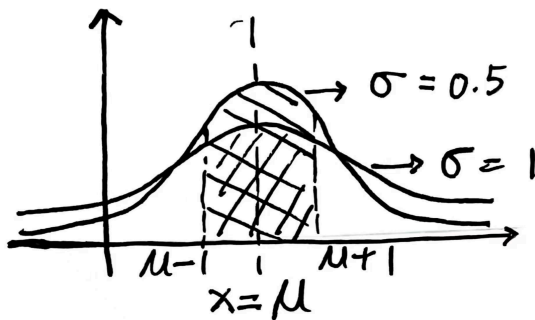
其中  $\mu, \sigma$  为常数, 则称  $X$  服从参数为  $\mu, \sigma$  的正态分布, 记作  $X \sim N(\mu, \sigma^2)$ .

当  $\mu = 0, \sigma = 1$  时, 称为标准正态分布, 概率密度函数写作  $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ .

$\mu$  决定概率密度函数图像的位置,  $\mu$  又称为位置参数。



$\sigma$  决定正态分布概率密度的峰值,  $\sigma$  越小, 峰值越大。 $\sigma$  越小,  $X$  落在关于对称轴对称的区间中的概率越大。



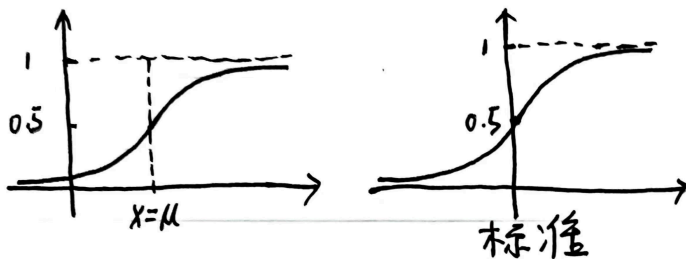
分布函数为

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

当  $\mu = 0, \sigma = 1$  时, 标准正态分布函数为



$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$



正态分布的性质:

1. 概率密度函数图像关于  $x = \mu$  对称
2.  $f(x)_{\max} = f(\mu) = \frac{1}{\sqrt{2\pi}\sigma}$ .  
对于等长区间, 越靠近  $x = \mu$ ,  $X$  落到该区间的概率越大。  
 $P\{\mu - h < X \leq \mu\} = P\{\mu < X \leq \mu + h\}$ .
3. 对于标准正态分布,  $\Phi(x) + \Phi(-x) = 1$
4. 引理:  $X \sim N(\mu, \sigma^2)$ , 令  $Z = \frac{x-\mu}{\sigma}$ ,  $Z \sim N(0, 1)$ .  
这个引理的作用是将一般正态分布转化为标准正态分布, 从而能够通过查表求值。

3σ法则:

- $X \sim N(\mu, \sigma^2)$ ,  $F(x) = P\{X \leq x\} = P\{\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\} = \Phi(\frac{x-\mu}{\sigma})$
- 对于  $(x_1, x_2]$ ,

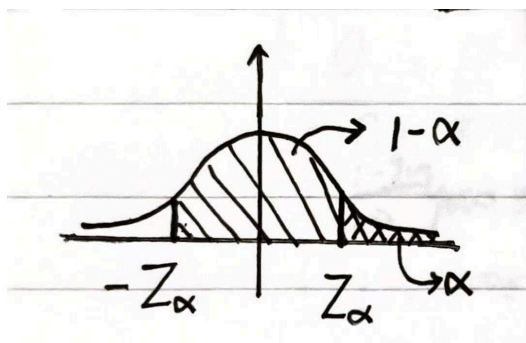
$$\begin{aligned} P\{x_1 < X \leq x_2\} &= P\{\frac{x_1 - \mu}{\sigma} < \frac{X - \mu}{\sigma} \leq \frac{x_2 - \mu}{\sigma}\} \\ &= P\{\frac{X - \mu}{\sigma} \leq \frac{x_2 - \mu}{\sigma}\} - P\{\frac{X - \mu}{\sigma} \leq \frac{x_1 - \mu}{\sigma}\} \\ &= \Phi(\frac{x_2 - \mu}{\sigma}) - \Phi(\frac{x_1 - \mu}{\sigma}) \end{aligned}$$

- $P\{\mu - \sigma < X \leq \mu + \sigma\} = \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 = 68.26\%$   
 $P\{\mu - 2\sigma < X \leq \mu + 2\sigma\} = 2\Phi(2) - 1 = 95.44\%$   
 $P\{\mu - 3\sigma < X \leq \mu + 3\sigma\} = 2\Phi(3) - 1 = 99.74\%$

上 $\alpha$ 分位点:  $X \sim N(0, 1)$ ,  $Z_\alpha$ 满足  $P\{X > Z_\alpha\} = \alpha$ ,  $0 < \alpha < 1$ ,  $Z_\alpha$ 是上 $\alpha$ 分位点。

上 $\alpha$ 分位点只针对标准正态分布。

性质:  $-Z_\alpha = Z_{1-\alpha}$



## 5. 随机变量的函数的分布

定理：设随机变量 $X$ 具有的概率密度为 $f_X(x)$ ,  $-\infty < x < +\infty$ , 又设函数 $g(x)$ 处处可导且恒有 $g'(x) > 0$  (或恒有 $g'(x) < 0$ ) , 则 $Y = g(x)$ 是连续性随机变量, 其概率密度为

$$f_Y(y) = \begin{cases} f_X[h(y)]|h'(y)|, & \alpha < y < \beta \\ 0, & \text{others} \end{cases}$$

其中 $\alpha = \min\{g(-\infty), g(\infty)\}$ ,  $\beta = \max\{g(-\infty), g(\infty)\}$ ,  $h(y)$ 是 $g(x)$ 的反函数。

注意：若 $f_X(x)$ 在 $[a, b]$ 以外等于0, 只需要要求 $g(x)$ 在 $[a, b]$ 恒有 $g'(x) > 0$ 或 $< 0$ ,  $\alpha = \min\{g(a), g(b)\}$ ,  $\beta = \max\{g(a), g(b)\}$ .

## 第三章 多维随机变量及分布

### 1. 二维随机变量

#### 1.1 二维随机变量及分布

二维随机变量： $E$ 的 $S = \{e\}$ ,  $e$ 由 $(X, Y)$ 构成,  $X, Y$ 是定义在 $S$ 里的, 称 $(X, Y)$ 为二维随机变量。

联合分布函数： $F(x, y) = P\{X \leq x, Y \leq y\}$ 称为 $(X, Y)$ 的联合分布函数。

性质：

1.  $F(x, y)$ 为不减函数

2.  $0 \leq F(x, y) \leq 1$

特殊地：

◦  $F(-\infty, y) = P\{X \leq -\infty, Y \leq y\} = 0$

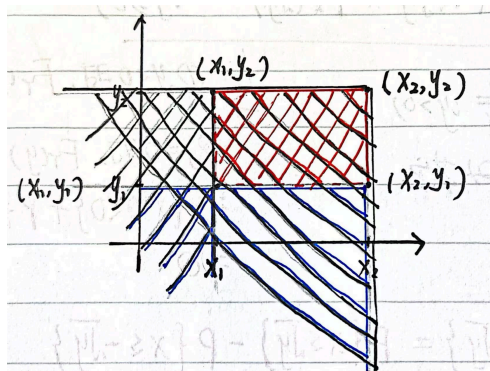
◦  $F(x, -\infty) = 0$

◦  $F(-\infty, -\infty) = 0$

◦  $F(+\infty, +\infty) = 1$

3.  $F(x, y)$ 是关于 $x, y$ 的右连续的函数

$P\{x_1 < X \leq x_2, y_1 < Y \leq y_2\} = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1)$ . 这是恒成立的等式, 不论是对于连续型还是离散型。



#### 1.2 离散型二维随机变量及联合分布律

定义： $(X, Y)$ 所有取值为有限对或可列无限对, 称为离散型二维随机变量。 $P\{X = x_i, Y = y_j\} = p_{ij}$ ,  $i, j = 1, 2, 3, \dots$

性质：

1.  $p_{ij} \geq 0$ , 非负性

$$2. \sum p_{ij} = 1$$

### 1.3 连续型二维随机变量及联合概率密度

定义：若  $(X, Y)$  的分布函数为  $F(x, y)$ ,  $f(x, y)$  为非负可积的,  $\int_{-\infty}^y \int_{-\infty}^x f(u, v) du dv = F(x, y)$ , 称  $(X, Y)$  为连续型二维随机变量, 称  $f(x, y)$  为  $(X, Y)$  的联合概率密度。

性质:

1.  $f(x, y) \geq 0$
2.  $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1$
3. 设  $G$  是平面  $oxy$  上的某个区域,  $(X, Y)$  落在  $G$  区域的概率  $P\{(X, Y) \in G\} = \iint_G f(x, y) dx dy$
4.  $f(x, y)$  在  $(x, y)$  点上连续,  $\frac{\partial^2 F(x, y)}{\partial x \partial y} = \frac{\partial^2 F(x, y)}{\partial y \partial x} = f(x, y)$

## 2. 边缘分布

### 2.1 边缘分布函数

二维随机变量  $(X, Y)$  具有分布函数  $F(x, y)$ , 而  $X$  和  $Y$  都是随机变量, 各自也有分布函数, 将它们分别记为  $F_X(x)$ ,  $F_Y(y)$ , 称为二维随机变量关于  $X$  和关于  $Y$  的边缘分布函数。

$$F_X(x) = P\{X \leq x\} = P\{X \leq x, Y \leq +\infty\} = F(x, +\infty)$$

$$\text{同理, } F_Y(y) = F(+\infty, y)$$

### 2.2 边缘概率密度

已知  $(X, Y)$  的联合概率密度  $f(x, y)$ , 求关于  $X$  和  $Y$  的边缘概率密度:

$$F_X(x) = F(x, +\infty) = \int_{-\infty}^x \int_{-\infty}^{+\infty} f(u, y) dy du$$

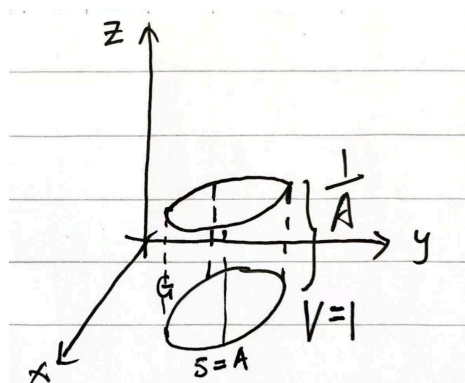
$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy$$

$$\text{同理, } f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx$$

### 2.3 常用二维分布

#### 1. 二维均匀分布

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



#### 2. 二维正态分布

注意：联合概率密度可以推出边缘概率密度；但是两个边缘概率密度无法推出联合概率密度。

## 3 条件分布

### 3.1 离散型二维随机变量的条件分布律

定义：\$(X, Y)\$是二维离散型随机变量，

$$P\{X = x|Y = y\} = \frac{P\{X = x, Y = y\}}{P\{Y = y\}}$$

称为在\$Y = y\$的条件下，\$X\$的条件分布律。

### 3.2 连续型二维随机变量的条件概率密度

定义：\$(X, Y)\$的概率密度为\$f(x, y)\$，\$(X, Y)\$关于\$Y\$的边缘概率密度为\$f\_Y(y)\$，在\$Y = y\$的条件下\$X\$的条件概率密度为

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}$$

\$X\$的条件概率分布函数为

$$F_{X|Y}(x|y) = \int_{-\infty}^x \frac{f(u, y)}{f_Y(y)} du$$

## 4. 相互独立的随机变量

定义：\$P\{X \leq x, Y \leq y\} = P\{X \leq x\}P\{Y \leq y\}\$，称随机变量\$X, Y\$相互独立。

\$F(x, y) = F\_X(x)F\_Y(y)\$，也相互独立。

如果是连续型随机变量：\$f(x, y) = f\_X(x)f\_Y(y)\$几乎处处成立，是\$X, Y\$相互独立的充要条件。用“几乎”限定，是因为偶尔在面积为0的点上不满足上述等式。但是这不影响相互独立的判定。

如果是离散型随机变量：对于样本空间内任一\$i, j\$，有\$P\_{ij} = P\_iP\_j\$，是\$X, Y\$相互独立的充要条件。

二维正态随机变量\$(X, Y)\$，若\$X, Y\$相互独立，则参数\$(\mu\_1, \mu\_2, \sigma\_1^2, \sigma\_2^2, \rho)\$中\$\rho = 0\$。

## 5. 两个随机变量的函数的分布

### 5.1 \$Z = X + Y\$

\$Z\$的概率密度函数为

$$f_Z(z) = \int_{-\infty}^{+\infty} f(z - y, y) dy = \int_{-\infty}^{+\infty} f(x, z - x) dx$$

注意：当\$X\$和\$Y\$相互独立时，\$f(z - y, y)\$可以拆成\$f\_X(z - y)f\_Y(y)\$，\$f(x, z - x)\$同理。

卷积公式：

$$f_X f_Y = \int_{-\infty}^{+\infty} f_X(z - y) f_Y(y) dy = \int_{-\infty}^{+\infty} f_X(x) f_Y(z - x) dx$$

结论：\$X, Y\$独立，\$X \sim N(\mu\_1, \sigma\_1^2), Y \sim N(\mu\_2, \sigma\_2^2)\$，\$X + Y = Z \sim N(\mu\_1 + \mu\_2, \sigma\_1^2 + \sigma\_2^2)\$；如果有三个这样的变量相加，\$Z \sim N(\mu\_1 + \mu\_2 + \mu\_3, \sigma\_1^2 + \sigma\_2^2 + \sigma\_3^2)\$。

有限个相互独立并服从正态分布的随机变量的线性组合也服从正态分布。

拓展：\$Z\$是\$X, Y\$的线性组合，\$Z = aX + bY\$，求\$f\_Z(z)\$

- 方法一：分布函数法。\$F\_Z(z) = P\{Z \leq z\} = P\{aX + bY \leq z\} = \iint\_{ax+by \leq z} f(x, y) dx dy\$。

分布函数法是求两个随机变量的函数的概率密度的一般方法，通用。

- 方法二：公式法。  $f_Z(z) = \int_{-\infty}^{+\infty} \frac{1}{|b|} f(x, \frac{z-ax}{b}) dx$  或  $f_Z(z) = \int_{-\infty}^{+\infty} \frac{1}{|a|} f(\frac{z-by}{a}, y) dy$

如果  $X, Y$  相互独立,  $f(\frac{z-by}{a}, y)$  可以拆成  $f_X(x)f_Y(\frac{z-ax}{b})$ ,  $f(\frac{z-by}{a}, y)$  同理。

## 5.2 Z = XY

$Z$  的概率密度函数为

$$f_Z(z) = \int_{-\infty}^{+\infty} \frac{1}{|y|} f(\frac{z}{y}, y) dy = \int_{-\infty}^{+\infty} \frac{1}{|x|} f(x, \frac{z}{x}) dx$$

当  $X$  和  $Y$  相互独立时, 可以将联合概率密度拆开, 同前, 之后不再特别说明。

## 5.3 Z = X/Y

$Z$  的概率密度函数为

$$f_Z(z) = \int_{-\infty}^{+\infty} |y| f(yz, y) dy$$

$Z = X/Y$  同理。

## 5.4 Z = max{X, Y}

已知  $X, Y$  独立, 分布函数为  $F_X(x), F_Y(y)$ 。

$Z = \max\{X, Y\}$ :

$$F_Z(z) = P\{Z \leq z\} = P\{X \leq Z, Y \leq Z\} = P\{X \leq z\}P\{Y \leq z\} = F_X(z)F_Y(z)$$

$Z = \max\{X_1, \dots, X_n\}$ :

$$F_Z(z) = F_{X_1}(z) \dots F_{X_n}(z)$$

若  $X_1, \dots, X_n$  独立同分布:

$$F_Z(z) = [F_X(z)]^n$$

$Z = \min\{X, Y\}$ :

$$F_Z(z) = 1 - P\{Z \geq z\} = 1 - P\{X \geq z\}P\{Y \geq z\} = 1 - [1 - F_X(z)][1 - F_Y(z)]$$

$Z = \min\{X_1, \dots, X_n\}$ :

$$F_Z(z) = 1 - [1 - F_{X_1}(z)] \dots [1 - F_{X_n}(z)]$$

若  $X_1, \dots, X_n$  独立同分布:

$$F_Z(z) = 1 - [1 - F_X(z)]^n$$

# 第四章 随机变量的数字特征

## 1. 数学期望

### 1.1 离散型

$$E(X) = \sum_{i=1}^{+\infty} p_i x_i$$

要求绝对收敛。

### 1.2 连续型

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

要求绝对收敛。

1.3 随机变量的函数的数学期望

$Y$ 是 $X$ 的函数,  $Y = g(X)$  ( $g$ 是连续函数) , 若 $X$ 为离散型,  $P\{X = x_k\} = p_k$ , 则  
 $E(Y) = E[g(x)] = \sum_{k=1}^{+\infty} g(x_k)p_k$ ; 若 $X$ 为连续型,  $E(Y) = E[g(x)] = \int_{-\infty}^{+\infty} g(x)f(x)dx$ .

要求绝对收敛。

$Z$ 是 $(X, Y)$ 的函数,  $Z = g(X, Y)$  ( $g$ 是连续函数) , 若 $X$ 为离散型,  $P\{X = x_i, Y = y_j\} = p_{ij}$ , 则  
 $E(Z) = \sum_{j=1}^{+\infty} \sum_{i=1}^{+\infty} p_{ij}g(x_i, y_i)$ ; 若 $X$ 为连续型,  $E(Z) = \iint f(x, y)g(x, y)dxdy$

要求绝对收敛。

1.4 数学期望的性质

- 1.  $E(C) = C$ .
- 2.  $X$ 是随机变量,  $E(CX) = CE(X)$ .
- 3.  $E(X + Y) = E(X) + E(Y)$ .
- 4.  $X, Y$ 是相互独立的随机变量,  $E(XY) = E(X)E(Y)$ .

1.5 特定分布的数学期望

分布类型	符号表示	期望
0-1分布	$B(1, p)$	$p$
二项分布	$B(n, p)$	$np$
泊松分布	$\Pi(\lambda)$	$\lambda$
几何分布	$G(p)$	$\frac{1}{p}$
均匀分布	$U(a, b)$	$\frac{a+b}{2}$
指数分布	$E(\lambda)$	$\frac{1}{\lambda}$
正态分布	$N(\mu, \sigma^2)$	$\mu$

2. 方差

2.1 定义

方差衡量随机变量和自身均值的偏离程度。

$X$ 是一个随机变量,  $[X - E(X)]^2$ 的期望存在, 称它的期望为 $X$ 的方差, 记作  
 $D(X) = E\{[X - E(X)]^2\}$

方差的本质是随机变量 $X$ 的函数的期望,  $E(X)$ 是一个常数。

标准差/均方差:  $\sigma(X) = \sqrt{D(X)}$

2.2 计算方法

- 1.  $X$ 为离散型:  $D(X) = \sum_{k=1}^{\infty} [x_k - E(X)]^2 p_k, p_k = P\{X = k\}$
- 2.  $X$ 为连续型:  $D(X) = \int_{-\infty}^{+\infty} [x - E(X)]^2 f(x)dx$
- 3. 无论离散型还是连续型:  $D(X) = E(X^2) - [E(X)]^2$

## 2.3 特殊分布的方差

分布类型	符号表示	方差
0-1分布	$B(1, p)$	$p(1 - p)$
二项分布	$B(n, p)$	$np(1 - p)$
泊松分布	$\Pi(\lambda)$	$\lambda$
几何分布	$G(p)$	$\frac{p^2}{1-p}$
均匀分布	$U(a, b)$	$\frac{(b-a)^2}{12}$
指数分布	$E(\lambda)$	$\frac{1}{\lambda^2}$
正态分布	$N(\mu, \sigma^2)$	$\sigma^2$

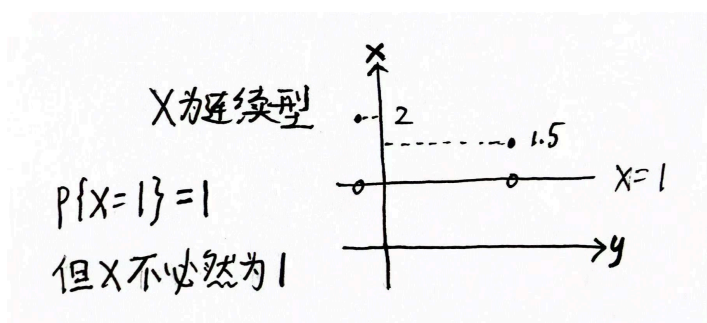
## 2.4 方差的性质

- $D(C) = 0$ .
- $D(CX) = C^2 D(X)$ ,  $D(C + X) = D(X)$ .
- $D(X + Y) = D(X) + D(Y) + 2E\{[X - E(X)][Y - E(Y)]\}$ .

特别地, 当 $X, Y$ 相互独立时,  $D(X + Y) = D(X) + D(Y)$ .

- $D(X) = 0$ 的充要条件为:  $X$ 以概率为1取常数 $E(X)$ .

“概率为1”不等同于“必然”



结论:  $X \sim N(\mu_1, \sigma_1^2)$ ,  $Y \sim N(\mu_2, \sigma_2^2)$ ,  $Z = CX + DY$ ,  $E(Z) = C\mu_1 + D\mu_2$ ,  $D(Z) = C^2\sigma_1^2 + D^2\sigma_2^2$ .

## 2.5 切比雪夫不等式

$E(X) = \mu$ ,  $D(X) = \sigma^2$ , 对于一个大于0的常数 $\epsilon$ , 都有以下不等式成立:  $P\{|X - \mu| \geq \epsilon\} \leq \frac{\sigma^2}{\epsilon^2}$  或  $P\{|X - \mu| \leq \epsilon\} \geq 1 - \frac{\sigma^2}{\epsilon^2}$

## 3. 协方差与相关系数

$E\{[X - E(X)][Y - E(Y)]\} \neq 0$ 时,  $X, Y$ 不相互独立, 即 $X, Y$ 存在一定关系。

- 线性关系:  $Y = aX + b$ ,  $a \neq 0$ .

$X, Y$ 线性相关, 简称相关。

- 非线性关系:  $Y = cX^2 + dX + e$ ,  $c \neq 0$ . 不一定是2次方, 也可以是3次方, 4次方.....

### 3.1 定义

$E\{[X - E(X)][Y - E(Y)]\} \neq 0$  叫做  $X$  与  $Y$  的协方差, 记作  $\text{cov}(X, Y)$ .

$\rho_{XY} = \frac{\text{cov}(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$  称为  $X, Y$  的相关系数, 描述  $X, Y$  的线性相关程度。

### 3.2 性质

1.  $\text{cov}(X, Y) = \text{cov}(Y, X)$ .

2.  $\text{cov}(X, X) = D(X)$ .

3.  $D(X + Y) = D(X) + D(Y) + 2\text{cov}(X, Y)$ .

4.  $\text{cov}(X, Y) = E(XY) - E(X)E(Y)$ .

$X, Y$  相互独立,  $E(XY) = E(X)E(Y)$ ,  $\text{cov}(X, Y) = 0$ .

5.  $\text{cov}(aX, bY) = ab\text{cov}(X, Y)$ .

6.  $\text{cov}(X_1 + X_2, Y) = \text{cov}(X_1, Y) + \text{cov}(X_2, Y)$ .

7.  $|\rho_{XY}| \leq 1$

8.  $|\rho_{XY}| = 1$  的充要条件: 存在常数  $a, b$ , 使得  $P\{aX + b\} = 1$ .

9. 若  $X, Y$  存在线性关系, 即  $Y = aX + b, a \neq 0$ : 当  $a > 0$  时,  $\rho_{XY} = 1$ ; 当  $a < 0$  时,  $\rho_{XY} = -1$ .

$|\rho_{XY}|$  越靠近 1, 线性相关性越大;  $|\rho_{XY}|$  越靠近 0, 线性相关性越小。

当  $\rho_{XY} = 0$  时,  $X, Y$  不相关。

辨析: 一般情况下, 独立  $\Rightarrow$  不相关; 不相关  $\nRightarrow$  独立。

• 特例:  $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ .

$\rho = 0 \Leftrightarrow X, Y$  不相关  $\Leftrightarrow X, Y$  独立。

## 第五章

大数定律略。

### 中心极限定理

#### 独立同分布中心极限定理

设  $\{X_i\}$  独立同分布,  $E(X_i) = \mu, D(X_i) = \sigma^2$ , 对于任意实数  $x$ , 都有等式成立:

$$\lim_{n \rightarrow \infty} P\left\{\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \leq x\right\} = \Phi(x)$$

#### 李雅普诺夫定理

设  $\{X_i\}$  独立同分布,  $E(X_i) = \mu, D(X_i) = \sigma^2$ , 对于任意实数  $x$ , 假设每个  $X_i$  都对  $\sum_{i=1}^n X_i$  影响不大, 记  $(\sum_{i=1}^n \sigma^2)^{\frac{1}{2}} = S_n$ , 则

$$\lim_{n \rightarrow \infty} P\left\{\frac{1}{S_n} \sum_{i=1}^n (X_i - \mu_i) \leq x\right\} = \Phi(x)$$

#### 棣莫弗-拉普拉斯中心极限定理

$Y_n$  服从参数  $n, p$  的二项分布, 对于任何实数  $x$ , 则

$$\lim_{n \rightarrow \infty} P\left\{\frac{Y_n - np}{\sqrt{np(1-p)}} \leq x\right\} = \Phi(x)$$



标准正态分布函数是二项分布的极限分布函数。

## 第六章 样本及抽样分布

### 1. 随机样本

总体：试验中全部的可能观测到的值。

个体：每一个可能被观测到的值。

有限总体：总体中的个体为有限的时候。

无限总体：总体中的个体为无限的时候。

总体 $X$ 或（总体的某一数量指标 $X$ ）服从的分布不知道，或者只知道服从的分布的形式，但不知道具体参数。就需要抽取部分个体来判断总体。

对总体 $X$ 进行一次随机观察并记录结果，即从总体中抽取了一个样本。在相同条件下，进行 $n$ 次独立重复抽取，产生了 $X_1, X_2, \dots, X_n$ 这些随机变量， $\{X_n\}$ 叫做样本。 $\{X_n\}$ 独立同分布，且与 $X$ 服从相同分布。 $x_1, x_2, \dots, x_n$ 叫做样本值/观测值。

有限总体使用放回抽样（若总体个数远大于抽取个数，可以不放回）；无限总体使用不放回抽样。

### 2. 抽样分布

#### 2.1 统计量

定义：设 $X_1, X_2, \dots, X_n$ 是来自总体 $X$ 的一个样本， $g(X_1, X_2, \dots, X_n)$ 是 $X_1, X_2, \dots, X_n$ 的函数，若 $g$ 不包含未知参数，则称 $g(X_1, X_2, \dots, X_n)$ 为一个统计量，称 $g(x_1, x_2, \dots, x_n)$ 为统计量的观测值。

典型的统计量：

1. 样本平均值。 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ .
2. 样本方差。 $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} (\sum_{i=1}^n X_i^2 - n\bar{X}^2)$ . 这里除以 $n-1$ 是为了满足样本的无偏性。
3. 样本标准差。
4. 样本 $k$ 阶原点矩。 $A_k = \frac{1}{n} \sum_{i=1}^n X_i^k$ .
5. 样本 $k$ 阶中心矩。 $A_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$ .
6. 经验分布函数（样本分布函数）。根据样本得到的分布函数。

定义：从总体 $X$ 中抽取一个容量为 $n$ 的样本 $X_1, X_2, \dots, X_n$ ，观测值为 $x_1, x_2, \dots, x_n$ ，升序排列为 $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ ，则经验分布函数可以写成：

$$F_n(x) = \begin{cases} 0, & x < x_{(1)} \\ \frac{k}{n}, & x_{(k)} \leq x < x_{(k+1)} \\ 1, & x \geq x_{(n)} \end{cases}$$

格列汶科定理： $n \rightarrow \infty$ ， $F_n(x)$ 依概率收敛于总体分布函数 $F(x)$ 。

#### 2.2 统计量的分布

##### 2.2.1 卡方分布

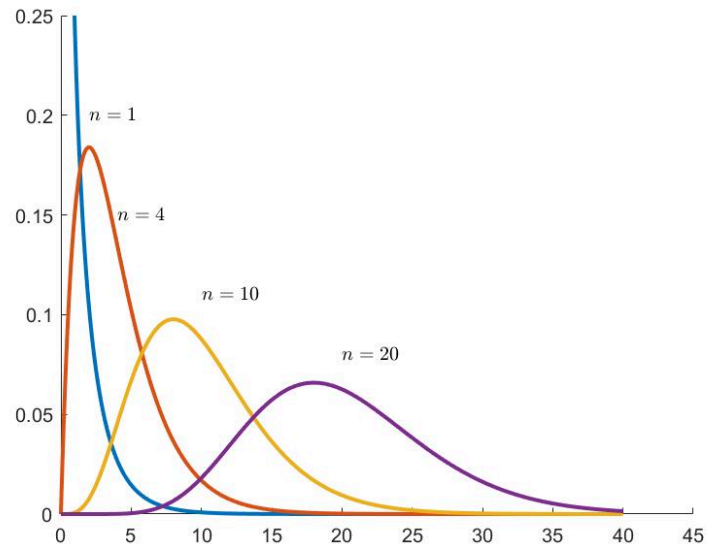
定义：样本 $X_1, X_2, \dots, X_n$ 来自总体 $X \sim N(0, 1)$ ，统计量 $\chi^2 = X_1^2 + X_2^2 + \dots + X_n^2$ ，称 $\chi^2$ 服从自由度为 $n$ 的卡方分布，记作 $\chi^2 \sim \chi^2(n)$ 。

自由度表示 $\chi^2$ 统计量右边独立变量的个数。

概率密度：

$$f(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \cdot \Gamma(\frac{n}{2})} \cdot x^{\frac{n}{2}-1} \cdot e^{-\frac{x}{2}}, & x > 0 \\ 0, & \text{others} \end{cases}$$

概率密度图像：



性质：

1.  $E[\chi^2(n)] = n$ ,  $n$ 为自由度。
2.  $D[\chi^2(n)] = 2n$ .
3.  $X \sim \chi^2(m), Y \sim \chi^2(n)$ ,  $X, Y$ 独立,  $X + Y \sim \chi^2(m + n)$ .