# 第 0 讲　概统基本概念一览

## 一　随机事件与概率

### 1.随机试验

· 对随机现象进行观察、记录或试验，称为随机试验

### 2.样本空间

· 称随机试验的所有可能结果构成的集合为样本空间，常用字母 $S$ 来表示

### 3.样本点

· 样本空间的每一个元素，即试验的每一个结果称为样本点

### 4.随机事件

· 样本空间的任一子集称为随机事件，用大写字母 $A$ 等表示；只含有一个样本点的事件称为基本事件

### 5.发生/不发生

· 一次试验完成时，若试验所出现的结果属于某一事件，就称该事件发生，否则称该事件不发生

### 6.必然事件与不可能事件

· 样本空间代表的事件必然发生，称为必然事件；空集称为不可能事件

### 7.划分

· 设 $S$ 为某一随机试验的样本空间， $B_1,B_2,\cdots,B_n$ 为该试验的一组事件，且满足

① $B_iB_j=\varnothing$， $i\neq j$

② $B_1\cup B_2\cup\cdots\cup B_n=S$

则称 $B_1,B_2,\cdots,B_n$ 为 $S$ 的一个划分，或称 $S$ 的一个**完备事件组**

### 8.概率

· 设某随机试验对应的样本空间为 $S$，对 $S$ 中的任一事件 $A$，定义一个实数 $P(A)$，若它满足：

① 非负性： $P(A)\geq 0$

② 规范性： $P(S)=1$

③ 可列可加性：对 $S$ 中的可列个两两互不相容的事件 $A_1,A_2,\cdots,A_n,\cdots$，有

$$P(\bigcup_{i=1}^{+\infty}A_j)=\sum_{i=1}^{+\infty}P(A_i)$$

则称 $P(A)$ 为事件 $A$ 发生的概率

### 9.等可能概型（古典概型）

· 如果一个随机试验满足下面两个条件：

① 样本空间中样本点数有限（有限性）

② 出现每一个样本点的概率相等（等可能性）

就称这个试验为**等可能概型**，又称古典概型

**10. 条件概率**

· 称在已知 $B$ 事件发生的条件下事件 $A$ 发生的概率为对应的条件概率

**11. 独立**

· 设 $A,B$ 为两随机事件，当 $P(A)P(B)=P(AB)$ 时，称 $A,B$ 相互独立

## 二 随机变量

**1. 随机变量**

· 设随机试验的样本空间为 $S$，若 $X$ 为定义在样本空间上的单实值函数，$e\in S$，则称 $X$ 为随机变量

**2. 离散型随机变量**

· 若随机变量 $X$ 的所有可能取值为有限个或可列个，则随机变量 $X$ 为离散型随机变量

**3. 概率分布律**

· 设 $X$ 为离散型随机变量，若其可能取值为 $x_1,x_2,\cdots,x_k,\cdots$ 则称

$$P(X=x_k)=p_k$$

为 $X$ 的概率分布律或**概率分布列**，简称 $X$ 的分布律

**4. 概率分布函数**

· 设 $x$ 为任意实数，函数 $F(x)=P(X\le x)$ 称为随机变量 $X$ 的概率分布函数，简称分布函数

**5. 连续型随机变量**

· 设随机变量 $X$ 的分布函数为 $F(x)$，若存在一个非负实值函数 $f(x)$，$-\infty<x<+\infty$，使得对任意实数 $x$，有

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

则称 $X$ 为连续型随机变量

**6. 概率密度函数**

· 称上个定义中的 $f(x)$ 为 $X$ 的概率密度函数，简称密度函数

## 三 多维随机变量

**1. 联合分布律**

· 称 $P(X=x_i,Y=y_j)=p_{ij}$ 为 $(X,Y)$ 的联合概率分布律，简称联合分布律

**2. 边际分布律**

· 称多维随机变量中单个随机变量的分布律为 $(X,Y)$ 的边际分布律

### 3. 条件分布律

- 已知 $(X, Y)$ 中一个随机变量取值确定的条件下，另一个随机变量的分布律称为 $(X, Y)$ 的条件分布律

### 4. 联合分布函数

- 设二维随机变量 $(X, Y)$，对于任意的实数 $x, y$，称函数

$$F(x, y) = P(X \le x, Y \le y)$$

为随机变量 $(X, Y)$ 的联合分布函数

### 5. 边际分布函数

- 二维随机变量中单个随机变量的分布函数为边际分布函数

### 6. 条件分布函数

- 已知 $(X, Y)$ 中一个随机变量取值确定的条件下，另一个随机变量的分布函数为条件分布函数

### 7. 概率密度函数

- 设二维随机变量 $(X, Y)$ 的联合分布函数 $F(x, y)$，若存在二元非负函数 $f(x, y)$，使对任意的实数 $x, y$ 有

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\,\mathrm{d}u\mathrm{d}v$$

则称 $(X, Y)$ 为二维连续型随机变量，称 $f(x, y)$ 为 $(X, Y)$ 的联合概率密度函数

### 8. 边际密度函数

- 二维随机变量中单个随机变量的密度函数为边际密度函数

### 9. 条件密度函数

- 某个随机变量取值确定的条件下，另一个随机变量的密度函数为条件密度函数

## 四　随机变量的数字特征

### 1. 数学期望

- 设离散型随机变量 $X$ 的概率分布律为 $p(X_i) = p_i$，$i = 1, 2, \cdots$，若级数

$$\sum_{i=1}^{+\infty} x_i p_i$$

绝对收敛，则称该级数为 $X$ 的数学期望，记为 $E(X)$

- 设连续型随机变量 $X$ 的概率密度函数为 $f(x)$，若

$$\int_{-\infty}^{+\infty} |x| f(x)\mathrm{d}x$$

存在，则称 $\int_{-\infty}^{+\infty} x f(x)\mathrm{d}x$ 为 $X$ 的数学期望，记为 $E(X)$

### 2. 方差

- 设随机变量 $X$ 的数学期望 $E(X)$ 存在，若 $E[(X - E(X))^2]$ 存在，则称其为 $X$ 的方差，记为 $\mathrm{Var}(X)$ 或 $D(X)$

3.**协方差**

· 对于数学期望都存在的随机变量 $X$ 和 $Y$，当 $X-E(X)$ 和 $Y-E(Y)$ 的数学期望都存在时，称

$$\mathrm{Cov}(X,Y)=E[(X-E(X))(Y-E(Y))]$$

为 $X$ 与 $Y$ 的协方差

4.**相关系数**

· 对于随机变量 $X$ 和 $Y$，当 $E(X^2)$ 与 $E(Y^2)$ 均在且 $\mathrm{Var}(X)\mathrm{Var}(Y)\neq 0$ 时，称

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

为 $X$ 与 $Y$ 的相关系数

5.**不相关**

· 若 $X$ 与 $Y$ 的相关系数 $\rho_{XY}=0$，则称 $X$ 与 $Y$ 不相关

## 五　大数定律

1.**依概率收敛**

· 设 $\{Y_n\}$ 为一随机变量序列，$c$ 为一常数，若对任意的 $\varepsilon>0$，都有

$$\lim_{n\to+\infty}P(|Y_n-c|\geq\varepsilon)=0 \quad 或 \quad \lim_{n\to+\infty}P(|Y_n-c|<\varepsilon)=1$$

则称 $\{Y_n\}$ 依概率收敛于 $c$，记为 $Y_n\xrightarrow{P}c$，$n\to+\infty$

## 六　数理统计基础

1.**总体　个体**

· 研究对象的全体称为总体 $X$，总体中的每个成员称为个体

2.**样本**

· 数理统计的任务是从总体中抽取一部分个体进行分析，称为样本 $X_i$

3.**样本容量**

· 称样本的数量为样本容量

4.**样本值**

· 称样本 $X_i$ 的观测值 $x_i$ 为样本值

5.**随机样本**

· 如果从总体中抽取样本是随机的，称为随机样本

6.**简单随机样本**

· 设总体 $X$ 是具有分布函数 $F$ 的随机变量，$X_1,\cdots,X_n$ 是来自总体 $X$ 的随机样本。若满足

①（独立性）$X_1,\cdots,X_n$ 是相互独立的随机变量

②（代表性）每个 $X_i$ 都与总体 $X$ 具有相同的分布函数

则称 $X_1,\cdots,X_n$ 为取自总体 $X$ 的**简单随机样本**

7. **统计量**

· 设 $X_1,\cdots,X_n$ 是来自总体 $X$ 的一个样本，$g(X_1,\cdots,X_n)$ 是样本 $X_1,\cdots,X_n$ 的函数，若 $g$ 不含未知参数，则称 $g(X_1,\cdots,X_n)$ 是一个**统计量**

## 七　参数估计

1. **点估计**

· 构造一个适当的统计量 $\hat{\theta}$ 用来估计分布中的未知参数 $\theta$ 的值，称为**点估计**

2. **无偏估计**

· 设 $\theta\in\Theta$ 是总体 $X$ 的待估参数，$X_1,\cdots,X_n$ 是来自总体 $X$ 的样本。若估计量 $\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$ 的数学期望存在，且满足 $E(\hat{\theta})=\theta$，$\forall\theta\in\Theta$，则称 $\hat{\theta}$ 是 $\theta$ 的**无偏估计**

3. **有效性**

· 设 $\hat{\theta}_1=\hat{\theta}_1(X_1,X_2,\cdots,X_n)$ 与 $\hat{\theta}_2=\hat{\theta}_2(X_1,X_2,\cdots,X_n)$ 都是参数 $\theta$ 的无偏估计，若 $\forall\theta\in\Theta$，$\mathrm{Var}(\hat{\theta}_1)\le\mathrm{Var}(\hat{\theta}_2)$，且至少存在一个 $\theta$ 使得不等号成立，则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ **有效**

4. **均方误差**

· 设 $\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$ 是总体参数 $\theta$ 的估计量，称 $E[(\hat{\theta}-\theta)^2]$ 为估计量 $\hat{\theta}$ 的**均方误差**，称为 $\mathrm{Mse}(\hat{\theta})$

5. **相合估计**

· 设 $\hat{\theta}_n=\hat{\theta}(X_1,X_2,\cdots,X_n)$ 是总体参数 $\theta$ 的估计量，若对任意 $\varepsilon>0$，有

$$\lim_{n\to+\infty}P(|\hat{\theta}_n-\theta|<\varepsilon)=1$$

即 $\hat{\theta}_n$ 依概率收敛于 $\theta$，则称 $\hat{\theta}_n$ 是 $\hat{\theta}$ 的**相合估计**，记为 $\hat{\theta}_n\xrightarrow{P}\theta$，$n\to+\infty$

6. **双侧置信区间**

· 设总体为 $X$，$\theta$ 为待估参数，$X_1,X_2,\cdots,X_n$ 是来自总体 $X$ 的样本，统计量 $\hat{\theta}_L=\hat{\theta}_L(X_1,X_2,\cdots,X_n)$ 和 $\hat{\theta}_U=\hat{\theta}_U(X_1,X_2,\cdots,X_n)$ 满足 $\hat{\theta}_L<\hat{\theta}_U$，且对给定的 $\alpha\in(0,1)$ 和任意的 $\theta$，有

$$P(\hat{\theta}_L<\theta<\hat{\theta}_U)\ge1-\alpha$$

则称随机区间 $\hat{\theta}_L<\hat{\theta}_U$ 是参数 $\theta$ 的**置信水平**为 $1-\alpha$ 的**置信区间**，$\hat{\theta}_L,\hat{\theta}_U$ 分别称为是 $\theta$ 的置信水平 $1-\alpha$ 的**双侧置信下限**和**双侧置信上限**

7. **单侧置信区间**

· 对给定的 $\alpha\in(0,1)$，如果统计量满足

$$P(\hat{\theta}_{\mathrm{L}} < \theta) \geq 1-\alpha$$
$$P(\theta < \hat{\theta}_{\mathrm{U}}) \geq 1-\alpha$$

则称 $\hat{\theta}_L,\hat{\theta}_U$ 分别是 $\theta$ 的置信水平 $1-\alpha$ 的单侧置信下限和单侧置信上限

9. **枢轴量**

- 设总体 $X$ 的密度函数 $f(x;\theta)$，其中 $\theta$ 为待估参数，并设 $X_1,X_2,\cdots,X_n$ 是来自总体 $X$ 的样本，如果样本和参数 $\theta$ 的函数 $G(X_1,X_2,\cdots,X_n,\theta)$ 的分布完全已知，且形式上不依赖其它未知参数，称 $G(X_1,X_2,\cdots,X_n,\theta)$ 为枢轴量

## 八　假设检验

1. **统计假设**

- 对总体的分布或其中的参数提出的假设称为统计假设，简称为假设，一般用 $H$ 表示

2. **原假设　备择假设**

- 根据样本资料想得到推翻的假设称为原假设 $H_0$，与其完全相反的假设称为备择假设 $H_1$

3. **检验形式**

- 单侧检验：$H_0$： $\theta \geq \theta_0$　　$H_1$： $\theta < \theta_0$（左侧检验）

  $H_0$： $\theta \leq \theta_0$　　$H_1$： $\theta > \theta_0$（右侧检验）
- 双侧检验：$H_0$： $\theta = \theta_0$　　$H_1$： $\theta \neq \theta_0$

4. **检验统计量**

- 取值与原假设是否成立有密切关系的统计量称为检验统计量

5. **拒绝域**

- 如果检验统计量的值落入了拒绝域，则应拒绝原假设

6. **两类错误**

- 如果拒绝了一个真实的原假设，则称为第 Ⅰ 类错误
- 如果接受了一个错误的原假设，则称为第 Ⅱ 类错误

7. **显著性水平**

- 假设检验的要求是犯 Ⅰ 类错误的概率不超过显著性水平 $\alpha$ 时，尽可能减少犯第 Ⅱ 类错误的概率

8. **$P\_$ 值**

- 当原假设 $H_0$ 为真时，检验统计量取比观察到的结果更为极端的数值的概率称为 $P-$值

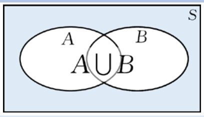
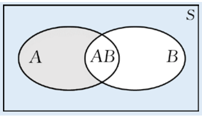
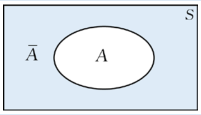
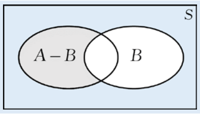
# 第 1 讲　随机事件与概率

## 知识梳理

### 一　事件关系与运算

**1.事件的运算及规律**

① 四种事件运算类型

| | 符号 | 定义 | Venn 图 |
|---|---|---|---|
| 和 | $A\cup B$ | $A\cup B=\{x:x\in A$ 或 $x\in B\}$ |  |
| 积 | $A\cap B$<br>$AB$ | $AB=\{x:x\in A$ 且 $x\in B\}$ |  |
| 逆 | $\bar{A}$ | $\bar{A}=\{x:x\notin A\}$ |  |
| 差 | $A-B$ | $A-B=A\cap\bar{B}=\{x:x\in A$ 且 $x\notin B\}$ |  |

② 事件运算的规律

| 规律名称 | 内容 | |
|---|---|---|
| 交换律 | $A\cup B=B\cup A$ | $AB=BA$ |
| 结合律 | $(A\cup B)\cup C=A\cup(B\cup C)$ | $(AB)C=A(BC)$ |
| 分配律 | $(A\cup B)C=AC\cup BC$ | $(AB)\cup C=(A\cup C)(B\cup C)$ |
| 德 · 摩根律 | $\overline{\bigcup_{j=1}^{n}A_j}=\bigcap_{j=1}^{n}\overline{A_j}$ | $\overline{\bigcap_{j=1}^{n}A_j}=\bigcup_{j=1}^{n}\overline{A_j}$ |

**2.事件之间的关系**

| | 符号 | 定义 | 说明 |
|---|---|---|---|
| 包含 | $B\subset A$ | 事件 B 发生一定导致事件 A 发生 | B 是 A 的子事件（子集）<br>Venn 图中 B 在 A 当中 |
| 相等 | $A=B$ | $B\subset A$ 且 $A\subset B$ | —— |
| 互斥 | $A\cap B=\varnothing$ | —— | Venn 图中二者没有重叠区域 |
| 独立 | —— | $P(AB)=P(A)P(B)$ 或 $P(A\mid B)=P(A)$ | 通过 定义 或 实际逻辑 判断 |

## 二　概率的运算性质

### 1. 概率的基本运算性质

① 和事件

**和事件的概率（双事件）**

$$\boxed{P(A\cup B)=P(A)+P(B)-P(AB)}$$

· 推论：若 $AB=\varnothing$，则 $P(A\cup B)=P(A)+P(B)$

若 $B=\overline{A}$，则 $P(\overline{A})=1-P(A)$

**和事件的概率（三事件）**

$$P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(AB)-P(BC)-P(AC)+P(ABC)$$

② 积事件

**积事件的概率**

$A$ 和 $B$ 相互独立 $\Leftrightarrow$ $P(AB)=P(A)P(B)$

③ 差事件

**差事件的概率**

$$\boxed{P(A-B)=P(A)-P(AB)}$$

### 2. 条件概率的基本运算性质

**条件概率**

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

① 乘法公式

**条件概率乘法公式**

$$\boxed{P(AB)=P(A)\cdot P(B\mid A)=P(B)\cdot P(A\mid B)}$$

② 全概率公式

**全概率公式**

$$\boxed{P(A)=\sum_{i=1}^{n}P(B_i)P(A\mid B_i)}$$

· $B_1,\cdots,B_n$ 是 S 的一个划分

③ 贝叶斯公式

**贝叶斯公式**

$$\boxed{P(B_k\mid A)=\frac{P(B_kA)}{P(A)}=\frac{P(B_k)P(A\mid B_k)}{\sum_{j=1}^{n}P(B_j)P(A\mid B_j)}}$$

## 三　古典概型

### 1. 古典概型事件概率

| 古典事件概率 |
| --- |
| $P(A)=\dfrac{A\text{ 样本数}}{S\text{ 样本数}}$ |

### 2. 取样模型

- 盒子里有 $N$ 个球，要取出 $n$ 个球

| 取法 | 总样本数 | 说明 |
| --- | --- | --- |
| 有放回地依次取出 $n$ 次 | $N^n$ | 相当于每次从 $N$ 个球里选一个 |
| 无放回地依次取出 $n$ 个 | $\mathrm{A}_N^n=N!\,/\,(N-n)!$ | 相当于从 $N$ 个球中选 $n$ 个并排列 |
| 一次性取出 $n$ 个 | $\mathrm{C}_N^n=N!/[(n!)(N-n)!]$ | —— |

- 事件样本数要**根据事件化为先后步骤**求排列组合

  如无放回依次取出时，里面有 $m$ 个红球，要求第 3 次取出红球

  ① 第一步：从 $m$ 个红球里选一个作为第 3 个球：$\mathrm{C}_m^1$

  ② 第二步：从剩下 $N-1$ 个球里选择 $n-1$ 个：$\mathrm{A}_{N-1}^{n-1}$

  则样本数为 $\mathrm{C}_m^1\mathrm{A}_{N-1}^{n-1}$

### 3. 抽签问题的结论

- 在无放回模型中，第 $k$ 次取到特定球的概率与 $k$ 无关，这使得我们可直接通过计算第 1 次的概率求解

# 题型解析

## 一　计算古典概型事件的概率

### 1.题型简述与解法

- 题干描述实际生活场景，且没有提供任何已知的概率
- 主要流程：求出所求事件的样本数和总样本数，相除后计算
- 计算样本数时需要灵活运用排列组合知识，将题目中的背景转换为熟悉的模型进行计算
- 如果发现计数需要复杂的分类讨论，不妨反过来求逆事件的样本数，可能有奇效

### 2.历年考试典型例题

**例 1**　（19—20 春夏）一盒中有 5 个球，其中 3 个红球，2 个白球，采用不放回抽样取 3 个球，则至少取到 2 个红球的概率为____________，第 2 次取到红球的概率为____________.

**解**　① 由于“至少取到 2 个红球”没有对抽取顺序提出要求，因此可以视为一次性取出

套用到抽球模型，$N=5$，$n=3$，总样本数 $C_5^3=20$

“至少取到 2 个红球”包括“2 个红球 1 个白球”和“3 个红球”两种情况

“2 个红球 1 个白球”相当于“3 个红球中选 2 个 且 2 个白球中选 1 个”，样本数 $C_3^2C_2^1=6$

“3 个红球”相当于“3 个红球中选 3 个”再全排列，样本数 $C_3^3=1$

→ 该事件样本数 $6+1=7$，概率 $7/10=\boxed{0.7}$

② 由于“第 2 次取到红球”对顺序提出要求，因此视为不放回依次取出，总样本数 $A_5^3=60$

“第 2 次取到红球”相当于“3 个红球选 1 个 且 4 个球中选 2 个全排列”，样本数 $C_3^1A_4^2=36$

→ 该事件概率为 $36/60=0.6$

实际上就是抽签问题，直接使用结论：$3/5=\boxed{0.6}$

**例 2**　（18—19 秋冬）某小区有 $a$ 个人申请小区停车位（$a\geq 3$），而小区的停车位只有 $b$ 个（$0<b\leq a-2$）。管理者决定采用随机抽签方法确定停车位使用权。则排在第 3 位的抽签者抽中停车位的概率为__________；前三个人中恰好有一人抽中停车位的概率为____________.

**解**　① 利用结论，第 3 位抽签抽中概率与第 1 位一致，即 $\boxed{b/a}$

② 由于“前三个人中恰好有一人抽中停车位”没有提出顺序要求，因此视为一次性抽取

总样本数 $C_a^3$，“前三个人中恰好有一人抽中停车位”相当于“从 $b$ 张中签中选择 1 张”且“$a-b$ 张未中签中选择 2 张”，所以样本数 $C_b^1C_{a-b}^2$，因此概率 $\boxed{(C_b^1C_{a-b}^2)/C_a^3}$

## 二　计算抽象事件的概率

### 1.题型简述与解法

- 题干直接使用 $A$ $B$ 等来表示事件，并提供已知概率）
- 盯住目标，利用概率公式将所求事件的概率转化为已知概率的表达式
- 需要熟练地由事件关系得到概率间的关系，初学者可借助韦恩图做题训练
- 遇到条件概率，可通过定义式转化为一般的概率，或直接利用条件概率性质

### 2.历年考试典型例题

**例 1**　（19—20 **秋冬**）设 $A,B$ 为两个随机事件，已知 $P(A)=0.45$，$P(A|B)=P(\overline{A}|\overline{B})=0.6$，则 $A$ 与 $B$ 是否独立？答：________；$P(B)=$________.

**解**　① 目标是判断 $A$ 和 $B$ 是否独立，等价于判断 $P(AB)=P(A)P(B)$ 或 $P(A|B)=P(A)$ 是否成立

而题干中直接就有 $P(A)=0.45$，$P(A|B)=0.6$，因此 $P(A|B)\neq P(A)$，即 $\boxed{\text{不独立}}$

② 方法一：将条件概率化为定义式

$$P(A|B)=\frac{P(AB)}{P(B)}=0.6 \rightarrow P(AB)=0.6P(B) \qquad (1)$$

$$P(\overline{A}|\overline{B})=\frac{P(\overline{A}\overline{B})}{P(\overline{B})}=\frac{P(\overline{A\cup B})}{1-P(B)}=\frac{1-P(A\cup B)}{1-P(B)}=\frac{1-P(A)-P(B)+P(AB)}{1-P(B)}=0.6 \qquad (2)$$

- （1） 及 $P(A)=0.45$ 代入 （2），得

$$\frac{0.55-0.4P(B)}{1-P(B)}=0.6\text{，解得 } P(B)=\boxed{0.25}$$

方法二：根据条件概率的性质

$$P(\overline{A}|\overline{B})=1-P(A|\overline{B})=0.6 \rightarrow P(A|\overline{B})=0.4$$

- 已知 $P(A|B)$ 和 $P(A|\overline{B})$，我们就能用全概率公式

$$P(A)=P(A|B)P(B)+P(A|\overline{B})P(\overline{B})$$

代入数据：　$0.45=0.6P(B)+0.4(1-P(B))$

解得 $P(B)=\boxed{0.25}$

**例 2**　（18—19 **春夏**）设 $A,B,C$ 为三个随机事件，已知 $A$ 发生时 $B$ 必定发生，$P(A)=0.3$，$P(B)=0.4$，$P(A|C)=0.4$，$P(B|C)=0.5$，$P(C)=0.6$，则 $P(C|A)=$________；$P(A\cup B\cup C)=$________.

**解**　① 从条件概率的定义出发

$$P(C|A)=\frac{P(AC)}{P(A)}=\frac{P(A|C)P(C)}{P(A)}=\frac{0.4\times0.6}{0.3}=\boxed{0.8}$$

② “已知 $A$ 发生时 $B$ 必定发生” $\rightarrow$ $A\subset B$ $\rightarrow$ $A\cup B=B$，因此

$$P(A\cup B\cup C)=P(B\cup C)=P(B)+P(C)-P(BC)=P(B)+P(C)-P(B|C)P(C)=\boxed{0.7}$$

**例 3** （15－16 **春夏**）设 $A,B,C$ 是三个事件，$P(A)=a$，$P(B)=b$，$P(C)=c$.（1）若 $A,B,C$ 相互独立，则 $P(\bar{B}\cup\bar{C}|AB)=$______；（2）若 $A\subset B$，且 $B$ 与 $C$ 不相容，则 $P(\bar{B}\cup\bar{C}|AB)=$______.

**解** 两问求的是同一个概率，形式复杂，首先简化它：

$$P(\bar{B}\cup\bar{C}|AB)=\frac{P((\bar{B}\cup\bar{C})AB)}{P(AB)}=\frac{P(AB\bar{B}\cup AB\bar{C})}{P(AB)}=\frac{P(AB\bar{C})}{P(AB)}$$

（1）$A,B,C$ 相互独立 $\rightarrow$ $AB$ 与 $C$ 独立

$$P(\bar{B}\cup\bar{C}|AB)=\frac{P(AB\bar{C})}{P(AB)}=P(\bar{C})=1-P(C)=1-c$$

（2）$A\subset B$ $\rightarrow$ $AB=A$，则

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

$B$ 与 $C$ 不相容 且 $A\subset B$ $\rightarrow$ $A$ 与 $C$ 也不相容 $\rightarrow$ $A\subset\bar{C}$ $\rightarrow$ $A\bar{C}=A$，因此

$$P(\bar{B}\cup\bar{C}|AB)=\frac{P(AB\bar{C})}{P(AB)}=\frac{P(A\bar{C})}{P(A)}=\frac{P(A)}{P(A)}=1$$

**例 4** （14－15 **秋冬**）设随机事件 $A$ 与 $B$ 独立，$P(A)=0.4$，$P(A\cup B)=0.64$，则 $P(B)=$________，$P(\bar{A}|A\cup B)=$ ________.

**解** ① $P(A\cup B)=0.64$ $\rightarrow$ 和事件公式：$P(A\cup B)=P(A)+P(B)-P(AB)=0.64$

$A$ 与 $B$ 独立 $\rightarrow$ $P(AB)=P(A)P(B)$ $\rightarrow$ $P(A\cup B)=P(A)+P(B)-P(A)P(B)=0.64$

代入 $P(A)=0.4$，解得 $P(B)=0.4$

② $$P(\bar{A}|A\cup B)=\frac{P(\bar{A}(A\cup B))}{P(A\cup B)}=\frac{P(\bar{A}B)}{P(A\cup B)}=\frac{P(\bar{A})P(B)}{P(A\cup B)}=\frac{0.6\times0.4}{0.64}=0.375$$

## 三 全概率与贝叶斯

### 1.题型简述与解法

· 题干描述的是实际生活场景，并且提供已知的概率，求全概率、贝叶斯事件

· 关键在于正确用字母描述事件，并将已知的条件转化为数学语言，主要靠日常逻辑

· 列全概率时，要确保划分与条件概率一一对应，不要有漏掉的划分事件

### 2.历年考试典型例题

**例 1** （19－20 **春夏**）设某地区居民患有某种疾病的概率为 0.001，用于疾病检测的方法存在误判，设患病者检测结果呈阳性的概率为 0.95，未患病者检测为阳性的概率为 0.002.若在该地区随机选一人进行检测，结果呈阳性，求 TA 的确患病的概率；若对 TA 独立进行两次检测，且假设两次检测

都处于相同的状态，如果结果都是阳性，求 TA 患病的概率。(保留 3 位小数)

解　（1）设事件 $A$ 为“检测阳性”，事件 $B$ 为“患病”，则事件 $\bar{B}$ 为“不患病”

“患有某种疾病的概率为 0.001” $\rightarrow$ $P(B)=0.001$，$P(\bar{B})=1-P(B)=0.999$

“患病者检测结果呈阳性的概率为 0.95” $\rightarrow$ $P(A|B)=0.95$

“未患病者检测为阳性的概率为 0.002” $\rightarrow$ $P(A|\bar{B})=0.02$

“结果呈阳性，求 TA 的确患病的概率” $\rightarrow$ $P(B|A)$

$\therefore$ $P(A)=P(A|B)P(B)+P(A|\bar{B})P(\bar{B})=0.95\times0.001+0.002\times0.999=$

$$\therefore\ P(B|A)=\frac{P(A|B)P(B)}{P(A|B)P(B)+P(A|\bar{B})P(\bar{B})}=\frac{0.95\times0.001}{0.95\times0.001+0.002\times0.999}=\boxed{0.322}$$

（2）设事件 $A_i$ 为“第 $i$ 次进行检测为阳性”，则 $A_1$、$A_2$ 相互独立

“两次检测都是阳性” $\rightarrow$ $P(A_1A_2)=P(A_1)P(A_2)=P^2(A)$（第 1 问可求出）

“患病者且 2 次检测都是阳性” $\rightarrow$

$$P(A_1A_2B)=P(A_1A_2|B)P(B)=P(A_1|B)P(A_2|B)P(B)$$

“2 次检测都是阳性，求 TA 患病的概率” $\rightarrow$ $P(B|A_1A_2)$

$$\therefore\ P(B|A_1A_2)=\frac{P(A_1A_2|B)P(B)}{P(A_1A_2)}=\frac{P(A_1|B)P(A_2|B)P(B)}{P^2(A)}=\boxed{0.996}$$

例 8　（17—18 **春夏**）小王喜欢玩某款一对一对战游戏，该游戏会根据玩家自身的等级分随机匹配等级分相近的玩家。假设在一局中小王遇到等级分高于，等于，低于自己的玩家的概率分别为 0.4，0.2，0.4，遇到等级分高的玩家，小王胜，平，负的概率分别 0.3，0.3，0.4，遇到等级分相同的玩家，小王胜，平，负的概率分别 0.4，0.4，0.2，遇到等级分高的玩家，小王胜，平，负的概率分别 0.5，0.3，0.2.

（1）求在一局中小王胜的概率；

（2）若已知小王胜了一局，求此局对手是等级分高的玩家的概率；

解　（1）设遇到等级分高中低的玩家分别为 $A_1A_2A_3$，胜平负分别是 $B_1B_2B_3$，则小王胜的概率

$$P(B_1)=P(A_1)P(B_1|A_1)+P(A_2)P(B_1|A_2)+P(A_3)P(B_1|A_3)$$

$$=0.4\times0.3+0.2\times0.4+0.4\times0.5=0.4$$

（2）“已知小王胜了一局，此局对手是等级分高的玩家” $\rightarrow$ $P(A_1|B_1)$

$$\therefore\ P(A_1|B_1)=\frac{P(A_1)P(B_1|A_1)}{P(B_1)}=\frac{0.4\times0.3}{0.4}=0.3$$

# 第 2 讲　一维随机变量的概率分布

## 知识梳理

### 一　一维随机变量的概率分布

**1.离散型：概率分布律**

| **X 的概率分布律** |
|---|
| $P(X=x_k)=p_k$ |

性质：$p_k \geq 0$ 且 $\sum_{k=1}^{+\infty} p_k = 1$

**2.连续型：概率密度函数**

| **概率密度函数** |
|---|
| $F(x)=\int_{-\infty}^{x} f(t)\mathrm{d}t$ |

性质：$f(x)\geq 0$ 且 $\boxed{\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1}$

· 可计算：$X$ 落在特定区间的概率

| **X 落在区间 $(a,b)$ 的概率** |
|---|
| $P(x_1<x<x_2)=\int_{x_1}^{x_2} f(t)\mathrm{d}t$ |

**3.概率分布函数**

| **概率分布函数** |
|---|
| $F(x)=P(X\leq x)$ |

性质：$0\leq F(x)\leq 1$ 且 $F(-\infty)=0$，$F(+\infty)=1$

· 对于离散型，$F(x)=\sum_{x_i\leq x} P(x=x_i)$，且 $F(x)$ 仅在 $x=x_i$ 处不连续

· 对于连续型，$F(x)=\int_{-\infty}^{x} f(t)\mathrm{d}t$，$F(x)$ 处处连续，且在 $f(x)$ 的连续点 $x$ 处，$\boxed{F'(x)=f(x)}$

### 二　一维随机变量的数字特征

**1.数学期望**　$E(X)$

| **离散型数学期望** | **连续型数学期望** |
|---|---|
| $E(g(X))=\sum_{i=1}^{+\infty} g(x_i)p_i$ | $E(g(X))=\int_{-\infty}^{+\infty} g(x)f(x)\mathrm{d}x$ |

· $g(X)$ 为随机变量 $X$ 的函数，包括 $X$ 本身

**2.方差**

| **方差** |
|---|
| $\mathrm{Var}(X)=E(X^2)-[E(X)]^2$ |

# 题型解析

## 四　一维离散型随机变量分布求解

### 1.题型简述与解法

① **求一维离散型随机变量的分布律**

- 题干描述实际生活场景，往往结合全概率、贝叶斯的知识
- 主要流程：列出 $X$ 的所有可能取值 → 计算每个取值的概率 → 列表

| $X$ | $x_1$ | $x_2$ | …… | $x_n$ |
|---|---|---|---|---|
| $P(X=x_i)$ | $p_1$ | $p_2$ | …… | $p_n$ |

- 也有考过根据分布函数求分布律，此时注意分布函数中的“阶跃点” $x_i$，其阶跃值就是对应概率

② **根据分布律求事件概率**

- 枚举出所有满足条件的 $x_i$，将它们的概率相加（如果枚举项太多，不妨反求逆事件的概率）

③ **求期望、方差**

- 代入公式计算即可，求 $g(X)$ 的期望时最好在草稿纸上把 $g(X)$ 以及对应概率列出来，方便计算

### 2.历年考试典型例题

**例 1**　（15—16 秋冬）一学徒工在一台机床上加工 2 个零件，第 1 个零件合格的概率为 0.6；在第 1 个是合格品的条件下，第 2 个合格的概率为 0.8；在第 1 个不合格的条件下，第 2 个合格的概率是 0.5. $X$ 表示合格零件数，求 $X$ 的分布律.

**解**　因为一共加工 2 个零件，所以 $X$ 的取值为 0，1，2

$X=0$ → 第 1 个零件不合格，第 2 个零件不合格，由条件概率：

$$P(X=0)=P(\text{第1次不合格,第2次不合格})=P(\text{第1次不合格})P(\text{第2次不合格}\mid\text{第1次不合格})$$
$$=(1-0.6)(1-0.5)=0.2$$

$X=2$ → 第 1 个零件合格，第 2 个零件合格，由条件概率：

$$P(X=2)=P(\text{第1次合格,第2次合格})=P(\text{第1次合格})P(\text{第2次合格}\mid\text{第1次合格})$$
$$=0.6\times0.8=0.48$$

可直接得到 $P(X=1)=1-0.2-0.48=0.32$，因此 $X$ 的分布律如下：

| $i$ | 0 | 1 | 2 |
|---|---|---|---|
| $P(X=i)$ | 0.2 | 0.32 | 0.48 |

**例 2**　（20—21 秋冬）设随机变量 $X$ 的分布函数 $F(x)=\begin{cases}0, & x<-1\\ 0.4, & -1\le x<0\\ 0.6, & 0\le x<1\\ 1, & x\ge1\end{cases}$ 则 $P(X=0)=$________.

**解**　$x=0$ 处为阶跃点，$F(x)$ 由 0.4 阶跃至 0.6，因此 $P(X=0)=0.6-0.4=\boxed{0.2}$

## 五　一维连续型随机变量分布求解

### 1.题型简述与解法

**① 已知密度函数或分布函数，求其中的未知常数**

· 密度函数：根据积分等于 1 解出

· 分布函数：根据 $F(x)$ 连续，在特定点（如分段函数点）求左右极限，解出

**② 求事件概率和分布函数**

· 求事件概率将事件转换成 $X$ 的取值范围，在对应的 $x$ 区间上对密度函数积分或分布函数相减

· 求分布函数要分类讨论，确保 $x$ 覆盖整个实数域，常用密度函数的分段点作为分类讨论的标准

**③ 求期望、方差**

· 代入公式积分即可，常考幂函数的积分

### 2.历年考试典型例题

**例 1**　（15—16 **春夏　节选**）设连续型随机变量 $X$ 的分布函数为

$$F(x)=\begin{cases}0, & x<0\\ c[(x+1)^2-1], & 0\le x<2\\ 1, & x\ge 2\end{cases}.$$

（1）求 $c$ 的值；（2）求 $P(X>1)$ 的值.

**解**　（1）由于分布函数连续，因此在 $x=2$ 处，$\lim\limits_{x\to 2^-}F(x)=F(2)$，即 $8c=1$，解得 $\boxed{c=1/8}$

（2）$P(X>1)=1-P(X\le 1)=1-F(1)=1-\dfrac{1}{8}[2^2-1]=\boxed{5/8}$

**例 2**　（16—17 **秋冬　节选**）设 $X$ 的密度函数 $f(x)=\begin{cases}x/4, 0<x<2\\ 1/4, 2\le x<4\\ 0, \quad 其他\end{cases}$，求 $X$ 的分布函数 $F(x)$；

**解**　$f(x)$ 为分段函数，因此 $F(x)$ 要分类讨论（根据 $f(x)$ 的分段点，分类讨论的界限分别为 0、2、4）

· 当 $x<0$ 时，$F(x)=0$；当 $x\ge 4$ 时，$F(x)=1$；

· 当 $0\le x<2$ 时，$F(x)=\displaystyle\int_0^x \frac{t}{4}\mathrm{d}t=\frac{x^2}{8}$

· 当 $2\le x<4$ 时，$F(x)=\displaystyle\int_0^2 \frac{t}{4}\mathrm{d}t+\int_2^x \frac{1}{4}\mathrm{d}t=\frac{4}{8}+\frac{x-2}{4}=\frac{x}{4}$（千万不要忘了前面的积分！）

· 综上所述：

$$F(x)=\begin{cases}0, & x<0\\ x^2/8, & 0\le x<2\\ x/4, & 2\le x<4\\ 1, & x\ge 4\end{cases}$$

**例 3** （16—17 **春夏 节选**）设 $X$ 的密度函数 $f(x)=\begin{cases} cx, 0<x<2 \\ 0, \quad 其他 \end{cases}$，求：

（1）$c$ 的值，$E(X)$，$D(X)$；

**解** （1）$\int_{-\infty}^{+\infty} f(x)\mathrm{d}x=\int_0^2 cx\mathrm{d}x=2c=1 \rightarrow \boxed{c=1/2}$

（2）$E(X)=\int_{-\infty}^{+\infty} xf(x)\mathrm{d}x=\int_0^2 \frac{1}{2}x^2\mathrm{d}x=\boxed{4/3}$

（3）$E(X^2)=\int_{-\infty}^{+\infty} x^2 f(x)\mathrm{d}x=\int_0^2 \frac{1}{2}x^3\mathrm{d}x=2$

$D(X)=E(X^2)-[E(X)]^2=\boxed{2/9}$

# 第 3 讲　二维随机变量的概率分布

## 知识梳理

### 一　二维离散型随机变量的分布律

1. **联合分布律** $p_{ij}$

**X，Y 的联合分布律**

$$\boxed{P(X=x_i,Y=y_j)=p_{ij}}$$

性质：$p_{ij}\geq 0$ 且 $\sum_{i=1}^{+\infty}\sum_{j=1}^{+\infty}p_{ij}=1$

2. **边际分布律** $p_i$　$p_j$

**X 的边际分布律**

$$\boxed{P(X=x_i)=\sum_{j=1}^{\infty}p_{ij}=p_i}$$

**Y 的边际分布律**

$$\boxed{P(Y=y_j)=\sum_{i=1}^{\infty}p_{ij}=p_j}$$

3. **条件分布律** $p_{ij}/p_i$　$p_{ij}/p_j$

**X 的条件分布律**

$$\boxed{P(X=x_i|Y=y_j)=\frac{p_{ij}}{p_j}}$$

**Y 的条件分布律**

$$\boxed{P(Y=y_j|X=x_i)=\frac{p_{ij}}{p_i}}$$

### 二　二维随机变量的分布函数

1. **联合分布函数** $F(x,y)$

**联合分布函数**

$$\boxed{F(x,y)=P(X\leq x,Y\leq y)}$$

2. **边际分布函数** $F_X(x)$　$F_Y(y)$

**X 的边际分布函数**

$$\boxed{F_X(x)=F(x,+\infty)}$$

**Y 的边际分布函数**

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

3. **条件分布函数** $F_{X|Y}(x\,|\,y_j)$　$F_{Y|X}(y\,|\,x_i)$

**X 的条件分布函数**

$$\boxed{F_{Y|X}(y\,|\,x_i)=P(Y\leq y\,|\,X=x_i)}$$

**Y 的条件分布函数**

$$\boxed{F_{X|Y}(x\,|\,y_j)=P(X\leq x\,|\,Y=y_j)}$$

### 三　二维连续型随机变量的密度函数

1. **联合密度函数** $f(x,y)$

**联合密度函数**

$$\boxed{F(x,y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x\mathrm{d}y}$$

性质：$f(x,y)\geq 0$ 且 $\iint\limits_{x,y\in(-\infty,+\infty)}f(x,y)\mathrm{d}x\mathrm{d}y=1$

- 通过在区域 $D$ 内积分来获得 $(X,Y)$ 落在某区域 $D$ 内的概率

**二维连续型概率求解**

$$P((X,Y)\in D)=\iint_D f(x,y)\mathrm{d}x\mathrm{d}y$$

2. **边际密度函数** $f_X(x)$ $f_Y(y)$

| X 的边际密度函数 | Y 的边际密度函数 |
| --- | --- |
| $f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y$ | $f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x$ |

3. **条件密度函数** $f_{X\mid Y}(x\mid y)$ $f_{Y\mid X}(y\mid x)$

| X 的条件密度函数 | Y 的条件密度函数 |
| --- | --- |
| $f_{Y\mid X}(y\mid x)=\dfrac{f(x,y)}{f_X(x)}$ | $f_{X\mid Y}(x\mid y)=\dfrac{f(x,y)}{f_Y(y)}$ |

## 四　二维随机变量的特征

1. **独立性**

**二维随机变量独立性判断**

二维随机变量 $X,Y$ 相互独立 $\Leftrightarrow$ $F(x,y)=F_X(x)F_Y(y)$ $p_{ij}=p_i p_j$ 或 $f(x,y)=f_X(x)f_Y(y)$

2. **数学期望**

| 离散型数学期望 | 连续型数学期望 |
| --- | --- |
| $E(h(X,Y))=\sum_{i=1}^{+\infty}h(x_i,y_j)p_{ij}$ | $E(h(X,Y))=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}h(x,y)f(x,y)\mathrm{d}x\mathrm{d}y$ |

3. **协方差**

**协方差**

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

4. **相关系数**

**相关系数**

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

5. **独立与不相关之间的关系**

- 独立和不相关看似是等价概念，但实际上这个相关指的是线性相关

**定理**　独立 $\Rightarrow$ 不相关　不相关 $\nRightarrow$ 独立

# 题型解析

## 六　求解二元离散型非常见随机变量分布

### 1. 题型简述与解法

① **根据实际背景求联合分布律**　例 1 (1)

· 一般使用二维表格来表示联合分布律，并额外加一行一列记录边际分布律

| $i/j$ | Y 的取值 | | | | $p_i$ |
|---|---|---|---|---|---|
| X 的取值 | $p_{11}$ | $p_{12}$ | …… | $p_{1n}$ | X 边际分布律 |
| | $p_{21}$ | $p_{22}$ | …… | $p_{2n}$ | |
| | …… | …… | …… | …… | |
| | $p_{m1}$ | $p_{m2}$ | …… | $p_{mn}$ | |
| $p_j$ | Y 的边际分布律 | | | | …… |

· 如果是根据实际背景，先列出 $X$ 和 $Y$ 的所有可能取值 $x_1,\cdots,x_m$， $y_1,\cdots,y_n$， 一一组合，列出表格

然后依次求出 $P(X=x_i,Y=y_j)$，填入表格，得到分布律

· 不要漏掉可能的取值情况，计算量大，不要出错，算完后可以看看加起来是否等于 1

② **根据联合分布律求事件概率、期望、协方差、相关系数**　例 1 (2)

· 求事件概率时，找出所有满足该事件的取值 $(X,Y)$，然后把对应的概率相加

· 求 $g(X,Y)$ 的数学期望时，先在联合分布律表格中计算并标记各个 $(X,Y)$ 对应的 $g(X,Y)$ 值

然后再列式 $E(g(X,Y))=\cdots$ 计算，如果 $g(X,Y)$ 为 0，可以直接跳过这一项，减少运算时间

③ **根据已知条件求联合分布律**　例 2

· 自带的已知条件：同行同列相加为边际分布律、边际分布律之和为 1

· 题目可能提供的条件：期望、协方差、相互独立等

· 解法：用未知数表示联合分布律，利用知识点中的公式表示出已知信息，联立求解

· $X$ 与 $Y$ 相互独立 $\Leftrightarrow$ 所有的 $P(X=x_i,Y=y_j)=P(X=x_i)P(Y=y_j)$

### 2. 历年考试典型例题

**例 1**　（14—15 **秋冬　节选**）盒中有 3 个红球，5 个白球，从中随机取一球，观察其颜色后放回，并从别处拿两个与取出的球同色的球放入盒中，搅匀后再从中取一球，$X_i=\begin{cases}1, & \text{第 } i \text{ 次取到红球}\\ 0, & \text{第 } i \text{ 次取到白球}\end{cases}$，$i=1,2$。

求：（1）$(X_1,X_2)$ 的联合分布律及 $X_2$ 的边际分布律；（2）求 $X_1$ 与 $X_2$ 的相关系数 $\rho_{X_1X_2}$ .

**解**　（1）两个随机变量为 $X_1$ 和 $X_2$，根据题干描述的背景，可以得到

$$P(X_1=1)=\frac{3}{8},\quad P(X_1=0)=\frac{5}{8}$$

如果第一次摸到的是红球，则第二次共有 5 红球 5 白球，此时摸到红球的概率为 1/2，即

$$P(X_2=1|X_1=1)=\frac{1}{2},\quad P(X_2=0|X_1=1)=\frac{1}{2}$$

如果第一次摸到的是白球，则第二次共有 3 红球 7 白球，此时摸到红球的概率为 3/10，即

$$P(X_2=1|X_1=0)=\frac{3}{10},\quad P(X_2=0|X_1=0)=\frac{7}{10}$$

因此，通过条件概率求出联合分布律（如 $P(X_1=1,X_2=1)=P(X_1=1)P(X_2=1|X_1=1)$）

**将每一列的概率相加就得到 $X_2$ 的边际分布律**：

| $X_1\setminus X_2$ | 0 | 1 | $P(X_1=i)$ |
|---|---|---|---|
| 0 | 7/16 | 3/16 | 5/8 |
| 1 | 3/16 | 3/16 | 3/8 |
| $P(X_2=j)$ | 5/8 | 3/8 | |

（2）根据求得的分布律为：

| $X_1\setminus X_2$ | 0 | 1 | $P(X_1=i)$ |
|---|---|---|---|
| 0 | 7/16 0 | 3/16 0 | 5/8 0 |
| 1 | 3/16 0 | 3/16 1 | 3/8 1 |
| $P(X_2=j)$ | 5/8 0 | 3/8 1 | |

$\to$ $E(X_1X_2)=1\times\frac{3}{16}=\frac{3}{16}$（见标红） $E(X_1)=E(X_2)=1\times\frac{3}{8}=\frac{3}{8}$

$\to$ $E(X_1^2)=E(X_2^2)=1\times\frac{3}{8}=\frac{3}{8}$（见标蓝） $\to$ $D(X_1)=D(X_2)=\frac{3}{8}-\left(\frac{3}{8}\right)^2=\frac{15}{64}$

$\to$ $\mathrm{Cov}(X_1,X_2)=\frac{3}{16}-\left(\frac{3}{8}\right)^2=\frac{3}{64}$ $\to$ $\rho_{X_1X_2}=\frac{\mathrm{Cov}(X_1,X_2)}{\sqrt{D(X_1)D(X_2)}}=\frac{3/64}{15/64}=\boxed{1/5}$

**例 2** （17—18 **春夏**）设 $(X,Y)$ 的联合分布律如下表所示.已知 $E(X)=0.6$， $E(Y)=0$.

| $X\setminus Y$ | $-1$ | 0 | 1 |
|---|---|---|---|
| 0 | 0.1 | $a_2$ | $a_3$ |
| 1 | $a_4$ | $a_5$ | $a_6$ |

（1）若 $a_6=0.1$，且 $X$ 与 $Y$ 不相关，求 $(X,Y)$ 的联合分布律；

（2）若 $X$ 与 $Y$ 相互独立，求 $(X,Y)$ 的联合分布律.

**解** （1）由 $X$ 与 $Y$ 不相关 $\to$ $E(XY)=E(X)E(Y)=0$

$E(XY)=-1\times a_4+1\times a_6=a_6-a_4=0$（见下表标红） $\to$ $a_4=a_6=0.1$

| $X\setminus Y$ | $-1$ | 0 | 1 |
|---|---|---|---|
| 0 | 0.1 0 | $a_2$ 0 | $a_3$ 0 |
| 1 | $a_4$ $-1$ | $a_5$ 0 | $a_6$ 1 |

$\therefore$ $E(X)=a_4+a_5+a_6=0.6$ $\to$ $a_5=0.4$

$E(Y)=-(0.1+a_4)+a_3+a_6=0$ $\to$ $a_3=0.1$

$a_2+\cdots+a_6=0.9 \rightarrow a_2=0.1$，联合分布律如下：

| $X\setminus Y$ | $-1$ | $0$ | $1$ |
|---|---|---|---|
| 0 | 0.1 | 0.1 | 0.1 |
| 1 | 0.1 | 0.4 | 0.1 |

（2）由 $E(X)=1\cdot P(X=1)=0.6$，得 $P(X=1)=0.6$，$P(X=0)=0.4$：

| $X\setminus Y$ | $-1$ | $0$ | $1$ | |
|---|---|---|---|---|
| 0 | 0.1 | $a_2$ | $a_3$ | 0.4 |
| 1 | $a_4$ | $a_5$ | $a_6$ | 0.6 |
| | | | | |

由于相互独立，此时 $P(X=0,Y=-1)=P(X=0)P(Y=-1)=0.4P(Y=-1)=0.1$

$\therefore\ P(Y=-1)=0.25$

$\because\ E(Y)=P(Y=1)-P(Y=-1)=0 \rightarrow P(Y=1)=0.25$，$P(Y=0)=0.5$

由此边际分布律全部求得，可求出联合分布律如下表所示

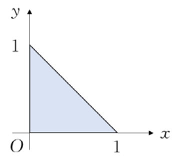
| $X\setminus Y$ | $-1$ | $0$ | $1$ | |
|---|---|---|---|---|
| 0 | 0.1 | 0.2 | 0.1 | 0.4 |
| 1 | 0.15 | 0.3 | 0.15 | 0.6 |
| | 0.25 | 0.5 | 0.25 | |

## 七　求解二元连续型非常见随机变量分布

### 1.题型简述与解法

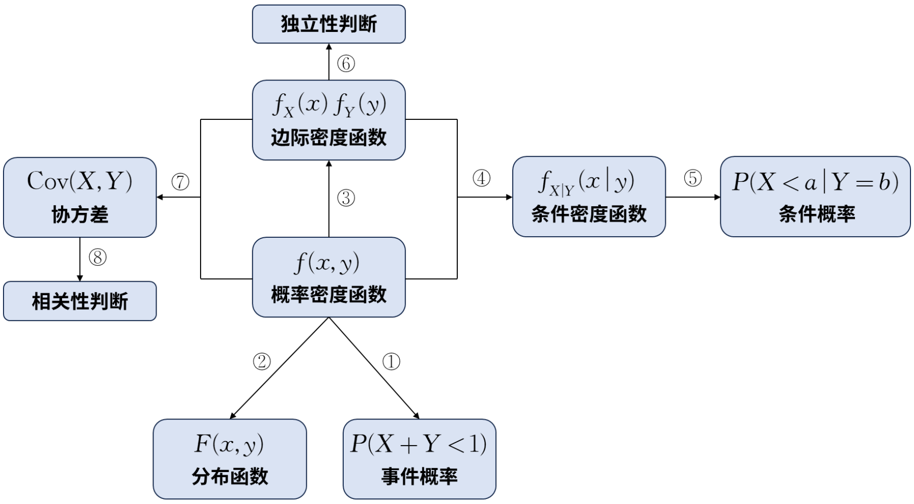
- 从二元随机变量 $(X,Y)$ 的联合密度函数出发，计算、判断多种特征
- 此类题型首先应**画出概率密度函数的非零区域**，如下面这个函数的非零区域就是一个等腰直角三角形

$$f(x,y)=\begin{cases}6x, x>0, y>0, x+y<1\\ 0,\quad 其他\end{cases} \rightarrow$$




有了这张图，我们在后续积分时思路会更加清晰，不易出错

- 计算的总体思路如下图所示：




① **求事件概率** 例 1 (1)

· 确认该事件对应的区域，与我们画好的非零区域取交集，得到要积分的区域，在该区域内积分即可

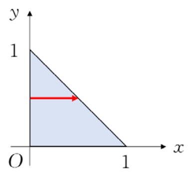
② **求分布函数的值** 例 2 (1)

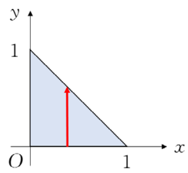
· 将分布函数转化为对应的事件，按①处理

③ **求边际密度函数** 例 1 (2) 例 2 (2)

· 代入边际密度函数的定义式进行积分

若 $f(x,y)$ 只在有限区域 $D$ 内不为 0，则只在不为 0 的 $x$ 或 $y$ 上积分即可




求 $f_Y(y)$：对 $x$ 从左往右积分　　求 $f_X(x)$：对 $y$ 从下往上积分

· 最终的结果要求 $x,y\in(-\infty,+\infty)$，因此**要写成分段函数的形式**（加上 0，其他）

④ **求条件密度函数** 例 1 (3) 例 2 (3)

· 求出边际密度后，代入条件密度函数的定义式即可

· 此时也要注意 $x$ 或 $y\in(-\infty,+\infty)$，写成分段函数的形式

⑤ **求条件概率** 例 1 (3) 例 2 (3)

· 将条件（“|”之后的那个）代入条件密度函数，在指定范围内对“|”前的变量作积分即可

⑥ **独立性判断**

· 检查是否 $f_X(x)f_Y(y)=f(x,y)$ 或 $F_X(x)F_Y(y)=F(x,y)$

⑦ **求协方差** 例 1 (4)

· 由 $f(x,y)$ 计算 $E(XY)$，由 $f_X(x),f_Y(y)$ 计算 $E(X),E(Y)$，然后代入公式计算

⑧ **相关性判断** 例 1 (4)

· 若协方差大于 0 则正相关，小于 0 则负相关，等于 0 则不相关

**2.历年考试典型例题**

**例 1** （19—20 春夏）设$(X,Y)$的联合密度函数

$$f(x,y)=\begin{cases}\dfrac{3y}{2},0<y<2x<2\\ 0,\quad 其他\end{cases}$$

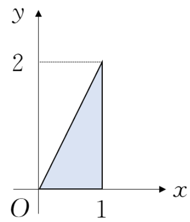
（1）求$P(\max(X,Y)<1)$；

（2）分别求$X,Y$的边际密度函数$f_X(x)$，$f_Y(y)$；

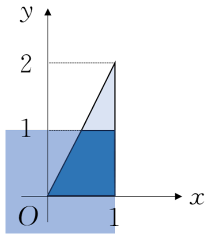
（3）$P(Y>0.5|X=0.5)$；

（4）判断$X$与$Y$是正相关，负相关，还是不相关？说明理由.

**解** 本题的非零区域如下图所示




（1）$P(\max(X,Y)<1)=P(X<1,Y<1)$，该区域与非零区域的交集为




对该区域进行积分：$P(\max(X,Y)<1)=\int_0^1 dy\int_{y/2}^1\frac{3y}{2}dx=\frac{1}{4}$

（2）① 求$f_X(x)$：仅$0<x<1$时，$f(x,y)$非 0，此时$f_X(x)=\int_0^{2x}\frac{3y}{2}dy=3x^2$

因此 $f_X(x)=\begin{cases}3x^2,\quad 0<x<1\\ 0,\quad 其他\end{cases}$

② 求$f_Y(y)$：仅$0<y<2$时，$f(x,y)$非 0，此时$f_Y(y)=\int_{y/2}^1\frac{3y}{2}dx=\frac{3y}{2}(1-\frac{y}{2})$

因此 $f_Y(y)=\begin{cases}\dfrac{3y}{2}(1-\dfrac{y}{2}),\quad 0<y<2\\ 0,\quad 其他\end{cases}$

（3）条件概率密度 $f_{Y|X}(y|x)=\dfrac{f(x,y)}{f_X(x)}=\begin{cases}\dfrac{y}{2x^2},0<y<2x<2\\ 0,\quad 其他\end{cases}$

$$\therefore\ P(Y>\frac{1}{2}|X=\frac{1}{2})=\int_{1/2}^{+\infty}f_{Y|X}(y|\frac{1}{2})\mathrm{d}y=\int_{1/2}^{1}2y\mathrm{d}y=\frac{3}{4}$$

（4）$E(XY)=\iint\limits_{D}xyf(x,y)\mathrm{d}x\mathrm{d}y=\int_0^1\mathrm{d}x\int_0^{2x}xy\frac{3y}{2}\mathrm{d}y=\frac{4}{5}$

$$E(X)=\int_{-\infty}^{+\infty}xf_X(x)\mathrm{d}x=\int_0^1 3x^3\mathrm{d}x=\frac{3}{4}$$

$$E(Y)=\int_{-\infty}^{+\infty}yf_Y(y)\mathrm{d}y=\int_0^2\frac{3y^2}{2}(1-\frac{y}{2})\mathrm{d}y=1$$

$$\therefore\ \mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)=\frac{4}{5}-\frac{3}{4}\times 1=\frac{1}{20}>0$$

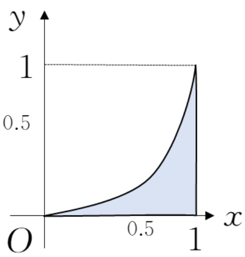
$\therefore$ $X$ 与 $Y$ 正相关

**例 2**　（17—18 **春夏　节选**）设 $(X,Y)$ 的联合密度函数 $f(x,y)=\begin{cases}4x, & 0<y<x^2, 0<x<1\\ 0, & 其他\end{cases}$：

（1）求 $(X,Y)$ 的联合分布函数值 $F(0.5,0.5)$；

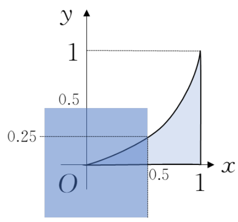
（2）分别求 $X$ 与 $Y$ 的边际概率密度函数 $f_X(x)$ 和 $f_Y(y)$，并判断 $X$ 与 $Y$ 是否相互独立；

**解**　本题的非零区域如下图所示：




（1）$F(0.5,0.5)=P(X\le 0.5,Y\le 0.5)$

$$=\int_0^{0.5}\mathrm{d}x\int_0^{x^2}4x\mathrm{d}y=\boxed{1/16}$$




（2）仅 $0<y<1$ 时，$f(x,y)$ 非 0，此时 $f_Y(y)=\int_{\sqrt{y}}^{1}4x\mathrm{d}x=2-2y$

因此　$f_Y(y)=\begin{cases}2-2y, & 0<y<1\\ 0, & 其他\end{cases}$

仅 $0<x<1$ 时，$f(x,y)$ 非 0，此时 $f_X(x)=\int_0^{x^2}4x\mathrm{d}y=4x^3$

因此　$f_X(x)=\begin{cases}4x^3, & 0<x<1\\ 0, & 其他\end{cases}$

· 显然 $f_X(x)f_Y(y)\neq f(x,y)$，因此 $X$ 与 $Y$ 不独立

# 第 4 讲　常见随机变量的概率分布

## 知识梳理

### 一　常见随机变量分布一览

1.常见一元离散分布

| 分布名称 | 符号 | 分布律 | 期望 | 方差 | 说明 |
| --- | --- | --- | --- | --- | --- |
| 0-1 分布 | $X \sim 0-1(p)$ | $P(X=0)=1-p \quad P(X=1)=p$ | $p$ | $p(1-p)$ | |
| 二项分布 | $X \sim B(n,p)$ | $P(X=k)=C_n^k p^k(1-p)^{n-k}$ | $np$ | $np(1-p)$ | 不直说 |
| 泊松分布 | $X \sim \pi(\lambda)$ | $P(X=k)=\dfrac{\lambda^k}{k!}\mathrm{e}^{-\lambda}$ | $\lambda$ | $\lambda$ | 记前几个 |

2.常见一元连续分布

| 分布名称 | 符号 | 密度函数 | 期望 | 方差 | |
| --- | --- | --- | --- | --- | --- |
| 均匀分布 | $X \sim U(a,b)$ | $f(x)=\begin{cases}1/(b-a), & x\in(a,b)\\ 0, & 其它\end{cases}$ | $\dfrac{a+b}{2}$ | $\dfrac{(b-a)^2}{12}$ | —— |
| 正态分布 | $X \sim N(\mu,\sigma^2)$ | $f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}},-\infty<x<+\infty$ | $\mu$ | $\sigma^2$ | 化成标准 |
| 指数分布 | $X \sim E(\lambda)$ | $f(x)=\begin{cases}\lambda\mathrm{e}^{-\lambda x}, x>0\\ 0, \quad x\le 0\end{cases}$ | $\dfrac{1}{\lambda}$ | $\dfrac{1}{\lambda^2}$ | 无记忆性<br>记住分布函数 |

3.常见二元连续分布

| 分布名称 | 符号 或 密度函数 | 备注 |
| --- | --- | --- |
| 二元均匀分布 | $f(x,y)=\begin{cases}c, & (x,y)\in D\\ 0, & 其它\end{cases}$ | $c$ 为区域 $D$ 面积的倒数 |
| 二元正态分布 | $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ | $\rho$ 就是 $X$ 和 $Y$ 的相关系数 |

### 二　随机变量数字特征的性质

1.数学期望的性质

**数学期望的线性组合（Xi 无需相互独立）**

$$E(c_0+\sum_{i=1}^{n}c_iX_i)=c_0+\sum_{i=1}^{n}c_iE(X_i)$$

**数学期望的乘积（Xi 必须相互独立）**

$$E(\prod_{i=1}^{n}X_i)=\prod_{i=1}^{n}E(X_i)$$

2.方差的性质

**方差的线性组合（Xi 无需相互独立）**

$$\mathrm{Var}(\sum_{i=1}^{n}c_iX_i+c_0)=\sum_{i=1}^{n}c_i^2\mathrm{Var}(X_i)+2\sum_{1\le i<j\le n}\mathrm{Cov}(X_i,X_j)$$

· 当 $X_i$ 两两独立时，协方差一项为 0

3. **协方差的性质**

| 协方差的性质 1 | 协方差的性质 2 | 协方差的性质 3 |
|---|---|---|
| $\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$ | $\mathrm{Cov}(X,X)=\mathrm{Var}(X)$ | $\mathrm{Cov}(aX,bY)=ab\mathrm{Cov}(X,Y)$ |

| 协方差的性质 1 | 协方差的性质 2 |
|---|---|
| $\mathrm{Cov}(X_1+X_2,Y)=\mathrm{Cov}(X_1,Y)+\mathrm{Cov}(X_2,Y)$ | 若 $X$ 和 $Y$ 相互独立，则 $\mathrm{Cov}(X,Y)=0$ |

## 三　正态分布的重要性质

1. **一元正态分布的性质**

① **标准正态分布 $Z\sim N(0,1)$ 的分布函数 $\Phi(x)$**

| 标准正态分布函数 | 标准正态分布函数的性质 | 标准正态分布函数的性质 |
|---|---|---|
| $\Phi(x)=P(Z\le x)$ | $\Phi(-x)=1-\Phi(x)$ | $P(a<Z<b)=\Phi(b)-\Phi(a)$ |

· $\Phi(x)$ 的值通过手册查阅，**考试会提供需要用到的 $\Phi(x)$ 值**

② **正态分布的线性函数依然是正态分布** → 可经线性函数转化为标准正态分布

| 正态分布 → 标准正态分布的转化 |
|---|
| $X\sim N(\mu,\sigma^2)\ \rightarrow\ Z=\dfrac{X-\mu}{\sigma}\sim N(0,1)$ |

· 因此求解正态分布问题，往往需要转化为标准正态分布处理

2. **二元正态分布的性质**

① **二元正态分布的 $X$ 和 $Y$ 各自依然遵循正态分布**

| 二元正态分布的分解 |
|---|
| $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)\ \rightarrow\ X\sim N(\mu_1,\sigma_1^2),\ \ Y\sim N(\mu_2,\sigma_2^2)$ |

② **$X$ 和 $Y$ 的线性函数 $aX+bY+c$ 依然是正态分布**

· 此时新正态变量的 $\mu$ 和 $\sigma^2$ 分别是 $aX+bY+c$ 的均值和方差

③ **对于两个正态变量，不相关和独立是等价的**

· 这使得判断正态变量是否独立时，只需计算协方差即可

## 四　其他分布的重要性质

1. **指数分布的性质**

| 指数分布的无记忆性 |
|---|
| $P(X>a+b\mid X>a)=P(X>b)$ |

# 题型解析

## 八　求解常见随机变量分布

### 1.题型简述与解法

- 通常以填空题的形式出现（偶尔也会有大题），考查常见随机变量分布的事件概率、期望、方差等
- 解题关键在于记住这些分布的期望、方差以及出众的特征，可以大幅减少时间和脑力消耗
- 与一般的分布不同，常见随机变量分布往往是通过 $D(X)$ 反求 $E(X^2)$，需要注意

① **0-1 分布**　例 1

- 该分布比较简单，只需要能反应过来它只有两种取值就可以了

② **二项分布**

- 该分布通常不会在题干中明确指出，往往通过“对 xx 进行独立考察”等字眼暗示
- 此时独立考察的次数就是参数 $n$， $p$ 则要根据题意自己求
- 通常会求事件概率，枚举复杂时可尝试求逆事件

③ **泊松分布**

- 该分布通常会求事件概率，枚举复杂时可尝试求逆事件
- 最好记住前 3 个取值的分布律，加快解题速度：

$$P(X=0)=\mathrm{e}^{-\lambda},\quad P(X=1)=\lambda\mathrm{e}^{-\lambda},\quad P(X=2)=\frac{\lambda^2}{2}\mathrm{e}^{-\lambda}$$

④ **均匀分布**

- 最好记住均值和方差，加快解题速度

⑤ **正态分布**

- 求事件概率要将正态分布 $X$ 转换为标准正态分布 $Z$，注意要对事件作一样的转换

$$P(x_1<X<x_2)=P(\frac{x_1-\mu}{\sigma}<Z<\frac{x_2-\mu}{\sigma})$$

- 如果试卷开头找不到需要的 $\Phi$，那你肯定算错了

⑥ **指数分布**

- 记住分布函数以及事件概率

$$F(x)=P(X\le x)=1-\mathrm{e}^{-\lambda x}\qquad P(X>x)=\mathrm{e}^{-\lambda x}$$

- 记住指数分布的无记忆性，可以将条件概率直接转为更简单的概率

⑦ **二元均匀分布**

- 求事件概率计算面积乘上概率密度函数即可，尤其是区域为圆的时候

⑧ **二元正态分布**

- 一般考察正态变量之间的相关性，新正态变量的期望与方差等

## 2.历年考试典型例题

**例 1** （**19－20 春夏**）设 $X$ 与 $Y$ 均服从 0－1 分布，$P(X=1)=\frac{3}{4}$，$P(X=1,Y=1)=\frac{1}{4}$

（1）若 $X$ 与 $Y$ 独立，则 $P(X=0,Y=1)=$______________；

（2）若 $X$ 与 $Y$ 的协方差为 $-\frac{1}{16}$，则 $P(X=0,Y=1)=$_____________.

**解** 因为 $X$ 与 $Y$ 均服从 0－1 分布，可直接打表，并得到 $P(X=1,Y=0)=1/2$

| X\Y | 0 | 1 | |
|---|---|---|---|
| 0 | | ? | 1/4 |
| 1 | 1/2 | 1/4 | 3/4 |
| | 1−p | p | |

（1）若 $X$ 与 $Y$ 独立，则 $P(X=1,Y=0)=P(X=1)P(Y=0)=1/2 \rightarrow P(Y=0)=2/3$

因此 $P(Y=1)=1/3$，$P(X=0,Y=1)=P(X=0)P(Y=1)=\boxed{1/12}$

（2）若 $X$ 与 $Y$ 的协方差为 $-\frac{1}{16}$，由 $E(X)=3/4$，$E(Y)=p$，$E(XY)=P(X=1,Y=1)=1/4$

即 $\mathrm{Cov}(X,Y)=\frac{1}{4}-\frac{3}{4}p=-\frac{1}{16}$，解得 $p=\frac{5}{12}$，因此 $P(X=1,Y=0)=\frac{5}{12}-\frac{1}{4}=\boxed{1/6}$

**例 2** （**14－15 春夏 节选**）某人喜欢长跑，基本上每天跑 10 公里，假设他跑 10 公里所花时间（分钟）为 $Y=40+20X$，其中 $X$ 的概率密度为 $f(x)=\begin{cases}2(1-x),0<x<1\\ 0, \quad 其他\end{cases}$.

（3）若一周跑 6 次，每天用时是相互独立的，求至少有 5 次用时少于 50 分钟的概率；

**解** 原题的前两问已经求出 $P(Y<50)$ 的概率，此处不再赘述过程，结果为 0.75

由题意可得这是一个二项分布，$n=6$，$p=0.75$

设少于 50 分钟的次数为 $Z$，则 $P(Z=k)=C_6^k(\frac{3}{4})^k(\frac{1}{4})^{6-k}$ （$k=0,\cdots,6$）

因此题目要求的 $P(Z\geq 5)=P(Z=5)+P(Z=6)=C_6^5(\frac{3}{4})^5\frac{1}{4}+C_6^6(\frac{3}{4})^6=\boxed{0.534}$

**例 3** 某公交车站单位时间内等车的人数 $X$ 服从泊松分布 $\pi(\lambda)$，$\lambda>0$

（1）（**18－19 春夏**）若 $E(X^2)=2\mathrm{Var}(X)$，则 $\lambda=$______________.

（2）（**17－18 秋冬**）$X$ 的分布函数值 $F(2)=$_______.

（3）（**15－16 秋冬**）若已知有人在等车，则至少有 2 人等车的概率为____________.

**解** （1）由 $X\sim\pi(\lambda)$，得 $E(X)=\mathrm{Var}(X)=\lambda$，则 $E(X^2)=D(X)+[E(X)]^2=\lambda+\lambda^2$

因此 $E(X^2)=2\mathrm{Var}(X)\ \to\ \lambda+\lambda^2=2\lambda$，$\lambda>0\ \to\ \boxed{\lambda=1}$

（2）$F(2)=P(X\le 2)=P(X=0)+P(X=1)+P(X=2)=(1+\lambda+\dfrac{\lambda^2}{2})e^{-\lambda}=\boxed{2.5e^{-1}}$

（3）已知有人在等车 $\to$ $X>0$，因此所求概率为

$$P(X\ge 2\mid X>0)=\frac{P(X\ge 2,X>0)}{P(X>0)}=\frac{1-P(X=1)-P(X=0)}{1-P(X=0)}=\frac{1-e^{-1}-e^{-1}}{1-e^{-1}}=\boxed{\frac{1-2e^{-1}}{1-e^{-1}}}$$

**例 4** （1）（17—18 **春夏**）设 $X\sim U(1,c)$（均匀分布），$E(X)=2$，则 $c=$____，$\mathrm{Var}(X)=$_____.

（2）（18—19 **秋冬**）设 $(X,Y)$ 在区域 $\{(x,y):0\le x\le 1,0\le y\le 1\}$ 上均匀分布，则

$P(X^2+Y^2\le 1)=$___________.

**解** （1）由 $E(X)=(1+c)/2=2$，解得 $c=3$，$\mathrm{Var}(X)=(3-1)^2/12=1/3$

（2）总的区域为 $1\times 1$ 的正方形，要求的区域为四分之一圆，因此概率为 $(\pi/4)/1=\pi/4$

**例 5** （16—17 **秋冬**）设 $(X,Y)\sim N(1,0,1,4,0.76)$，则 $P(X<Y-0.4)=$_________；当 $c=$_______ 时，$cX-Y$ 与 $X-Y$ 相互独立.

**解** 由 $(X,Y)\sim N(1,0,1,4,0.76)$，得 $X\sim N(1,1)$，$Y\sim N(0,4)$，$X$ 与 $Y$ 的相关系数为 0.76（注意参数顺序是 $\mu_1,\mu_2,\sigma_1^2,\sigma_2^2$），则

$$E(X-Y)=1-0=1$$

$$D(X-Y)=D(X)+D(Y)-2\rho\sqrt{D(X)D(Y)}=1.96=1.4^2$$

因此 $X-Y\sim N(1,1.4^2)$：

$$P(X<Y-0.4)=P(X-Y<-0.4)=P(\frac{X-Y-1}{1.4}<\frac{-0.4-1}{1.4})=\Phi(-1)=1-\Phi(1)=\boxed{0.16}$$

由于 $\mathrm{Cov}(cX-Y,X-Y)=cD(X)-(c+1)\mathrm{Cov}(X,Y)+D(Y)=c-1.52(c+1)+4$

由正态分布的性质，当 $c-1.52(c+1)+4=0$ 时，相互独立 $\to$ $\boxed{c=4.77}$

**例 6** （14—15 **春夏**）某小店开门到首个顾客到达所用的时间 $X$（单位：分钟）的概率密度函数为

$$f(x)=\begin{cases}0.2e^{-0.2x}, & x>0\\ 0, & x\le 0\end{cases}$$

则在开门后 10 分钟内首个顾客到达的概率为___________，若开门后 5 分钟内没有顾客到达，则在接下来的 5 分钟内仍没有顾客到达的概率为____________.

**解** 由题意 $X\sim E(0.2)$，即 $X$ 服从 $\lambda=0.2$ 的指数分布，因此第一空 $P(X<10)=1-e^{-0.2\times 10}=1-e^{-2}$

第二空 $P(X>5+5\mid X>5)=P(X>5)=1-e^{-0.2\times 5}=1-e^{-1}$

# 第 5 讲　随机变量函数的概率分布

## 知识梳理

### 一　求解随机变量函数概率分布的基本思路

1.→ **离散**

- 包括一维连续 → 多维离散，多维离散 → 多维离散，多维连续 → 一维离散等

① **列出因变量的所有可能取值(组合)，并与自变量的取值一一对应**

- 如果自变量是连续变量，则对应的是取值范围

② **求因变量各个取值(组合)的概率：等于对应的自变量取值(组合)的概率**

- 如果一个因变量的取值对应多个自变量的取值，则其概率为所有对应的自变量取值概率之和
- 如果一个因变量的取值没有对应的自变量取值，则该取值对应的概率为 0

③ **列表，得到分布律**

2.→ **连续**

- 包括一维连续 → 一维连续，多维连续 → 一维连续，多维混合 → 一维连续等

① **写出因变量的分布函数，用取值范围的概率表示**

② **将因变量用自变量表示，转化为自变量的取值范围**

- 常用的等价事件转化

**最大值分布函数**

$$\boxed{P(\max(X_1,\cdots,X_n)\le y)=P(X_1\le y,\cdots,X_n\le y)}$$

**最小值分布函数**

$$\boxed{P(\min(X_1,\cdots,X_n)\le y)=1-P(X_1> y,\cdots,X_n> y)}$$

③ **用自变量的分布函数或密度函数算出因变量的分布函数，若有需要则求导得到密度函数**

- 如果自变量是离散与连续组合出的混合变量，则计算分布函数时要根据离散变量的取值分类讨论

3.→ **混合**

- 包括一维连续 → 一维连续，多维连续 → 一维连续，多维混合 → 一维连续等

① **首先求出离散因变量的可能取值**

② 再写出联合分布函数，根据离散因变量的可能取值分段讨论，转化为用自变量表示的事件

③ 求出概率，得到分布函数

# 题型解析

## 九　随机变量的函数分布求解

### 1.题型简述与解法

- 已知某随机变量的分布，求该随机变量的函数的分布函数、密度函数、分布律，判断独立等
- 求解方法见知识梳理
- 一句话总结就是以分布函数或分布律为核心，将因变量的事件转化回自变量的事件

### 2.历年考试典型例题

① **一维连续 → 多维离散**

**例 1**　（16—17 春夏）设 $X$ 的密度函数 $f(x)=\begin{cases}0.5x, & 0<x<2\\ 0, & \text{其他}\end{cases}$，令 $Y_1=\begin{cases}1, X>1\\ 0, X\le 1\end{cases}$，$Y_2=\begin{cases}1, X\le 1.5\\ 0, X>1.5\end{cases}$，

求 $(Y_1,Y_2)$ 的联合分布律；

**解**　$Y_1$ 和 $Y_2$ 的可能取值都为 0 或 1，因此总共有四种情况：

- $P(Y_1=0,Y_2=0)=P(X\le 1,X>1.5)=0$
- $P(Y_1=0,Y_2=1)=P(X\le 1,X\le 1.5)=P(X\le 1)=1/4$
- $P(Y_1=1,Y_2=0)=P(X>1,X>1.5)=P(X>1.5)=7/16$
- $P(Y_1=1,Y_2=1)=P(X>1,X\le 1.5)=P(1<X\le 1.5)=5/16$

| $Y_1\backslash Y_2$ | 0 | 1 |
|---|---|---|
| 0 | 0 | 1/4 |
| 1 | 7/16 | 5/16 |

② **多维离散 → 多维离散**

**例 2**　（20—21 秋冬）设 $(X,Y)$ 的联合分布律如下表所示，令 $Z=\min(X,Y)$，求 $(X,Z)$ 的联合分布律。

| $X\backslash Y$ | 0 | 1 | 2 |
|---|---|---|---|
| 0 | $a$ | 0 | $b$ |
| 1 | $b$ | $a$ | $b$ |
| 2 | 0 | $b$ | $2a$ |

**解**　由于 $X$ 和 $Z$ 的可能取值均为 0、1 和 2，所以一共有 9 种可能的取值组合

- 可以得到每个 $(X,Y)$ 对应的 $Z$ 值如下表所示（红色），由此可以得到剩余 $(X,Z)$ 的概率

| $X\backslash Y$ | 0 | 1 | 2 |
|---|---|---|---|
| 0 | $a$ 0 | 0 0 | $b$ 0 |
| 1 | $b$ 0 | $a$ 1 | $b$ 1 |
| 2 | 0 0 | $b$ 1 | $2a$ 2 |

- $P(X=0,Z=0)=P(X=0,Y=0)+P(X=0,Y=1)+P(X=0,Y=2)=a+b$

$P(X=0,Z=1)=P(X=0,Z=2)=0$

$P(X=1,Z=0)=P(X=1,Y=0)=b$

$P(X=1,Z=1)=P(X=1,Y=0)+P(X=1,Y=1)=a+b$

$P(X=2,Z=0)=P(X=2,Y=0)=0$

$P(X=2,Z=1)=P(X=1,Y=1)=b$

$P(X=2,Z=2)=P(X=2,Y=2)=2a$

· 分布律如下表所示

| $X\setminus Z$ | 0 | 1 | 2 |
|---|---|---|---|
| 0 | $a+b$ | 0 | 0 |
| 1 | $b$ | $a+b$ | 0 |
| 2 | 0 | $b$ | $2a$ |

③ **一维连续 → 一维连续**

**例 3** （14—15 **春夏**）某人喜欢长跑，基本上每天跑 10 公里，假设他跑 10 公里所花时间（分钟）为 $Y=40+20X$，其中 $X$ 的概率密度为 $f(x)=\begin{cases}2(1-x),0<x<1\\ 0, \quad 其他\end{cases}$.求 $Y$ 的概率密度 $f_Y(y)$

**解** 思路：$f(x)\to F(x)\to P(X\le x)\Leftrightarrow P(Y\le y)\to F_Y(y)\to f_Y(y)$

$$F_Y(y)=P(Y\le y)=P(40+20X\le y)=P(X\le\frac{y-40}{20})=F_X(\frac{y-40}{20})$$

因此 $f_Y(y)=F_Y'(y)=F_X'(\frac{y-40}{20})\frac{1}{20}=f(\frac{y-40}{20})\frac{1}{20}$，即

$$f_Y(y)=\begin{cases}2(1-\frac{y-40}{20})\cdot\frac{1}{20},0<\frac{y-40}{20}<1\\ 0, \quad 其他\end{cases}\to f_Y(y)=\begin{cases}0.3-\frac{y}{200},40<y<60\\ 0, \quad 其他\end{cases}$$

· 教材上有一个直接得到密度函数的公式，本题可以使用，但是要注意本题能使用的原因是 $Y$ 是严格单调的，而考试更喜欢考不严格单调的，因此不用去记那个公式

④ **多维连续 → 一维连续**

**例 4** （15—16 **秋冬**）设随机变量 $X$ 的概率密度为 $f(x)=\begin{cases}1-x,0<x<1\\0.5, \quad 1\le x<2\\0, \quad 其他\end{cases}$，若对 $X$ 独立观察 3 次，结果记为 $X_1,X_2,X_3$，设 $Y=\max\{X_1,X_2,X_3\}$，求 $Y$ 的概率密度 $f_Y(y)$.

**解** 本题类型为 **多维连续 → 一维连续**

$F_Y(y)=P(Y\le y)=P(\max\{X_1,X_2,X_3\}\le y)=P(X_1\le y,X_2\le y,X_3\le y)$

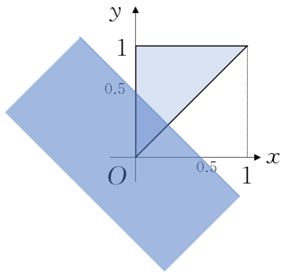
$\because$ $X_1,X_2,X_3$ 相互独立

$\therefore$ $F_Y(y)=P(X_1\le y)P(X_2\le y)P(X_3\le y)=[P(X<y)]^3=F^3(y)$

$\therefore$ $f_Y(y)$

**例 5** （18—19 **春夏**）设 $(X,Y)$ 的联合密度函数 $f(x,y)=\begin{cases}8xy, 0<x<y<1\\ 0, \quad 其他\end{cases}$，求 $Z=X+Y$ 的概率密度函数 $f_Z(z)$.

**解** $z$ 的分布函数 $F(z)=P(Z\le z)=P(X+Y\le z)$，如图所示




当 $z<0$ 时，$P(X+Y\le z)=0$

当 $0\le z<1$ 时，$P(X+Y\le z)=\int_0^{z/2}\mathrm{d}x\int_x^{z-x}8xy\mathrm{d}y=\dfrac{z^4}{6}$

当 $1\le z<2$ 时，$P(X+Y\le z)=1-\int_{z/2}^1\mathrm{d}y\int_{z-y}^{y}8xy\mathrm{d}x=1-\dfrac{8}{3}z+2z^2-\dfrac{1}{6}z^4$

当 $z\ge 2$ 时，$P(X+Y\le z)=1$

$$\therefore\ F_Z(z)=\begin{cases}0, & z<0\\ \dfrac{z^4}{6}, & 0\le z<1\\ 1-\dfrac{8}{3}z+2z^2-\dfrac{1}{6}z^4, & 1\le z<2\\ 1, & z\ge 2\end{cases}\ \rightarrow\ f_Z(z)=\begin{cases}\dfrac{2z^3}{3}, & 0\le z<1\\ -\dfrac{8}{3}+4z-\dfrac{2}{3}z^3, & 1\le z<2\\ 0, & 其他\end{cases}$$

⑤ **多维混合 → 一维连续**

**例 6** （18—19 **秋冬**）将一枚硬币独立抛 2 次，$X$ 表示正面朝上的次数；$Y$ 服从 $(0,2)$ 区间上的均匀分布，设 $X$ 与 $Y$ 相互独立，$M=\max(X,Y)$，$Z=X+Y$，求 $M,Z$ 的分布函数

**解**
$$F_M(m)=P(M\le m)=P(\max(X,Y)\le m)=P(X\le m,Y\le m)$$
$$=P(X\le m)P(Y\le m)=F_X(m)F_Y(m)$$
$$F_Z(z)=P(Z\le z)=P(X+Y\le z)$$
$$=P(X=0,Y\le z)+P(X=1,Y\le z-1)+P(X=2,Y\le z-2)$$

⑥ **多维连续 → 多维混合 （包含多维连续 → 一维离散）**

**例 7** （19—20 **秋冬**）设 $(X,Y)$ 的联合密度函数 $f(x)=\begin{cases}0.75, & y^2<x<1\\ 0, & 其他\end{cases}$，令 $Z=\begin{cases}0, & 0\le Y<\sqrt{X}<1\\ 1, & 其他\end{cases}$，判断 $X$ 与 $Z$ 是否独立？说明理由.

**解** 由于 $X$ 连续，$Z$ 离散，因此只能通过分布函数判断

$$F_Z(z)=\begin{cases}0, & z<0\\ 1/2, & 0\le z<1\\ 1, & z\ge 1\end{cases}$$

$$F(x,z)=P(Z\le z,X\le x)=\begin{cases}0, & z<0\\ P(Z=0,X\le x), & 0\le z<1\\ P(X\le x), & z\ge 1\end{cases}$$

$\because$ $P(Z=0,X\le x)=P(0\le Y<\sqrt{X}<1,X\le x)$，由密度函数关于 $x$ 轴对称，可以得到

$P(0\le Y<\sqrt{X}<1,X\le x)=\dfrac{1}{2}P(X\le x)=\dfrac{1}{2}F_X(x)$

$$F(x,z)=P(Z\le z,X\le x)=\begin{cases}0(\cdot F_X(x)), & z<0\\ \dfrac{1}{2}F_X(x), & 0\le z<1\\ (1\cdot)F_X(x), & z\ge 1\end{cases}=F_Z(z)F_X(x)$$

$\therefore$ $X$ 与 $Z$ 相互独立

# 第 6 讲　大数定律与中心极限定理

## 知识梳理

### 一　切比雪夫不等式

· 设随机变量 $X$ 的数学期望和方差存在，分别记为 $\mu,\sigma^2$，则对任意的 $\varepsilon>0$，有

**切比雪夫不等式**

$$P(|X-\mu|\geq\varepsilon)\leq\frac{\sigma^2}{\varepsilon^2}$$

· 由于近十年只考过一次，所以没有收录进题型中

### 二　辛钦大数定律

1. **大数定律基本内容**

**辛钦大数定律**

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i\xrightarrow{P}E(X)\ (n\to\infty)$$

· $X_i$ 为独立同分布的随机变量序列，在下一讲会学到 $\overline{X}$ 为样本均值

2. **大数定律的推论**

**辛钦大数定律推论 1**

$$\frac{1}{n}\sum_{i=1}^{n}h(X_i)\xrightarrow{P}E[h(X)]\ (n\to\infty)$$

**辛钦大数定律推论 2**

$$f(\overline{X})\xrightarrow{P}f[E(X)]$$

### 三　中心极限定理

**中心极限定理**

$n$ 足够大时，$\sum_{i=1}^{n}X_i\sim N(n\mu,n\sigma^2)$　或　$\frac{1}{n}\sum_{i=1}^{n}X_i\sim N(\mu,\frac{\sigma^2}{n})$

· $X_i$ 为独立同分布的随机变量序列

· 推论：$n$ 比较大的二项分布（可视为 $n$ 个 0–1 分布变量之和）可以近似为正态分布

# 题型解析

## 十　求解依概率收敛

### 1.题型简述与解法

· 题干中出现“依概率收敛”或“$\xrightarrow{P}$”符号

· 首先求出所需的数学期望，然后代入辛钦大数定律（或两个推论）

### 2.历年考试典型例题

**例 1**　（19－20 **秋冬**）设随机变量 $X$ 的概率密度函数

$$f(x)=\begin{cases}\dfrac{x^2}{9}, & 0<x<3\\ 0, & 其他\end{cases}$$

对 $X$ 独立重复观测 $n$ 次，结果记为 $X_1,\cdots,X_n$ .

（3）当 $n\to+\infty$ 时，$\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^{-2}\mathrm{e}^{-X_i}$ 依概率收敛到何值?

**解**　· 求出 $E(X^{-2}\mathrm{e}^{-X})$

$$E(X^{-2}\mathrm{e}^{-X})=\int_0^3\frac{x^2}{9}\cdot x^{-2}\mathrm{e}^{-x}\mathrm{d}x=\frac{1}{9}(1-\mathrm{e}^{-3})$$

· 代入辛钦大数定律

$$当\ n\to+\infty\ 时，\frac{1}{n}\sum_{i=1}^{n}X_i^{-2}\mathrm{e}^{-X_i}\xrightarrow{P}E(X^{-2}\mathrm{e}^{-X})=\frac{1}{9}(1-\mathrm{e}^{-3})$$

**例 2**　（14－15 **春夏**）设总体 $X\sim N(\mu,\sigma^2)$，$X_1,\cdots,X_n$ 为来自 $X$ 的简单随机样本，当 $n\to\infty$ 时，$\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^2$ 依概率收敛到________.

**解**　对于常见分布，应通过 $E(X)$ 和 $D(X)$ 反推 $E(X^2)$，因此 $E(X^2)=\mu^2+\sigma^2$

**例 3**　设总体 $X$ 的分布律为 $P(X=-1)=\dfrac{\theta}{3}$，$P(X=0)=\dfrac{2\theta}{3}$，$P(X=1)=\dfrac{2(1-\theta)}{3}$，$P(X=2)=\dfrac{1-\theta}{3}$，未知参数 $\theta\in(0,1)$，$X_1,\cdots,X_n$ 是 $X$ 的简单随机样本，$\overline{X}$ 是样本均值，当 $n\to+\infty$ 时，$(\overline{X}-\dfrac{4}{3})^2\xrightarrow{P}$__________.

**解**　$E(X)=\dfrac{4}{3}-\dfrac{5}{3}\theta$，根据大数定律的推论，$(\overline{X}-\dfrac{4}{3})^2\xrightarrow{P}[E(X)-\dfrac{4}{3}]^2=(\dfrac{4}{3}-\dfrac{5}{3}\theta-\dfrac{4}{3})^2=\dfrac{25}{9}\theta^2$

## 十一　求解近似分布

### 1.题型简述与解法

- 题干中出现“近似”或约等于符号，且含有一些比较大的数字
- 求出对应分布的期望和方差，代入中心极限定理后，用正态分布的性质求解

**例 1**　（16－27 春夏）设 $X$ 的密度函数 $f(x)=\begin{cases}0.5x, 0<x<2\\ 0, \quad 其他\end{cases}$

（3）若对 $X$ 独立重复观察 72 次，结果记为 $X_1,\cdots,X_{72}$，求 $P(\frac{1}{72}\sum_{i=1}^{72}X_i>\frac{25}{18})$ 的近似值.

**解**　$E(X)=\int_0^2\frac{1}{2}x^2\mathrm{d}x=\frac{4}{3}$，$D(X)=E(X^2)-[E(X)]^2=\int_0^2\frac{1}{2}x^3\mathrm{d}x-\frac{16}{9}=2-\frac{16}{9}=\frac{2}{9}$

$\therefore$ 由中心极限定理，$\frac{1}{72}\sum_{i=1}^{72}X_i\sim N(\frac{4}{3},\frac{1}{18^2})$

$$\therefore\ P(\frac{1}{72}\sum_{i=1}^{72}X_i>\frac{25}{18})=P(\frac{\frac{1}{72}\sum_{i=1}^{72}X_i-\frac{4}{3}}{1/18}>\frac{\frac{25}{18}-\frac{4}{3}}{1/18})=P(Z>1\frac{\frac{1}{72}\sum_{i=1}^{72}X_i-\frac{4}{3}}{1/18}>1)\approx1-\Phi(1)\approx1.6$$

**例 2**　（19－20 春夏）设随机变量 $X$ 的概率密度函数

$$f(x)=\begin{cases}x^2, 0<x<1\\ \frac{2}{3},\ 1<x<2\\ 0,\quad 其他\end{cases}$$

对 $X$ 独立重复观测 $n$ 次，结果记为 $X_1,\cdots,X_n$.

（4）若 $n=450$，$Z$ 表示 450 次观测中 $\{X_i<1\}$ 出现的次数，求 $P(Z>160)$ 的近似值.

**解**　注意这里并不是求 $X$ 的近似概率，而是 $Z$ 的

由题意，$P(X<1)=\int_0^1x^2\mathrm{d}x=\frac{1}{3}$，因此 $Z\sim B(450,\frac{1}{3})$，则 $E(Z)=150$，$D(Z)=100$

因此 $Z$ 可以近似为 $N(150,100)$

$$\therefore\ P(Z>160)=P(\frac{Z-150}{10}>\frac{160-150}{10})\approx1-\Phi(1)\approx0.16$$

# 第 7 讲　统计量与抽样分布

## 知识梳理

· 事先说明：数理统计部分中的 $X_1,\cdots,X_n$ 都是独立同分布的随机变量

### 一　常见统计量

| 样本均值 | 样本方差 |
|---|---|
| $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ | $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ |

| 样本标准差 | 样本 $k$ 阶矩 | 样本 $k$ 阶中心矩 |
|---|---|---|
| $S=\sqrt{S^2}$ | $A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$ | $B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k$ |

· 比较重要的是样本均值和样本方差

**1.样本均值的性质**

| 样本均值的期望 | 样本均值的方差 |
|---|---|
| $E(\overline{X})=E(X)$ | $D(\overline{X})=\frac{D(X)}{n}$ |

**2.样本方差的性质**

| 样本方差的期望 |
|---|
| $E(S^2)=D(X)$ |

### 二　三个抽样分布

**1. $\chi^2$ 分布**

**$\chi^2$ 分布**

$$X_i\sim N(0,1)\to \boxed{Y=\sum_{i=1}^{n}X_i^2\sim\chi^2(n)}$$

① 期望和方差

| $\chi^2$ 分布的期望 | $\chi^2$ 分布的方差 |
|---|---|
| $E(Y)=n$ | $D(Y)=2n$ |

② 性质

· 若 $Y_1\sim\chi^2(m)$，$Y_2\sim\chi^2(n)$ 且相互独立，则 $Y_1+Y_2\sim\chi^2(m+n)$

2. $t$ 分布

**$t$ 分布**

$X \sim N(0,1)$，$Y \sim \chi^2(n)$ 且相互独立 $\rightarrow$ $\boxed{t = \dfrac{X}{\sqrt{Y/n}} \sim t(n)}$

2. $F$ 分布

**$t$ 分布**

$Y_1 \sim \chi^2(n_1)$，$Y_2 \sim \chi^2(n_2)$ 且相互独立 $\rightarrow$ $\boxed{F = \dfrac{Y_1/m}{Y_2/n} \sim F(m,n)}$

- $\dfrac{1}{F} \sim F(n,m)$，若 $X \sim t(n)$，则 $X^2 \sim F(1,n)$

## 三　正态总体下抽样分布的特有性质

### 1.均值和方差满足分布

- 若 $X_i \sim N(\mu,\sigma^2)$，则 $\overline{X}$ 和 $S^2$ 相互独立，且

| **正态总体的样本均值** | **正态总体的样本方差** |
| --- | --- |
| $\boxed{\overline{X} \sim N(\mu,\dfrac{\sigma^2}{n})}$ | $\boxed{\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)}$ |

### 2.正态总体的样本方差的方差

**正态总体的样本方差的方差**

$$\boxed{D(S^2) = \frac{2\sigma^4}{n-1}}$$

# 题型解析

## 十二　求解统计量的性质

### 1.题型简述与解法

· 通常以填空题的形式考查由样本、样本均值、样本方差等构成的统计量的事件概率、期望、方差等

· 经常用到的知识点：

① 期望、方差、协方差的运算性质

② 正态分布的性质，以及正态变量的线性组合依然是正态变量

③ 样本均值、样本方差的期望和方差，以及正态总体下的特殊性质

· 题干经常会有一大串的式子，其中一些实际上是样本方差的展开式，要能看出来

· 需要注意：不是所有的题目都是正态分布，要看清楚题干再使用正态总体下的结论

### 2.历年考试典型例题

**① 计算协方差或相关系数**

**例 1**　（15－16 **秋冬**）设总体 $X\sim N(0,4)$，$X_1,\cdots,X_{100}$ 为来自 $X$ 的简单随机样本，则 $\sum_{i=1}^{60}X_i$ 与 $\sum_{i=51}^{100}\left|X_i\right|$ 的相关系数为__________。

**解**　根据协方差的性质：

$$\mathrm{Cov}\left(\sum_{i=1}^{60}X_i,\sum_{i=51}^{100}\left|X_i\right|\right)=\sum_{i=1}^{60}\sum_{j=51}^{100}\mathrm{Cov}\left(X_i,\left|X_j\right|\right)$$

由于 $X_i$ 之间相互独立，因此

$$\sum_{j=51}^{60}\mathrm{Cov}\left(X_i,\left|X_i\right|\right)$$

由于

$$\mathrm{Cov}\left(X_i,\left|X_i\right|\right)=E\left(X_i\left|X_i\right|\right)-E(X_i)E\left(\left|X_i\right|\right)$$

$$E\left(X_i\left|X_i\right|\right)=\int_{-\infty}^{+\infty}x\left|x\right|f(x)\,\mathrm{d}x=0\text{（}x\left|x\right|f(x)\text{ 为奇函数）}$$

$$E(X_i)=0$$

因此 $\mathrm{Cov}\left(X_i,\left|X_i\right|\right)=0\ \rightarrow\ \mathrm{Cov}\left(\sum_{i=1}^{60}X_i,\sum_{i=51}^{100}\left|X_i\right|\right)=0\ \rightarrow$ 相关系数为 0

**② 计算事件概率**

**例 2**　（19－20 **秋冬**）设总体 $X\sim N(\mu,1)$，$X_1,\cdots,X_{16}$ 是 $X$ 的简单随机样本，$\overline{Y_1}=(X_1+\cdots+X_4)/4$，$\overline{Y_2}=(X_5+\cdots+X_{16})/12$，则 $P\left(\left|\overline{Y_1}-\mu\right|<1\right)=$__________.

**解**　根据正态总体的性质，$\overline{Y_1}\sim N(\mu,\frac{1}{4})$，则由正态分布的性质，$\overline{Y_1}-\mu\sim N(0,\frac{1}{4})$

因此$P(\left|\overline{Y_1}-\mu\right|<1)=P(|\frac{\overline{Y_1}-\mu}{1/2}|<\frac{1}{1/2})=2\Phi(2)-1$

**例 3** （**16－17 春夏**）设总体$X\sim N(\mu,\sigma^2)$，$\mu,\sigma^2$未知，$X_1,\cdots,X_{16}$为来自$X$的简单随机样本，$\overline{X}$和$S^2$分别为样本均值与样本方差，则$P\left(\frac{X_1+X_2}{2}-\overline{X}>\sigma\right)=$________；

**解** 由于$\overline{X}$中包含了$X_1$和$X_2$，需要注意减号左右两边不是相互独立的

$\because\ \frac{X_1+X_2}{2}\sim N(\mu,\frac{\sigma^2}{2})$，$\overline{X}\sim N(\mu,\frac{\sigma^2}{16})$，记$Y=\frac{X_1+X_2}{2}-\overline{X}$

$\therefore\ E(Y)=\mu-\mu=0$，$D(Y)=\frac{\sigma^2}{2}+\frac{\sigma^2}{16}-2\mathrm{Cov}(\frac{X_1+X_2}{2},\overline{X})$

$\because\ \mathrm{Cov}(\frac{X_1+X_2}{2},\overline{X})=\mathrm{Cov}(\sum_{i=1}^{2}\frac{1}{2}X_i,\sum_{j=1}^{16}\frac{1}{16}X_j)=\frac{1}{32}\mathrm{Cov}(\sum_{i=1}^{2}X_j,\sum_{i=1}^{16}X_j)=\frac{D(X_1)+D(X_2)}{32}=\frac{\sigma^2}{16}$

$\therefore\ D(Y)=\frac{\sigma^2}{2}+\frac{\sigma^2}{16}-2\frac{\sigma^2}{16}=\frac{7}{16}\sigma^2$ $\quad\therefore\ Y\sim N(0,\frac{7}{16}\sigma^2)$

$\therefore\ P(Y>\sigma)=1-\Phi(4/\sqrt{7})=1-\Phi(1.51)=0.07$

② **计算均值和方差**

**例 4** （**20－21 秋冬**）设总体$X\sim N(\mu,\sigma^2)$，$X_1,\cdots,X_n$是$X$的简单随机样本，$\overline{X}$，$S^2$分别是样本均值和样本方差，若$\mu=0$，则$\mathrm{Var}(S^2-\overline{X}^2)=$________.

**解** 根据正态总体的性质，$S^2$与$\overline{X}$相互独立

$\because\ \overline{X}\sim N(0,\frac{\sigma^2}{n})\ \rightarrow\ \frac{\overline{X}}{\sigma/\sqrt{n}}\sim N(0,1)\ \rightarrow\ \frac{n}{\sigma^2}\overline{X}^2\sim\chi^2(1)$

$\therefore\ \mathrm{Var}(\frac{n}{\sigma^2}\overline{X}^2)=2\ \rightarrow\ \frac{n^2}{\sigma^4}\mathrm{Var}(\overline{X}^2)=2\ \rightarrow\ \mathrm{Var}(\overline{X}^2)=\frac{2\sigma^4}{n^2}$

$\because\ \mathrm{Var}(S^2)=\frac{2\sigma^4}{n-1}\ \rightarrow\ \mathrm{Var}(S^2-\overline{X}^2)=\frac{2\sigma^4}{n-1}-\frac{2\sigma^4}{n^2}$

**例 5** （**19－20 秋冬**）设总体$X\sim N(\mu,1)$，$X_1,\cdots,X_{16}$是$X$的简单随机样本，$\overline{Y_1}=(X_1+\cdots+X_4)/4$，$\overline{Y_2}=(X_5+\cdots+X_{16})/12$，则$\mathrm{Var}\left[\sum_{i=1}^{4}(X_i-\overline{Y_1})^2+\sum_{i=5}^{16}(X_i-\overline{Y_2})^2\right]=$________.

**解** 注意到$\sum_{i=1}^{4}(X_i-\overline{Y_1})^2=(4-1)S_1^2$，$\sum_{i=5}^{16}(X_i-\overline{Y_2})^2=11S_2^2$，且两者相互独立（没有重叠的$X_i$）

由正态总体$\mathrm{Var}(S^2)=\frac{2\sigma^4}{n-1}$，得

$$\mathrm{Var}\left[\sum_{i=1}^{4}(X_i-\overline{Y_1})^2+\sum_{i=5}^{16}(X_i-\overline{Y_2})^2\right]=\mathrm{Var}\left[\sum_{i=1}^{4}(X_i-\overline{Y_1})^2\right]+\mathrm{Var}\left[\sum_{i=5}^{16}(X_i-\overline{Y_2})^2\right]$$

$$=\mathrm{Var}(3S_1^2)+\mathrm{Var}(11S_2^2)=3^2\times\frac{2}{3}+11^2\times\frac{2}{11}=28$$

## 十三　判断统计量函数的分布

### 1. 题型简述与解法

· 填空题，判断统计量函数属于哪种抽样分布，或已知含未知数的统计量函数满足某分布，求该未知数

· 既然是判断抽样分布，则总体一定是正态分布，以下结论都是适用的：

| 常用结论 1 | 常用结论 2 | 常用结论 3 |
| --- | --- | --- |
| $\dfrac{n(\overline{X}-\mu)^2}{\sigma^2}\sim\chi^2(1)$ | $\dfrac{1}{\sigma^2}\sum\limits_{i=1}^{n}(X_i-\mu)^2\sim\chi^2(n)$ | $\dfrac{1}{\sigma^2}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2\sim\chi^2(n-1)$ |

常用结论 3 实际上是 $S^2$ 的展开式，因为题目通常给出方差的展开式而不是 $S^2$

· 如果 $\mu=0$，此时要意识到 $X_i^2$ 可以除以 $\sigma^2$ 变成 $\chi^2$ 分布（或者用常用结论 2）

· 大多数题目会考察 $F$ 分布，即会把以上分布作比值处理，此时 $\sigma^2$ 、$n$ 等参数都会被合并或约掉，使我们一眼看不出来，这时就要根据以上结论将给的式子拆成两个 $\chi^2$ 分布

### 2. 历年考试典型例题

**例 1**　（15－16 **春夏**）设总体 $X\sim N(\mu,\sigma^2)$， $X_1,\cdots,X_n$ 为来自 $X$ 的简单随机样本，则

$$\sum_{i=1}^{3}\left(X_i-\frac{X_1+X_2+X_3}{3}\right)^2\Bigg/\sum_{i=4}^{5}(X_i-\mu)^2\sim\underline{\qquad\qquad}\text{分布}.$$

**解**　$\sum\limits_{i=1}^{3}\left(X_i-\dfrac{X_1+X_2+X_3}{3}\right)^2$ 具有 3 个样本的方差形式，根据常用结论 3：

$$\frac{1}{\sigma^2}\sum_{i=1}^{3}\left(X_i-\frac{X_1+X_2+X_3}{3}\right)^2\sim\chi^2(2)$$

根据常用结论 2：

$$\frac{1}{\sigma^2}\sum_{i=4}^{5}(X_i-\mu)^2\sim\chi^2(2)$$

两者相互独立，因此

$$\sum_{i=1}^{3}\left(X_i-\frac{X_1+X_2+X_3}{3}\right)^2\Bigg/\sum_{i=4}^{5}(X_i-\mu)^2=\frac{\dfrac{1}{\sigma^2}\sum\limits_{i=1}^{3}\left(X_i-\dfrac{X_1+X_2+X_3}{3}\right)^2/2}{\dfrac{1}{\sigma^2}\sum\limits_{i=4}^{5}(X_i-\mu)^2/2}\sim F(2,2)$$

**例 2**　（15－16 **秋冬**）设总体 $X\sim N(0,4)$， $X_1,\cdots,X_{100}$ 为来自 $X$ 的简单随机样本，则 $\sum\limits_{i=1}^{100}|X_i|$ 近似服从

分布，$\left(\sum\limits_{i=1}^{50}X_i\right)^2/\sum\limits_{i=51}^{100}X_i^2\sim\underline{\qquad\qquad}$分布.

**解**　$\sum\limits_{i=1}^{50}X_i$ 具有前 50 个样本的均值形式，根据常用结论 1：

$$\frac{50(\frac{1}{50}\sum_{i=1}^{50}X_i)^2}{\sigma^2}=\frac{\frac{1}{50}(\sum_{i=1}^{50}X_i)^2}{\sigma^2}\sim\chi^2(1)$$

注意到$\mu=0$，因此$\sum_{i=51}^{100}X_i^2=\sum_{i=51}^{100}(X_i-0)^2$，根据常用结论 2：

$$\frac{1}{\sigma^2}\sum_{i=51}^{100}X_i^2\sim\chi^2(50)$$

∴ 两者相互独立：$\left(\sum_{i=1}^{50}X_i\right)^2/\sum_{i=51}^{100}X_i^2=\dfrac{50\cdot\dfrac{\frac{1}{50}(\sum_{i=1}^{50}X_i)^2}{\sigma^2}}{\dfrac{1}{\sigma^2}\sum_{i=51}^{100}X_i^2}=\dfrac{\dfrac{\frac{1}{50}(\sum_{i=1}^{50}X_i)^2}{\sigma^2}}{\dfrac{1}{\sigma^2}\sum_{i=51}^{100}X_i^2/50}\sim F(1,50)$

**例 3** （15—16 **秋冬**）设总体$X\sim N(\mu,1)$，$\mu$未知，$X_1,\cdots,X_{25}$为来自$X$的简单随机样本，$\overline{X}$是样本均值，若$a(\overline{X}-X_1)^2\sim\chi^2(1)$，则$a=$________________.

**解** ∵ $\overline{X}\sim N(\mu,\frac{1}{25})$ $X_1\sim N(\mu,1)$ $\mathrm{Cov}(\overline{X},X_1)=\frac{1}{25}D(X_1)=\frac{1}{25}$

∴ $D(\overline{X}-X_1)=\frac{1}{25}+1-2\times\frac{1}{25}=\frac{24}{25}$ $\rightarrow$ $\overline{X}-X_1\sim N(0,\frac{24}{25})$

∴ 由常用结论 2，$\dfrac{1\times(\overline{X}-X_1)^2}{24/25}\sim\chi^2(1)$，比对得到$a=\dfrac{25}{24}$

# 第 8 讲　点估计

## 知识梳理

### 一　点估计的基本流程

**1.矩估计法**

① 求出分布（含待估参数$\theta$）的期望（含参数$\theta$）

② 变换期望的表达式为$\theta = f(E(X))$的形式

③ 将$E(X)$换成$\overline{X}$，得到矩估计量$\hat{\theta} = f(\overline{X})$

**2.极大似然估计法**

· 假设有简单随机样本$X_1,\cdots,X_n$，分布律或密度函数中含有待估参数$\lambda$

① 写出极大似然函数$L(\lambda)$：所有$X_i$对应取值的概率或概率密度之积

② 取对数$\ln L(\lambda)$（将乘积式化为加和式，方便求导）

③ 对$\ln L(\lambda)$求导（若参数有多个，则分别对各个参数求偏导）

④ 如果存在极值，找到极大值对应的$\lambda$，作为极大似然估计量$\hat{\lambda}$

如果$L(\lambda)$是单调的，找到$\lambda$取值范围的最值使$L(\lambda)$达到最大值，作为极大似然估计量$\hat{\lambda}$

### 二　点估计量的评价

**1.无偏性准则**

| 无偏性准则 |
| --- |
| $\hat{\theta}$ 是 $\theta$ 的无偏估计 $\Leftrightarrow$ $E(\hat{\theta}) = \theta$ |

**2.有效性准则**

| 有效性准则 |
| --- |
| 若估计量都是$\theta$的无偏估计，$\mathrm{Var}(\hat{\theta})$越小的更有效（越小越好） |

**3.均方误差准则**

| 均方误差 |
| --- |
| $\mathrm{Mse}(\hat{\theta}) = E[(\hat{\theta}-\theta)^2] = \mathrm{Var}(\hat{\theta}) + [E(\hat{\theta})-\theta]^2$ |

**4.相合性准则**

| 相合性准则 |
| --- |
| $\hat{\theta}$ 是 $\theta$ 的相合估计 $\Leftrightarrow$ $\hat{\theta}_n \xrightarrow{P} \theta, n \to +\infty$ |

# 题型解析

## 十四　估计量评价

### 1.题型简述与解法

- 以填空题或大题的形式，判断某估计量是否满足四种准则之一，或已知估计评价反求参数
- 代入计算即可，注意要运用上一讲的结论以及期望方差的运算性质

### 2.历年考试典型例题

① 无偏估计

**例 1**　（15—16 春夏）设总体 $X \sim N(\mu,\sigma^2)$，$X_1,\cdots,X_n$ 为来自 $X$ 的简单随机样本；若 $aX_1^2+bX_1X_2$ 是 $\sigma^2$ 的无偏估计，则 $(a-b)=$ ＿＿＿＿＿＿.

**解**　由无偏估计 $\rightarrow$ $E(aX_1^2+bX_1X_2)=aE(X_1^2)+bE(X_1X_2)=a(\mu^2+\sigma^2)+b\mu^2=\sigma^2$

比较系数：$a+b=0$，$a=1$ $\rightarrow$ $b=-1$，因此 $a-b=2$

② 有效性准则

**例 2**　（16—17 春夏）总体 $X$ 的密度函数 $f(x;\theta)=\begin{cases}\dfrac{2x}{\theta^2}, 0<x<\theta\\ 0, \quad 其他\end{cases}$，$\theta>0$ 未知，$X_1,X_2,X_3$ 是 $X$ 的简单随机样本，设 $T=aX_1+bX_2+cX_3$，其中 $a,b,c$ 是实数.

（1）求 $T$ 是 $\theta$ 的无偏估计的充分必要条件；

（2）问 $a,b,c$ 取什么值时，$T$ 是 $\theta$ 的有效估计量？说明理由.

**解**　（1）$T$ 是 $\theta$ 的无偏估计 $\Leftrightarrow$ $E(T)=aE(X_1)+bE(X_2)+cE(X_3)=(a+b+c)E(X)=\theta$

$$\because\ E(X)=\int_0^\theta \frac{2x^2}{\theta^2}\mathrm{d}x=\frac{2}{3}\theta \quad \therefore\ (a+b+c)\frac{2}{3}\theta=\theta \ \Leftrightarrow\ a+b+c=\frac{3}{2}$$

（2）有效的前提是无偏，因此 $a+b+c=\dfrac{3}{2}$

$$\because\ \mathrm{Var}(T)=a^2\mathrm{Var}(X)+b^2\mathrm{Var}(X)+c^2\mathrm{Var}(X)=(a^2+b^2+c^2)\mathrm{Var}(X)$$

由基本不等式，$a^2+b^2+c^2\geq\dfrac{(a+b+c)^2}{3}$

$\therefore$ 当 $a=b=c=\dfrac{1}{2}$ 时，$\mathrm{Var}(T)$ 取到最小值，此时 $T$ 是 $\theta$ 的有效估计量

③ 均方误差准则

**例 3**　（17—18 春夏）设总体 $X \sim N(\mu,\sigma^2)$，$X_1,\cdots,X_{16}$ 是总体 $X$ 的简单随机样本，$\overline{X}$ 是样本均值.若 $\mu=0$，用 $T=16(\overline{X})^2$ 估计 $\sigma^2$，则均方误差 $\mathrm{Mse}(T)=$ ＿＿＿＿＿＿

**解**　$\because\ \overline{X}\sim N(0,\sigma^2)$ $\rightarrow$ $\dfrac{16}{\sigma^2}\overline{X}^2\sim\chi^2(1)$ $\rightarrow$ $E(\overline{X}^2)=\dfrac{\sigma^2}{16}$，$D(\overline{X}^2)=\dfrac{2\sigma^4}{16^2}$

$\therefore\ E(T)=16E(\overline{X}^2)=\sigma^2,\quad D(T)=2\sigma^4$

$\because\ \mathrm{Mse}(T)=D(T)+[E(T)-\sigma^2]^2=2\sigma^4$

$\therefore\ E(Y)=\mu-\mu=0,\quad D(Y)=\dfrac{\sigma^2}{2}+\dfrac{\sigma^2}{16}-2\mathrm{Cov}(\dfrac{X_1+X_2}{2},\overline{X})$

$\because\ \mathrm{Cov}(\dfrac{X_1+X_2}{2},\overline{X})=\mathrm{Cov}(\sum_{i=1}^{2}\dfrac{1}{2}X_i,\sum_{j=1}^{16}\dfrac{1}{16}X_j)=\dfrac{1}{32}\mathrm{Cov}(\sum_{i=1}^{2}X_j,\sum_{i=1}^{16}X_j)=\dfrac{D(X_1)+D(X_2)}{32}=\dfrac{\sigma^2}{16}$

$\therefore\ D(Y)=\dfrac{\sigma^2}{2}+\dfrac{\sigma^2}{16}-2\dfrac{\sigma^2}{16}=\dfrac{7}{16}\sigma^2\quad\therefore\ Y\sim N(0,\dfrac{7}{16}\sigma^2)$

$\therefore\ P(Y>\sigma)=1-\Phi(4/\sqrt{7})=1-\Phi(1.51)=0.07$

④ **相合性准则**

**例 4**　（14—15 秋冬）随机变量 $X$ 的概率密度为 $f(x;\theta)=\begin{cases}2\theta^2x, 0<x<\theta^{-1}\\ 0,\quad 其他\end{cases}$．（2）若 $\theta>0$ 是未知参数，设 $X_1,\cdots,X_n$ 是来自总体 $X$ 的样本，$\dfrac{1}{\overline{X}}$ 是 $\theta$ 的相合估计吗？答：__________（是或不是）．

**解**　$E(X)=\int_0^{\theta^{-1}}2\theta^2x^2\mathrm{d}x=\dfrac{2}{3}\theta^{-1}\quad\therefore\ \dfrac{1}{\overline{X}}\xrightarrow{P}\dfrac{1}{E(X)}=\dfrac{3}{2}\theta\neq\theta$ → 不是相合估计

## 十五　求矩估计并评价

### 1.题型简述与解法

· 按照求矩估计的流程依葫芦画瓢即可：**求 $E(X)$ → 转换成 $\theta=$ 的形式 → 将 $E(X)$ 换成 $\overline{X}$**

### 2.历年考试典型例题

**例 1**　（15—16 秋冬）设总体 $X$ 的分布律如下表，其中 $0<\theta<1$ 为未知参数.

| $X$ | 0 | 1 | 3 | 6 |
|---|---|---|---|---|
| 概率 | $\dfrac{\theta}{2}$ | $\dfrac{\theta}{2}$ | $\dfrac{2(1-\theta)}{3}$ | $\dfrac{1-\theta}{3}$ |

（1）$X_1,\cdots,X_n$ 为来自 $X$ 的简单随机样本，求 $\theta$ 的矩估计量 $\hat{\theta}_1$，并判断其是否为无偏估计量，是否为相合估计量，说明理由；

**解**　$E(X)=0\times\dfrac{\theta}{2}+1\times\dfrac{\theta}{2}+3\times\dfrac{2(1-\theta)}{3}+6\times\dfrac{1-\theta}{3}=4-\dfrac{7\theta}{2}$ → $\theta=\dfrac{2}{7}(4-E(X))$

$\therefore$ 矩估计量 $\hat{\theta}_1=\dfrac{2}{7}(4-\overline{X})$，其中 $\overline{X}=\dfrac{1}{n}\sum_{i=1}^{n}X_i$

$\because\ E(\hat{\theta}_1)=\dfrac{2}{7}(4-E(\overline{X}))=\dfrac{2}{7}(4-E(X))=\dfrac{2}{7}(4-4+\dfrac{7}{2}\theta)=\theta\quad\therefore$ 是无偏估计

$n\to\infty,\quad\hat{\theta}_1=\dfrac{2}{7}(4-\overline{X})\xrightarrow{P}\dfrac{2}{7}(4-E(X))=\theta\quad\therefore$ 是相合估计

## 十六 求极大似然估计并评价

### 1.题型简述与解法

- 按照求极大估计的流程依葫芦画瓢即可：**列出 $L(\lambda)\to$ 取对数 $\to$ 求导 $\to$ 找最大值**
- 注意：如果 $L(\lambda)$ 是单调函数，就要找 $\lambda$ 的取值范围，此时如果密度函数给出 $0<x<\lambda$ 之类的

  就能得到 $\lambda$ 的最值，即 $\max\{X_i\}$ 之类
- 因此在求期望时就需要用到第 5 讲求随机变量函数的分布函数的技巧了
- 此外，有时还要求 $P(X?)$ 的极大似然估计，将参数的极大似然估计代入即可

### 2.历年考试典型例题

**例 1** （16—17 **秋冬**）大学新生报到时，无家长陪同，1 位家长陪同，2 位家长陪同的概率分别为 $\theta$，$(1-\theta)/4$，$3(1-\theta)/4$（这里 $\theta$ 未知）.现按简单随机抽样调查了 100 名新生，设新生中陪同的家长数是 0 位、1 位、2 位的人数分别是 $n_0,n_1,n_2$，$n_0+n_1+n_2=100$.设 $Y$ 表示 100 个新生中无家长陪同的人数.

（2）求 $P(Y\le 13)$ 的近似值（用 $\theta$ 表示）

（3）（节选）求 $\theta$ 的极大似然估计量；

（4）（节选）若 $n_0=10$，$n_1=26$，$n_2=64$，求 $\theta$ 的极大似然估计值，以及 $P(Y\le 13)$ 的极大似然估计近似值.

**解** （2）由题意 $Y\sim B(100,\theta)$，因此 $Y\sim N(100\theta,100\theta(1-\theta))$

$$\therefore\ P(Y\le 13)=P\left(\frac{Y-100\theta}{\sqrt{100\theta(1-\theta)}}\le\frac{13-100\theta}{\sqrt{100\theta(1-\theta)}}\right)=\Phi\left(\frac{13-100\theta}{\sqrt{100\theta(1-\theta)}}\right)$$

（3）$L(\theta)=\prod_{i=1}^{100}P(X_i=x_i)=\theta^{n_0}\cdot\left(\frac{1-\theta}{4}\right)^{n_1}\cdot\left(\frac{3(1-\theta)}{4}\right)^{n_2}$

取对数：$\ln L(\theta)=n_0\ln\theta+(n_1+n_2)\ln(1-\theta)+C$（$C$ 为与 $\theta$ 无关的值）

求导：$\frac{\mathrm{d}\ln L(\theta)}{\mathrm{d}\theta}=\frac{n_0}{\theta}-\frac{n_1+n_2}{1-\theta}=0\ \to\ \hat{\theta}=\frac{n_0}{n_0+n_1+n_2}=\frac{n_0}{100}$

（4）代入 $n_0=10$：$\hat{\theta}=\frac{10}{100}=0.1$

$P(Y\le 13)=\Phi(1)=0.84$

**例 2** （20—21 **秋冬**）设总体 $X$ 的概率密度函数为 $f(x,\theta)=\begin{cases}\frac{2\theta^2}{x^3}, & x\ge\theta\\ 0, & x<\theta\end{cases}$，其中 $\theta>0$ 是未知参数，$X_1,\cdots,X_n$ 是总体 $X$ 的简单随机样本，

（2）求 $\theta$ 的极大似然估计量 $\hat{\theta}_2$，并判断其是否为 $\theta$ 的无偏估计量，说明理由.

**解** $L(\theta)=\prod_{i=1}^{n}f(x_i;\theta)=\prod_{i=1}^{n}\frac{2\theta^2}{X_i^3}=\frac{2^n\theta^{2n}}{\prod_{i=1}^{n}X_i^3}(X_1,\cdots,X_n\geq\theta)$

取对数：$\ln L(\theta)=L(\theta)=n\ln 2+2n\ln\theta-3\sum_{i=1}^{n}\ln X_i\ \to\ L(\theta)$ 单调递增

$\therefore$ 当$\theta$取到最大值时，$L(\theta)$达到最大值 $\because\ X_1,\cdots,X_n\geq\theta\ \to\ \hat{\theta}_2=\min\{X_i\}$

下面判断$\hat{\theta}_2$是否为无偏估计

$P(\hat{\theta}_2\leq z)=P(\min\{X_i\}\leq z)=1-P(X_1>z,\cdots,X_n>z)=1-P^n(X>z)=1-[1-F_X(z)]^n$

$\therefore\ F_X(x)=\begin{cases}0, & x<\theta\\ 1-\dfrac{\theta^2}{x^2}, & x\geq\theta\end{cases}\ \to\ F_Z(z)=1-[1-F_X(z)]^n=\begin{cases}0, & z<\theta\\ 1-\dfrac{\theta^{2n}}{z^{2n}}, & z\geq\theta\end{cases}$

$\therefore\ f_Z(z)=F_Z'(z)=\begin{cases}0, & z<\theta\\ \dfrac{2n\theta^{2n}}{z^{2n+1}}, & z\geq\theta\end{cases}\quad E(\hat{\theta}_2)=\int_{\theta}^{+\infty}\frac{2n\theta^{2n}}{z^{2n}}\mathrm{d}z=\frac{2n}{2n-1}\theta\neq\theta$

$\therefore\ \hat{\theta}_2$不是$\theta$的无偏估计

**例 3** （18—19 **春夏**）设总体 $X$ 的分布律为 $P(X=0)=a$，$P(X=1)=b$，$P(X=2)=a+b$，$P(X=3)=1-2(a+b)$.未知参数$a>0$，$b>0$，$a+b<0.5$，$X_1,\cdots,X_{400}$是总体$X$的简单随机样本，其中 0，1，2，3 分别出现 60，100，140，100 次.（1）求$a,b$的极大似然估计值；

**解** $L(a,b)=\prod_{i=1}^{n}P(X=x_i;a,b)=a^{60}b^{100}(a+b)^{140}[1-2(a+b)]^{100}$

$\ln L(a,b)=60\ln a+100\ln b+140\ln(a+b)+100\ln[1-2(a+b)]$

含有两个参数，因此求偏导：

$$\begin{aligned}\frac{\partial\ln L(a,b)}{\partial a}&=\frac{60}{a}+\frac{140}{a+b}-\frac{200}{1-2(a+b)}\\ \frac{\partial\ln L(a,b)}{\partial b}&=\frac{100}{b}+\frac{140}{a+b}-\frac{200}{1-2(a+b)}\end{aligned}\ \to\ \text{解得}\begin{cases}\hat{a}=9/64\\ \hat{b}=15/64\end{cases}\quad\text{（严格上讲要求二阶导验证）}$$

# 第 9 讲　区间估计与假设检验

## 知识梳理

### 一　枢轴量与抽样分布的上α分位数

1. 常用正态总体枢轴量

| 分布名称 | 待估参数 | 条件 | 枢轴量 |
|---|---|---|---|
| 单个正态总体 | $\mu$ | $\sigma^2$ 已知 | $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ |
| | $\mu$ | $\sigma^2$ 未知 | $\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$ |
| | $\sigma^2$ | —— | $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ |
| 两个正态总体 | $\mu_1-\mu_2$ | $\sigma^2$ 已知 | $\dfrac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}} \sim N(0,1)$ |
| | $\mu_1-\mu_2$ | $\sigma^2$ 未知 | $\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{1/n_1+1/n_2}} \sim t(n_1+n_2-2)$ |
| | $\sigma_1^2/\sigma_2^2$ | $\sigma^2$ 已知 | $\dfrac{(n_1-1)S_1^2/\sigma_1^2}{(n_2-1)S_2^2/\sigma_2^2} \sim F(n_1-1,n_2-1)$ |

· 注：$S_w=\sqrt{\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$

2. 上 α 分位数

· 设 $Y$ 遵从某种分布（标准正态分布、$t$ 分布、$\chi^2$ 分布、$F$ 分布中的一种）

若存在实数 $a$ 使得 $P(Y>a)=\alpha$，则称 $a$ 为该分布的上 $\alpha$ 分位数

| 分布 | 分位数符号 | 性质 |
|---|---|---|
| $N(0,1)$ | $z_\alpha$ | $z_{1-\alpha}=-z_\alpha$ |
| $\chi^2(n)$ | $\chi_\alpha^2(n)$ | —— |
| $t(n)$ | $t_\alpha(n)$ | $t_{1-\alpha}(n)=-t_\alpha(n)$ |
| $F(n_1,n_2)$ | $F_\alpha(n_1,n_2)$ | $F_{1-\alpha}(n_1,n_2)=\dfrac{1}{F_\alpha(n_1,n_2)}$ |

· 相当于分布函数的反函数，已知概率反求取值

## 二　区间估计与参数假设检验的基本原理

### 1.区间估计的基本原理

· 以估计$\sigma^2$未知下的$\mu$为例

由于简单随机样本$X_1,\cdots,X_n$的样本均值$\overline{X}$和样本方差$S^2$都是随机变量

→ 枢轴量$\dfrac{\overline{X}-\mu}{S/\sqrt{n}}$也是一个一元连续型随机变量，且满足$t(n)$分布

→ $t(n)$分布函数完全已知，可以计算$\dfrac{\overline{X}-\mu}{S/\sqrt{n}}$落入某个区间$(a,b)$的概率

→ 现在我们选择一个区间$(a,b)$，使得$\dfrac{\overline{X}-\mu}{S/\sqrt{n}}$落入$(a,b)$的概率为 95%， $a$ 和$b$是完全确定的

→ 则有$P(a<\dfrac{\overline{X}-\mu}{S/\sqrt{n}}<b)=95\%$，变换内部的不等式得到$P(\overline{X}-\dfrac{\sqrt{n}}{S}b<\mu<\overline{X}-\dfrac{\sqrt{n}}{S}a)=95\%$

也就是说，当我们随机取样获得样本统计值时，待估参数$\mu$落在上述区间的概率为 95%

由于区间可以任意选择，因此我们通常会选择区间长度最短的$(a,b)$

### 2.参数假设检验的基本原理

· 以检验$\sigma^2$未知下的$\mu$为例

原假设的$\mu$已知的前提下，我们可以直接计算$\dfrac{\overline{X}-\mu}{S/\sqrt{n}}$落入某个区间$(a,b)$的概率

比如通过计算，可以找到一个区间$(a,b)$， $\dfrac{\overline{X}-\mu}{S/\sqrt{n}}$落在区间$(a,b)$外的概率为 5%

显然这个概率非常小，如果我们取样统计，发现这个小概率事件发生了，那么我们就拒绝原假设

因此只要$\dfrac{\overline{X}-\mu}{S/\sqrt{n}}$落在区间$(a,b)$外就拒绝原假设，这个外区间就是拒绝域

· $P_$值的解释

原假设的$\mu$已知的前提下，我们可以直接计算$\dfrac{\overline{X}-\mu}{S/\sqrt{n}}$落入某个区间$(a,b)$的概率

现在我们取样统计，得到了一个枢轴量$T$，这个$T$可能比较偏僻

按照原假设的分布，我们可以计算出枢轴量取值比现在这个$T$更偏僻更极端的概率

如果这个概率非常小（如小于规定的 5%），那我们也认为这次抽样属于小概率事件发生了

因此拒绝原假设

# 题型解析

## 十七　求待估参数的置信区间

**1.题型简述与解法**

- 以填空题或大题的形式，求出某正态总体的待估参数的置信区间
- 按照如下步骤进行（考试直接写出置信区间的表达式即可，不用写推导过程）：

① **确定区间估计类型**（如$\sigma^2$未知求$\mu$），在下表中**选择对应的枢轴量**$G(\theta)$

这个$G(\theta)$应当含有待估参数$\theta$，且不能含有其他未知参数

| 分布名称 | 待估参数 | 条件 | 枢轴量 |
| --- | --- | --- | --- |
| 单个正态总体 | $\mu$ | $\sigma^2$已知 | $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$ |
| | $\mu$ | $\sigma^2$未知 | $\dfrac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$ |
| | $\sigma^2$ | —— | $\dfrac{(n-1)S^2}{\sigma^2}\sim \chi^2(n-1)$ |
| 两个正态总体 | $\mu_1-\mu_2$ | $\sigma^2$已知 | $\dfrac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}\sim N(0,1)$ |
| | $\mu_1-\mu_2$ | $\sigma^2$未知 | $\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{1/n_1+1/n_2}}\sim t(n_1+n_2-2)$ |
| | $\sigma_1^2/\sigma_2^2$ | $\sigma^2$已知 | $\dfrac{(n_1-1)S_1^2/\sigma_1^2}{(n_2-1)S_2^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$ |

② **确认置信区间的类型**，写出对应的置信区间（此处用$g_\alpha(n)$代表分位数）

| 检验类型 | 置信区间原始式 |
| --- | --- |
| 双侧置信区间 | $\boxed{g_{1-\alpha/2}(n)<G(\theta)<g_{\alpha/2}(n)}$ |
| 单侧置信上限 | $\boxed{G(\theta)>g_{1-\alpha}(n)}$ |
| 单侧置信下限 | $\boxed{G(\theta)<g_\alpha(n)}$ |

③ **对$P$内部的不等式进行变换**，将待估参数$\theta$分离出来，得到置信区间表达式

④ **将统计值、$n$、$g_\alpha$或$g_{1-\alpha}$代入**，就得到置信区间或单侧置信上下限

- 这种方法只需记住枢轴量以及置信区间类型对应的不等式形式，无需记忆大量的置信区间

**2.历年考试典型例题**

**例 1**　设总体$X\sim N(\mu,\sigma^2)$，$X_1,\cdots,X_n$为来自$X$的简单随机样本，设样本均值为$\overline{x}$，样本标准差为$s$，置信水平为95%

（1）（15—16 **春夏**）若$n=16$，$\overline{x}=5.098$，$s=1.200$，则$\sigma^2$的单侧置信上限为_____________；

（2）（15—16 **秋冬**）若$n=9$，$\overline{x}=2.15$，$s=0.9$，则$\sigma^2$的双侧置信区间为_______________；

（3）（17－18 **秋冬**）若 $n=9$，$\overline{x}=1.8$，$\sum_{i=1}^{16}(x_i-\overline{x})^2=30.6$，则 $\mu$ 的置信区间为________。

**解** （1）（2） 判断类型：求 $\sigma^2$，因此选择枢轴量 $\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$

（1） · 判断区间类型：单侧置信上限，选择 $\frac{(n-1)S^2}{\sigma^2}>\chi^2_{1-\alpha}(n)$

· 变换：$\frac{(n-1)S^2}{\sigma^2}>\chi^2_{1-\alpha}(n)\Rightarrow\sigma^2<\frac{(n-1)S^2}{\chi^2_{1-\alpha}(n)}$

· 代入数据，得单侧置信上限为 $\frac{15s^2}{\chi^2_{0.95}(16)}=2.975$

（2） · 判断区间类型：双侧置信区间，选择 $\chi^2_{1-\alpha/2}(n)<\frac{(n-1)S^2}{\sigma^2}<\chi^2_{\alpha/2}(n)$

· 变换：$\chi^2_{1-\alpha/2}(n)<\frac{(n-1)S^2}{\sigma^2}<\chi^2_{1-\alpha/2}(n)\Rightarrow\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n)}<\sigma^2<\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n)}$

· 代入数据，得置信区间为 $\left(\frac{8\times0.9^2}{17.5},\frac{8\times0.9^2}{2.18}\right)=(0.37,2.97)$

（3） 判断类型：$\sigma^2$ 未知，求 $\mu$，选择枢轴量 $\frac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$

· 判断区间类型：双侧置信区间，选择 $t_{1-\alpha/2}(n-1)<\frac{\overline{X}-\mu}{S/\sqrt{n}}<t_{\alpha/2}(n-1)$

· 变换：$\overline{X}-\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1)<\mu<\overline{X}-\frac{S}{\sqrt{n}}t_{1-\alpha/2}(n-1)$

· 代入数据，得单侧置信上限为 $(1.8-\frac{\sqrt{30.6}}{16}\times2.13,1.8+\frac{\sqrt{30.6}}{16}\times2.13)$，即 $(1.04,2.56)$

**例 2** （18－19 **秋冬**）为验证某汽车厂生产的汽车平均每公升汽油行驶里程是否达到 15km 以上，随机选取 16 辆车，记录下每辆车每公升汽油行驶的千米数，得到样本均值 $\overline{x}=14.22$，样本方差 $s^2=1.2^2$，假设数据来自正态分布 $N(\mu,\sigma^2)$。

（2）现有另一汽车厂生产的同类型汽车，其每公升汽油行驶的千米数 $Y\sim N(\mu_Y,\sigma^2)$，随机选取该类型汽车 11 辆车，测得样本均值 $\overline{y}=14.97$，样本方差 $s_y^2=1.4^2$，求 $\mu-\mu_Y$ 的置信度为 95%的双侧置信区间。（保留两位小数）

**解** · 判断类型：$\mu_1-\mu_2$（$\sigma^2$ 均未知），因此选择 $\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{1/n_1+1/n_2}}\sim t(n_1+n_2-2)$

· 判断区间类型：双侧置信区间，选择 $t_{1-\alpha/2}(n_1+n_2-2)<\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{1/n_1+1/n_2}}<t_{\alpha/2}(n_1+n_2-2)$

· 变换：$\left|\mu-\mu_Y-(\overline{X}-\overline{Y})\right|<+S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}t_{\alpha/2}(n_1+n_2-2)$

· 代入数据，解得置信区间为 $(-1.79,0.29)$

## 十八　求拒绝域并检验假设

### 1. 题型简述与解法

· 实际上考试从来没有求过拒绝域，检验假设通常都会要求计算 $P\_$ 值，$P\_$ 值都算出来了谁还算拒绝域啊！

① 根据要检验的参数，选择对应的检验统计量 $G$（下表以双边检验为例）

| 分布名称 | 原假设 | 条件 | 检验统计量 |
|---|---|---|---|
| 单个正态总体 | $\mu=\mu_0$ | $\sigma^2$ 已知 | $Z=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$ |
| | $\mu=\mu_0$ | $\sigma^2$ 未知 | $T=\dfrac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$ |
| | $\sigma^2=\sigma_0^2$ | —— | $\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$ |
| 两个正态总体 | $\mu_1=\mu_2$ | $\sigma^2$ 已知 | $\dfrac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}\sim N(0,1)$ |
| | $\mu_1=\mu_2$ | $\sigma^2$ 未知 | $\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{1/n_1+1/n_2}}\sim t(n_1+n_2-2)$ |
| | $\sigma_1^2=\sigma_2^2$ | $\sigma^2$ 已知 | $\dfrac{(n_1-1)S_1^2/\sigma_1^2}{(n_2-1)S_2^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$ |

② 根据检验类型，选取对应的拒绝域公式并计算

| 检验类型 | 条件 | 拒绝域 |
|---|---|---|
| 左侧检验 | $H_0$： $\theta\ge\theta_0$ | $W=\{G_0<g_{1-\alpha}\}$ |
| 右侧检验 | $H_0$： $\theta\le\theta_0$ | $W=\{G_0>g_\alpha\}$ |
| 双侧检验 | $H_0$： $\theta=\theta_0$ | $W=\{\lvert G_0\rvert>g_{\alpha/2}\}$（$\mu$ 相关）<br>$W=\{G_0<g_{1-\alpha/2}\cup G_0>g_{\alpha/2}\}$（$\sigma^2$ 相关） |

③ 计算枢轴量 $G_0$，如果落在了拒绝域内，就拒绝原假设

### 2. 历年考试典型例题

**例 1**　（19—20 春夏）为了解某县粮食产量情况，随机调查该县 64 个乡当年的粮食产量，得到样本均值为 1120 吨，样本方差 108900，设乡粮食产量（单位：吨）$X\sim N(\mu,\sigma^2)$，$\mu,\sigma^2$ 未知.

在显著水平 0.05 下检验假设 $H_0:\mu\le 1000, H_1:\mu>1000$

**解**　· 确定类型：$\sigma^2$ 未知，检验 $\mu$，取枢轴量 $T=\dfrac{\overline{X}-\mu}{S/\sqrt{n}}$

· 确定检验类型：右侧检验，取拒绝域 $W=\{T_0>t_\alpha(n-1)\}$ $T_0>t_\alpha(n-1)$

· 计算 $T_0=\dfrac{1120-1000}{\sqrt{108900}/\sqrt{64}}=2.909$，$t_{0.05}(63)=1.669$ $\rightarrow$ $T_0>t_{0.05}(63)$，拒绝原假设

## 十九　求 P 值并检验假设

### 1.题型简述与解法

· 以填空题或大题的形式，根据样本值求出某正态总体的 $P\_$，并判断是否拒绝原假设

① **根据要检验的参数，选择并计算对应的检验统计量 $G$**

| 分布名称 | 原假设 | 条件 | 检验统计量 |
|---|---|---|---|
| 单个正态总体 | $\mu=\mu_0$ | $\sigma^2$ 已知 | $Z=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$ |
| | $\mu=\mu_0$ | $\sigma^2$ 未知 | $T=\dfrac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$ |
| | $\sigma^2=\sigma_0^2$ | —— | $\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$ |
| 两个正态总体 | $\mu_1=\mu_2$ | $\sigma^2$ 已知 | $\dfrac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}\sim N(0,1)$ |
| | $\mu_1=\mu_2$ | $\sigma^2$ 未知 | $\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{1/n_1+1/n_2}}\sim t(n_1+n_2-2)$ |
| | $\sigma_1^2=\sigma_2^2$ | $\sigma^2$ 已知 | $\dfrac{(n_1-1)S_1^2/\sigma_1^2}{(n_2-1)S_2^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$ |

② **根据检验类型，选择对应的 $P\_$ 值公式并计算**

| 检验类型 | 条件 | P 值公式 |
|---|---|---|
| 左侧检验 | $H_0:\ \theta\geq\theta_0$ | $P\_=P(g<G_0)$ |
| 右侧检验 | $H_0:\ \theta\leq\theta_0$ | $P\_=P(g>G_0)$ |
| 双侧检验 | $H_0:\ \theta=\theta_0$ | $P\_=2P(g>\lvert G_0\rvert)$（$\mu$ 相关）<br>$P\_=2\min\{P(g>G_0),P(g<G_0)\}$（$\sigma^2$ 相关） |

具体方法：在试卷第一页抬头寻找 $g_A=G_0$，则 $P(g<G_0)=1-A$，$P(g>G_0)=A$

③ **如果 $P\_<\alpha$，则拒绝原假设**

### 2.历年考试典型例题

**例 1**　（18—19 **春夏**）为调查某减肥药的疗效，随机选择 16 位服药一个疗程的使用者，记录他们的减肥重量 $X$（单位：公斤），假设 $X\sim N(\mu,\sigma^2)$，已测得样本均值 $\overline{x}=1.18$，样本标准差 $s=1.6$.

（1）对于假设：$H_0:\mu\leq 0,H_1:\mu>0$，求 $P\_$ 值并进行检验（取 $\alpha=0.05$）；

**解**　（1）$\sigma^2$ 未知检验 $\mu$ $\rightarrow$ 取检验统计量

$$T=\frac{\overline{X}-\mu}{S/\sqrt{n}} \rightarrow T_0=\frac{1.18-0}{1.6/\sqrt{16}}=2.95$$

右侧检验 $\rightarrow$ $P\_=P(t(15)>T_0)$

$\because$ $t_{0.005}(15)=2.95$ $\rightarrow$ $P\_=0.005<0.05$ $\rightarrow$ 拒绝原假设

**例 2** （**14－15 春夏**）设总体 $X\sim N(\mu,\sigma^2)$， $X_1,\cdots,X_n$ 为来自 $X$ 的简单随机样本，若取得容量是 9 的样本，计算得样本均值 $\overline{x}=1.896$，样本标准差为 $s=0.8$，则假设 $H_0:\mu=1,H_1:\mu\neq1$ 的 $P\_$ 值为________，若显著水平为 0.05，则应该拒绝还是接受原假设______________.

**解** $\sigma^2$ 未知检验 $\mu$ $\rightarrow$ 取检验统计量

$$T=\frac{\overline{X}-\mu}{S/\sqrt{n}} \rightarrow T_0=\frac{1.896-1}{0.8/\sqrt{9}}=3.36$$

双侧检验 $\rightarrow$ $P\_=P(t(8)>T_0)$

$\because$ $t_{0.005}(8)=3.36$ $\rightarrow$ $P\_=0.005<0.05$ $\rightarrow$ 拒绝原假设

**例 3** （**19－20 秋冬**） 为了解某市两所高校学生的消费情况，在两所高校各随机调查 100 人，调查结果为甲校学生月平均消费 2583 元，样本方差 882669，乙校学生月平均消费 2439 元，样本方差 678976，设甲校学生月平均消费额 $X\sim N(\mu,\sigma_1^2)$，乙校学生月平均消费额 $Y\sim N(\mu_2,\sigma_2^2)$，$\mu_1,\mu_2,\sigma_1^2,\sigma_2^2$ 未知，两个样本独立.

（1）在显著水平 0.05 下检验假设 $H_0:\sigma_1^2=\sigma_2^2,H_1:\sigma_1^2\neq\sigma_2^2$，并计算相应的 $P\_$ 值；

**解** 检验 $\sigma_1^2=\sigma_2^2$ $\rightarrow$ 取检验统计量

$$F=\frac{S_1^2}{S_2^2}\sim F(99,99) \rightarrow F_0=1.30$$

双侧检验 $\rightarrow$ $P(F(99,99)\geq1.30)=0.1$ $\rightarrow$ $P\_=2P(F(99,99)\geq1.30)=0.2>0.05$ $\rightarrow$ 接受

**例 4** （**17－18 秋冬**）设总体 $X\sim N(\mu,\sigma^2)$， $\mu,\sigma^2$ 均未知， $X_1,\cdots,X_{16}$ 是总体 $X$ 的简单随机样本， $\overline{X}$ 是样本均值，若计算得 $\overline{x}=1.8$， $\sum_{i=1}^{16}(x_i-\overline{x})^2=30.6$，为检验假设 $H_0:\sigma^2\leq1,H_1:\sigma^2>1$，$P_=$_______,若显著水平 $\alpha=0.05$,应该拒绝还是接受原假设？答:____________________.

**解** 检验 $\sigma^2$ $\rightarrow$ 取检验统计量

$$\chi^2=\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1) \rightarrow \chi_0^2=30.6$$

右侧检验 $\rightarrow$ $P\_=P(\chi^2>\chi_0^2)$，由 $\chi_{0.01}^2(15)=30.6$ $\rightarrow$ $P\_=0.01<0.05$ $\rightarrow$ 拒绝原假设

## 二十 拟合优度检验

### 1.题型简述与解法

· 已知总体 $X$ 的大量样本数据，使用拟合优度检验判断 $X$ 是否服从指定分布

① **将 $X$ 的取值范围划分为 $k$ 个区间，获得各个区间内数据数量 $n_i$（实际频数）**

· 如果 $X$ 是连续型，则要将将 $X$ 的取值范围划分为 $k$ 个区间（题目已经划分好了）

② **假设 $X$ 服从该分布，计算 $X$ 在各个区间内的理论频数 $np_i$**

· $n$ 为样本数，$p_i$ 为 $X$ 落在区间 $i$ 的概率

· 如果分布含有未知参数，则先用极大似然法估出参数（这种情况题目一般会设两问）

③ **计算统计量 $\chi^2$，若 $\chi^2 \le \chi_\alpha^2(k-1)$，则接受原假设，否则拒绝原假设**

**拟合优度统计量**

$$\chi^2=\sum_{i=1}^{k}\frac{(n_i-np_i)^2}{np_i}$$

### 2.历年考试典型例题

**例 1** （18—19 春夏）设总体 $X$ 取值 0，1，2，3，4，5，对总体进行 100 次观察，其中 0，1，2，3，4，5 分别观察到 11，18，19，21，16，15 次.

（1）若总体的分布律如下表所示，未知参数 $p\in(0,1)$，求参数 $p$ 的极大似然估计值 $\hat{p}$；

| $X$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 概率 | $0.25p$ | $0.5p(1-p)$ | $0.5p(1-p)$ | $(1-p)^2$ | $0.5p$ | $0.25p$ |

（2）在显著水平 0.05 下，用 $\chi^2$ 拟合优度检验法检验假设：$H_0$: $X$ 的分布律如上表所示.

**解** （1）$L(p)=(0.25p)^{20}[0.5p(1-p)]^{37}(1-p)^{42}(0.5p)^{36}$

$\to$ $\ln L(p)=79\ln p+79\ln(1-p)+C$

$\to$ $\dfrac{\mathrm{d}\ln L(p)}{\mathrm{d}p}=79(\dfrac{1}{p}+\dfrac{1}{1-p})=0\Rightarrow p=\dfrac{1}{2}$

$\to$ 当 $p=\dfrac{1}{2}$ 时，$L(p)$ 取极大值 $\to$ $\hat{p}=\dfrac{1}{2}$

（2）取 $\hat{p}=\dfrac{1}{2}$，计算理论频数如下表所示

| $X$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 理论频数 | 12.5 | 12.5 | 12.5 | 25 | 25 | 12.5 |
| 实际频数 | 11 | 18 | 19 | 21 | 16 | 15 |

$$\chi^2=\frac{(12.5-11)^2}{12.5}+\frac{(12.5-18)^2}{12.5}+\frac{(12.5-19)^2}{12.5}+\frac{(25-21)^2}{25}+\frac{(25-16)^2}{12.5}+\frac{(12.5-15)^2}{12.5}$$

$=10.36>9.49=\chi_{0.05}^2(5-1)$ $\to$ 拒绝原假设