概率论与数理统计

小助手

出品：仲英学辅

2020 年 9 月 1 日

仲英书院学业辅导中心

ZHONG YING XUE FU

XI'AN JIAOTONG UNIVERSITY

作品信息

- **标题：**概率论与数理统计 - 小助手
- **主编作者：**电气 813 赖启平、电气 713 刘沁阳
- **校对排版：**电气 86 刘菁锐、能动 B81 梁佳佳
- **出品时间：**2020 年 9 月 1 日
- **总页数：**40

许可证说明

前言

编写人员：电气 813 赖启平、电气 713 刘沁阳

排版人员：电气 86 刘菁锐、能动 B81 梁佳佳

感谢学业辅导中心各位工作人员与志愿者的努力工作，使本资料可以按时完工。由于编者们的能力与精力限制，难免有错误之处。如果同学们在本资料中发现错误，请联系仲英学业辅导中心：XJTUzyxuefu@163.com，我们将在修订时予以更正。

从第 3 周开始，每晚 19:30-21:30，学辅志愿者在东 21 舍 118 学辅办公室值班，当面为学弟学妹们答疑。

同时，我们也有线上答疑平台——学粉群。

19 级学粉群：902493560，756433480;

20 级学粉群：598243135, 1137961185.

期中考试与期末考试前，我们还会举办考前讲座。学辅还有新生专业交流会，转专业交流会，英语考试讲座等活动，消息会在学粉群和公众号上公布，欢迎同学们参与。

仲英书院学业辅导中心

2020 年 9 月 1 日

学粉群 6.0
QQ 群号：598243135

学粉群 6.1
QQ 群号：1137961185

微信公众号
仲英学业辅导中心及薪火工作室

目录

概率统计与随机过程

一、随机事件与概率

确定性现象、随机现象（统计规律性）。

概率论与数理统计是研究和揭示随机现象的统计规律性的学科。

【重点：随机事件及其概率、概率的公理化定义及其性质、条件概率、乘法公式、全概率公式、贝叶斯公式、事件的相互独立性】

1.1 随机事件

1.1.1 随机现象与随机试验

1. 随机试验
   (1) 可以在相同条件下重复进行。
   (2) 每次试验的可能结果不止一个，且能事先明确试验的所有可能结果。
   (3) 进行一次试验之前不能确定哪一个结果会出现。

1.1.2 样本空间与随机事件

1. 样本空间：随机试验 E 的所有可能结果组成的集合称为 E 的样本空间，记为 S（Ω）。
2. 样本点：样本空间的元素，即 E 的每个结果，称为样本点（基本结果），记为 ω。
3. 随机事件（事件）：试验 E 的样本空间 S 的子集为 E 的随机事件。在每次试验中，当且仅当这一子集中的一个样本点出现时，称这一事件发生。
   (1) 基本事件：一个样本点组成的单点集称为基本事件。
   (2) 必然事件：如 S（Ω）。
   (3) 不可能事件：如 ∅ 。

1.1.3 事件间的关系与事件的运算

1. 事件的包含与相等。
2. 和事件（并事件）。
3. 积事件（交事件）。
4. 差事件。$A-B=A\underline{B}$，$A-B=A-AB$。
5. 互斥事件（互不相容）。
6. 对立事件（补事件）。
7. 运算律：交换、结合、分配、对偶律（德摩根律）。

1.2 概率

1.1.1 概率的古典定义

1. 等可能概型（古典概型）
   (1) 有限性：试验的样本空间只包含有限个元素。
   (2) 等可能性：试验中每个基本事件发生的可能性相同。

$$
\begin{aligned}
1 = P(S) &= P(\{e_1\} \cup \{e_2\} \cup \cdots \cup \{e_n\}) \\
&= P(\{e_1\}) + P(\{e_2\}) \\
&\quad + \cdots + P(\{e_n\}) = nP(\{e_i\}),
\end{aligned}
$$

$$
P(\{e_i\}) = \frac{1}{n}, \quad i = 1,2,\cdots,n.
$$

2. 等可能事件中事件 A 概率的计算公式：

$$P(A)=\sum_{j=1}^{k}P(\{e_{i_j}\})=\frac{k}{n}=\frac{A\text{ 包含的基本事件数}}{S\text{ 中基本事件的总数}}.$$

附：由古典概型可以构造典型的概率模型——

- 袋中有 n 个球，1 个红球，其余都为黑球，从袋中无放回地摸出 r 个球，令 A={摸出的球中有红球}，$\overline{A}$={摸出的球中没有红球}；请计算 A 及 $\overline{A}$ 的概率，并观察计算的结果会得出什么数学公式？$C_n^r=C_{n-1}^{r-1}+C_{n-1}^r$
- 袋中有 n 个球，m 个红球，其余都为黑球，从袋中无放回地摸出 r 球(r<=m)，令 Ai={摸出的 r 个球中有 i 个红球}，(i=0,1,2,...r)，请计算 Ai（i=0,1,2,...r）的概率，并观察你计算的结果，会得出什么数学公式?$C_n^r=\sum_{i=0}^{r}C_m^iC_{n-m}^{r-i}$

3. 几何概型

将古典概型中的有限性推广到无限性，而保留等可能性，就得到几何概型。

P(A)=A 的几何度量/S 的几何度量=L(A)/L(S)

- 有一个可度量的几何图形（几何度量：长度、面积、体积）
- 试验 E 可看作在 S 中随机地投掷一点
- 事件 A 就是所投掷的点落在 S 中的可度量图形 A 中

1.2.2 概率的统计定义

1. 频率：事件发生的频繁程度。相同条件下，进行 n 次试验，n 次试验中事件 A 发生的次数 n_A 称为事件 A 发生的频数。n_A/n 称为事件 A 发生的频率，记作 $f_n(A)$。

2. 频率的稳定性：随机事件 A 在相同条件下重复多次时，事件 A 发生的频率在一个 固定的数值 p 附近摆动，随试验次数的增加更加明显。

3. 频率的性质

(1) $0\le f_n(A)\le 1$。

(2) $f_n(S)=1$。

(3) 若 A_k 互不相容，则 $f_n(A_1\cup\cdots A_k)=f_n(A_1)+\cdots f_n(A_k)$。

4. 概率的统计定义：对于任意事件 A，在相同条件下重复进行 n 次试验，事件 A 发生的频率 m/n，随着试验次数 n 的增大而稳定地在某个常数附近摆动，则称 p 为事件 A 的概率，当试验次数足够大时，可以用事件 A 发生的频率近似代替事件 A 的概率。

1.2.3 概率的公理化定义

1. 事件在一次试验中发生的可能性大小。概率的公理化定义：给定一个随机试验，S 是它的样本空间，对于任意一个事件 A，赋予一个实数 P(A)，如果 P(·)满足下列三条公理，则称 P(A)为事件 A 的概率。

- 非负性：$0\le \mathrm{P}(A)\le 1$。

- 规范性：$P(S)=1$。
- 可列可加性：若 A_k 互不相容，则 $P(A_1 \cup A_2 \cdots) = P(A_1) + P(A_2) + \cdots$。

1.2.4 概率性质

1. **性质 i** $P(\varnothing)=0.$

2. **性质 ii（有限可加性）** 若 $A_1, A_2, \cdots, A_n$ 是两两互不相容的事件，则有

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n). \tag{3.2}$$

3. **性质 iii** 设 A, B 是两个事件，若 $A \subset B$，则有

$$P(B-A) = P(B) - P(A);$$

$$P(B) \geqslant P(A).$$

推论：对于任意事件 A 有，P(B-A)=P(B)-P(AB)

4. **性质 iv** 对于任一事件 A，

$$P(A) \leqslant 1.$$

5. **性质 v（逆事件的概率）** 对于任一事件 A，有

$$P(\overline{A}) = 1 - P(A).$$

6. **性质 vi（加法公式）** 对于任意两事件 A, B 有

$$P(A \cup B) = P(A) + P(B) - P(AB).$$

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1A_2) - P(A_1A_3) - P(A_2A_3) + P(A_1A_2A_3).$$

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{1 \leqslant i < j \leqslant n} P(A_iA_j) + \sum_{1 \leqslant i < j < k \leqslant n} P(A_iA_jA_k) + \cdots + (-1)^{n-1} P(A_1A_2 \cdots A_n).$$

7. $P\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} P(A_i)$

8. 概率的连续性

$$A_n \subset A_{n+1}, n = 1, 2, \cdots, \ P(\bigcup_{n=1}^{\infty} A_n) = \lim_{n \to \infty} P(A_n)$$

$$A_n \supset A_{n+1}, n = 1, 2, \cdots, \ P(\bigcap_{n=1}^{\infty} A_n) = \lim_{n \to \infty} P(A_n)$$

1.3 条件概率 全概率公式与贝叶斯公式

1.3.1 条件概率与乘法公式

1. 引入:考虑一个有两个子女的家庭,问两个孩子都为男孩的概率为多少？若已知有男孩，则都为男孩的概率为多少？若已知第一个孩子为男孩，则都为男孩的概率为多少？

2. 定义：设 A，B 为同一个随机试验中的两个随机事件，在已知事件 A 已发生的条件下，事件 B 发生的概率称为条件概率。

$$P(B|A)=\frac{k}{m}=\frac{k/n}{m/n}=\frac{P(AB)}{P(A)}.$$

3. 乘法定理

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

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

$$P(A_1A_2\cdots A_n)=P(A_n|A_1A_2\cdots A_{n-1})P(A_{n-1}|A_1A_2\cdots A_{n-2})\cdots P(A_2|A_1)P(A_1)$$

1.3.2 全概率公式和贝叶斯公式

1. 样本空间的划分

定义 设 S 为试验 E 的样本空间，$B_1,B_2,\cdots,B_n$ 为 E 的一组事件.若

(i) $B_iB_j=\varnothing, i\neq j, i,j=1,2,\cdots,n$;

(ii) $B_1\cup B_2\cup\cdots\cup B_n=S$,

则称 $B_1,B_2,\cdots,B_n$ 为样本空间 S 的一个划分.

2. 全概率公式

定理 设试验 E 的样本空间为 S，A 为 E 的事件，$B_1,B_2,\cdots,B_n$ 为 S 的一个划分，且 $P(B_i)>0\ (i=1,2,\cdots,n)$，则

$$P(A)=P(A|B_1)P(B_1)+P(A|B_2)P(B_2)+\cdots+P(A|B_n)P(B_n). \tag{5.6}$$

3. 贝叶斯公式（后验公式）

定理 设试验 E 的样本空间为 S. A 为 E 的事件，$B_1,B_2,\cdots,B_n$ 为 S 的一个划分，且 $P(A)>0, P(B_i)>0\ (i=1,2,\cdots,n)$，则

$$P(B_i\mid A)=\frac{P(A\mid B_i)P(B_i)}{\sum_{j=1}^{n}P(A\mid B_j)P(B_j)},\quad i=1,2,\cdots,n. \tag{5.7}$$

常用形式：

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

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

1.4 随机事件的独立性

1.4.1 两个事件的独立性

1. 定义：设 A、B 为任意两个随机事件，若 P(B|A)=P(B)，则称 B 相对于 A 独立，并且可进一步证明，A, B 是相互独立的。事件 A 与事件 B 独立的充分必要条件是 P(AB)=P(A)P(B)。

1.4.2 多个事件的独立性

1. 多个事件独立性的概念

定义 设 A,B,C 是三个事件，如果满足等式

$$\left.\begin{aligned}&P(AB)=P(A)P(B),\\&P(BC)=P(B)P(C),\\&P(AC)=P(A)P(C),\\&P(ABC)=P(A)P(B)P(C),\end{aligned}\right\}\qquad(6.2)$$

则称事件 A,B,C 相互独立.

一般，设 $A_1,A_2,\cdots,A_n$ 是 n ($n\geqslant 2$) 个事件，如果对于其中任意 2 个，任意 3 个，…，任意 n 个事件的积事件的概率，都等于各事件概率之积，则称事件 A_1，$A_2,\cdots,A_n$ 相互独立.

注：共有 2^n-n-1 个等式；两两独立不一定相互独立。

2. 定理

由定义，可以得到以下两个推论.

1°若事件 $A_1,A_2,\cdots,A_n(n\geqslant 2)$ 相互独立，则其中任意 k ($2\leqslant k\leqslant n$) 个事件也是相互独立的.

2°若 n 个事件 $A_1,A_2,\cdots,A_n(n\geqslant 2)$ 相互独立，则将 $A_1,A_2,\cdots,A_n$ 中任意多个事件换成它们各自的对立事件，所得的 n 个事件仍相互独立.

3°若 n 个事件相互独立，则有 $P(\bigcup_{i=1}^{n}A_i)=1-P(\bigcap_{i=1}^{n}\overline{A_i})$。

二、随机变量及概率分布

为了更方便地研究随机事件及其概率，我们引入概率论中另一个重要的概念——随机变量，这一变量是依赖于随机试验结果的实值函数。因为随机变量是实值函数，那么就可以借助微积分知识与研究方法，对随机现象的概率规律进行分门别类的研究，总结出重要的几类概率模型。

2.1 一维随机变量

2.1.1 随机变量与分布函数

1. 随机变量

将样本空间数量化，即用数值来表示试验的结果。

定义 2.1.1 设 E 为一随机试验，Ω 为其样本空间，若 $X=X(\omega),\omega\in\Omega$ 为单值实函数，且对于任意实数 x，集合 $\{\omega|X(\omega)\leqslant x\}$ 都是随机事件，则称 X 为**随机变量**. 随机变量经常用 X、Y、Z 等表示.

2. 随机变量的分布函数

对于非离散型随机变量X，其可能取值不能一一列举出来，因而不能像离散型随机变量那样用分布律描述。非离散型随机变量取任意一指定实数值的概率等于0。

(1) 定义：

定义 设 X 是一个随机变量，x 是任意实数，函数

$$F(x)=P\{X\leqslant x\},\quad -\infty<x<\infty$$

称为 X 的**分布函数**.

对于任意实数 $x_1,x_2(x_1<x_2)$，有

$$\begin{aligned}P\{x_1<X\leqslant x_2\}&=P\{X\leqslant x_2\}-P\{X\leqslant x_1\}\\&=F(x_2)-F(x_1),\end{aligned}\tag{3.1}$$

因此，若已知 X 的分布函数，我们就知道 X 落在任一区间 $(x_1,x_2]$ 上的概率，从这个意义上说，分布函数完整地描述了随机变量的统计规律性.

如果将 X 看成是数轴上的随机点的坐标，那么，分布函数 $F(x)$ 在 x 处的函数值就表示 X 落在区间 $(-\infty,x]$ 上的概率.

(2) 基本性质

1° $F(x)$ 是一个不减函数.

事实上，由(3.1)式对于任意实数 $x_1,x_2(x_1<x_2)$，有

①

$$F(x_2)-F(x_1)=P\{x_1<X\leqslant x_2\}\geqslant 0.$$

2° $0\leqslant F(x)\leqslant 1$，且

$$F(-\infty)=\lim_{x\to-\infty}F(x)=0,$$

②

$$F(\infty)=\lim_{x\to\infty}F(x)=1.$$

③ 3° $F(x+0)=F(x)$，即 $F(x)$ 是右连续的.

一般，设离散型随机变量 X 的分布律为

$$P\{X=x_k\}=p_k,\quad k=1,2,\cdots.$$

由概率的可列可加性得 X 的分布函数为

$$F(x)=P\{X\leqslant x\}=\sum_{x_k\leqslant x}P\{X=x_k\},$$

即

④

$$F(x)=\sum_{x_k\leqslant x}p_k,$$

注:如果一个函数满足上述性质,那么该函数一定是某个随机变量的分布函数。

2.1.2 离散型随机变量

1. 离散型随机变量的定义：全部可能的取值是有限个或可列无限多个，这种随机变量称为称为离散型随机变量。

设离散型随机变量 X 所有可能取的值为 $x_k(k=1,2,\cdots)$，X 取各个可能值的概率，即事件 $\{X=x_k\}$ 的概率，为

$$P\{X=x_k\}=p_k,\ k=1,2,\cdots. \tag{2.1}$$

由概率的定义，p_k 满足如下两个条件：

$$1^\circ\quad p_k\geqslant 0,\ k=1,2,\cdots; \tag{2.2}$$

$$2^\circ\quad \sum_{k=1}^{\infty}p_k=1. \tag{2.3}$$

2. 离散型随机变量的分布律

$$P\{X=x_k\}=p_k,\ k=1,2,\cdots.$$

X	x_1	x_2	$\cdots$	x_n	$\cdots$
p_k	p_1	p_2	$\cdots$	p_n	$\cdots$

3. 三种重要的离散型随机变量

(1) （0-1）分布

设随机变量 X 只可能取 0 与 1 两个值，它的分布律是

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

X	0	1
p_k	$1-p$	p

对于一个随机试验，如果它的样本空间只包含两个元素，即 $S=\{e_1,e_2\}$，我们总能在 S 上定义一个服从 (0－1) 分布的随机变量

$$X=X(e)=\begin{cases}0, & 当\ e=e_1,\\ 1, & 当\ e=e_2\end{cases}$$

(2) 伯努利试验、二项分布

设试验 E 只有两个可能结果：A 及 $\bar{A}$，则称 E 为伯努利(Bernoulli)试验．设 $P(A)=p\ (0<p<1)$，此时 $P(\bar{A})=1-p$．将 E 独立重复地进行 n 次，则称这一串重复的独立试验为 n 重伯努利试验．

以 X 表示 n 重伯努利试验中事件 A 发生的次数，X 是一个随机变量，我们来求它的分布律．X 所有可能取的值为 $0,1,2,\cdots,n$．由于各次试验是相互独立的，因此事件 A 在指定的 $k\ (0\leqslant k\leqslant n)$ 次试验中发生，在其他 $n-k$ 次试验中 A 不发生（例如在前 k 次试验中 A 发生，而后 $n-k$ 次试验中 A 不发生）的概率为

$$\underbrace{p\cdot p\cdot\cdots\cdot p}_{k\text{个}}\cdot\underbrace{(1-p)\cdot(1-p)\cdot\cdots\cdot(1-p)}_{n-k\text{个}}=p^k(1-p)^{n-k}.$$

这种指定的方式共有$\binom{n}{k}$种，它们是两两互不相容的，故在n次试验中A发生k次的概率为$\binom{n}{k}p^k(1-p)^{n-k}$，记$q=1-p$，即有

$$P\{X=k\}=\binom{n}{k}p^kq^{n-k}, k=0,1,2,\cdots,n. \tag{2.6}$$

(3) 泊松分布

体积相对小的物质在较大空间的稀疏分布，都可以看作泊松分布，参数λ可以用观测值的平均值求出。

设随机变量X所有可能取的值为$0,1,2,\cdots$，而取各个值的概率为

$$P\{X=k\}=\frac{\lambda^k e^{-\lambda}}{k!}, k=0,1,2,\cdots,$$

其中$\lambda>0$是常数.则称X服从参数为λ的**泊松分布**，记为$X\sim\pi(\lambda)$.

泊松定理 设$\lambda>0$是一个常数，n是任意正整数，设$np_n=\lambda$，则对于任一固定的非负整数k，有

$$\lim_{n\to\infty}\binom{n}{k}p_n^k(1-p_n)^{n-k}=\frac{\lambda^k e^{-\lambda}}{k!}.$$

$$\binom{n}{k}p^k(1-p)^{n-k}\approx\frac{\lambda^k e^{-\lambda}}{k!} \quad (其中\ \lambda=np). \tag{2.7}$$

也就是说以n,p为参数的二项分布的概率值可以由参数为$\lambda=np$的泊松分布的概率值近似.上式也能用来作二项分布概率的近似计算.

- X 服从参数为 p 的几何分布$X\sim G(p)$，$P(X=m)=p(1-p)^{m-1}, m=1,2,3\cdots$
- Y 服从参数为 r 和 p 的帕斯卡分布$Y\sim P(r,p)$，

  $P(Y=n)=C_{n-1}^{r-1}p^r(1-p)^{n-r}, n=r,r+1,\cdots N,\cdots$
- Z 服从参数为 n，p 的二项分布$Z\sim B(n,p)$，

  $P(Z=k)=C_n^k p^k(1-p)^{n-k}, k=0,1,2,\cdots n$
- 超几何分布$X\sim H(n,N,M)$，$P(X=k)=\dfrac{C_M^K C_{N-M}^{n-k}}{C_N^n}, k=0,1,2\cdots\min(n,M)$

2.1.3 连续型随机变量

1. 连续型随机变量及其概率密度

   (1) 定义

一般，如上节例 2 中的随机变量那样，如果对于随机变量 X 的分布函数 $F(x)$，存在非负函数 $f(x)$，使对于任意实数 x 有

$$F(x)=\int_{-\infty}^{x} f(t)\mathrm{d}t, \tag{4.1}$$

则称 X 为**连续型随机变量**，其中函数 $f(x)$ 称为 X 的**概率密度函数**，简称**概率密度**①.

由(4.1)式，据数学分析的知识知连续型随机变量的分布函数是连续函数.

(2) 性质

1° $f(x)\geqslant 0$；

2° $\int_{-\infty}^{\infty} f(x)\mathrm{d}x=1$；

3° 对于任意实数 $x_1, x_2 (x_1\leqslant x_2)$，

$$P\{x_1<X\leqslant x_2\}=F(x_2)-F(x_1)=\int_{x_1}^{x_2} f(x)\mathrm{d}x;$$

4° 若 $f(x)$ 在点 x 处连续，则有 $F'(x)=f(x)$.

需要指出的是，对于连续型随机变量 X 来说，它取任一指定实数值 a 的概率均为 0，即 $P\{X=a\}=0$. 事实上，设 X 的分布函数为 $F(x)$，$\Delta x>0$，则由 $\{X=a\}\subset\{a-\Delta x<X\leqslant a\}$ 得

$$0\leqslant P\{X=a\}\leqslant P\{a-\Delta x<X\leqslant a\}=F(a)-F(a-\Delta x).$$

在上述不等式中令 $\Delta x\to 0$，并注意到 X 为连续型随机变量，其分布函数 $F(x)$ 是连续的. 即得

$$P\{X=a\}=0. \tag{4.4}$$

据此，在计算连续型随机变量落在某一区间的概率时，可以不必区分该区间是开区间或闭区间或半闭区间. 例如有

$$P\{a<X\leqslant b\}=P\{a\leqslant X\leqslant b\}=P\{a<X<b\}.$$

在这里，事件 $\{X=a\}$ 并非不可能事件，但有 $P\{X=a\}=0$. 这就是说，若 A 是不可能事件，则有 $P(A)=0$；反之，若 $P(A)=0$，并不一定意味着 A 是不可能事件.

以后当我们提到一个随机变量 X 的"概率分布"时，指的是它的分布函数；或者，当 X 是连续型随机变量时，指的是它的概率密度，当 X 是离散型随机变量时，指的是它的分布律.

2. 重要的连续型随机变量

(1) 均匀分布

若连续型随机变量 X 具有概率密度

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

则称 X 在区间 (a,b) 上服从**均匀分布**. 记为 $X\sim U(a,b)$.

四舍五入精确到第 k 位小数，所产生随机误差的概率服从均匀分布

$$U\sim\left(-\frac{1}{2}10^{-k},\frac{1}{2}10^{-k}\right)$$

(2) 指数分布

若连续型随机变量 X 的概率密度为

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

其中 $\theta>0$ 为常数，则称 X 服从参数为 θ 的**指数分布**.

无记忆性：

$$P\{X>s+t \mid X>s\}=P\{X>t\}.$$

事实上

$$\begin{aligned}P\{X>s+t \mid X>s\}&=\frac{P\{(X>s+t)\cap(X>s)\}}{P\{X>s\}}\\&=\frac{P\{X>s+t\}}{P\{X>s\}}=\frac{1-F(s+t)}{1-F(s)}\\&=\frac{e^{-(s+t)/\theta}}{e^{-s/\theta}}=e^{-t/\theta}\\&=P\{X>t\}.\end{aligned}$$

随机服务系统中的服务时间、电话问题中的通话时间、无线电元件的寿命、动物的寿命。

(3) 正态分布（高斯分布）

正态变量条件：受众多相互独立的随机因素影响、每一因素的影响都是微小的、且正负影响可以相互抵消。（各种测量误差、人体生理特征、工厂产品尺寸、农作物收获量、学生的考试成绩）

① 定义

若连续型随机变量 X 的概率密度为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},-\infty<x<\infty, \tag{4.10}$$

其中 $\mu,\sigma(\sigma>0)$ 为常数，则称 X 服从参数为 μ,σ 的**正态分布**或**高斯**(Gauss)分布，记为 $X\sim N(\mu,\sigma^2)$.

② 性质

1° 曲线关于 $x=\mu$ 对称. 这表明对于任意 $h>0$ 有(图 2－12)

$$P\{\mu-h<X\leqslant\mu\}=P\{\mu<X\leqslant\mu+h\}.$$

2° 当 $x=\mu$ 时取到最大值

$$f(\mu)=\frac{1}{\sqrt{2\pi}\sigma}.$$

x 离 μ 越远，$f(x)$ 的值越小. 这表明对于同样长度的区间，当区间离 μ 越远，X 落在这个区间上的概率越小.

在 $x=\mu\pm\sigma$ 处曲线有拐点. 曲线以 Ox 轴为渐近线.

另外，如果固定 σ，改变 μ 的值，则图形沿着 Ox 轴平移，而不改变其形状(如图 2－12)，可见正态分布的概率密度曲线 $y=f(x)$ 的位置完全由参数 μ 所确定. μ 称为位置参数.

如果固定 μ，改变 σ，由于最大值 $f(\mu)=\frac{1}{\sqrt{2\pi}\sigma}$，可知当 σ 越小时图形变得越尖(如图 2－13)，因而 X 落在 μ 附近的概率越大.

③ 标准正态分布

1) 定义

特别，当 $\mu=0,\sigma=1$ 时称随机变量 X 服从标准正态分布。其概率密度和分布函数分别用 $\varphi(x),\Phi(x)$ 表示，即有

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-t^2/2}, \tag{4.13}$$

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-t^2/2}dt. \tag{4.14}$$

易知

$$\Phi(-x)=1-\Phi(x) \tag{4.15}$$

2) 转化

引理 若 $X\sim N(\mu,\sigma^2)$，则 $Z=\frac{X-\mu}{\sigma}\sim N(0,1)$.

证 $Z=\frac{X-\mu}{\sigma}$ 的分布函数为

$$P\{Z\leqslant x\}=P\left\{\frac{X-\mu}{\sigma}\leqslant x\right\}=P\{X\leqslant\mu+\sigma x\}$$
$$=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\mu+\sigma x}e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt,$$

令 $\frac{t-\mu}{\sigma}=u$，得

$$P\{Z\leqslant x\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-u^2/2}du=\Phi(x),$$

由此知 $Z=\frac{X-\mu}{\sigma}\sim N(0,1)$.

于是，若 $X\sim N(\mu,\sigma^2)$，则它的分布函数 $F(x)$ 可写成

$$F(x)=P\{X\leqslant x\}=P\left\{\frac{X-\mu}{\sigma}\leqslant\frac{x-\mu}{\sigma}\right\}=\Phi\left(\frac{x-\mu}{\sigma}\right).$$

对于任意区间 $(x_1,x_2]$，有

$$P\{x_1<X\leqslant x_2\}=P\left\{\frac{x_1-\mu}{\sigma}<\frac{X-\mu}{\sigma}\leqslant\frac{x_2-\mu}{\sigma}\right\}$$
$$=\Phi\left(\frac{x_2-\mu}{\sigma}\right)-\Phi\left(\frac{x_1-\mu}{\sigma}\right).$$

3) 3σ法则

设 $X\sim N(\mu,\sigma^2)$，由 $\Phi(x)$ 的函数表还能得到（图 2－16）：

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

我们看到，尽管正态变量的取值范围是 $(-\infty,\infty)$，但它的值落在 $(\mu-3\sigma,\mu+3\sigma)$ 内几乎是肯定的事．这就是人们所说的"3σ"法则．

4) α分位点

为了便于今后在数理统计中的应用，对于标准正态随机变量，我们引入上 α 分位点的定义.

设 $X\sim N(0,1)$，若 z_α 满足条件

$$P\{X>z_\alpha\}=\alpha,\ 0<\alpha<1, \tag{4.18}$$

则称点 z_α 为标准正态分布的**上 α 分位点**（如图 2－17）．下面列出了几个常用的 z_α 的值．

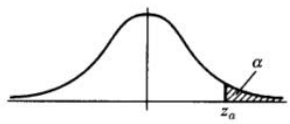

图 2－17

α	0.001	0.005	0.01	0.025	0.05	0.10
z_α	3.090	2.576	2.326	1.960	1.645	1.282

另外，由 $\varphi(x)$ 图形的对称性知道 $z_{1-\alpha}=-z_\alpha$.

3. 随机变量的函数的分布

由已知随机变量 X 的概率分布求得其函数的概率分布。

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

$$f_Y(y)=\begin{cases} f_X[h(y)]|h'(y)|, & \alpha<y<\beta, \\ 0, & \text{其他}, \end{cases} \tag{5.2}$$

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

2.2 二维随机变量

2.2.1 二维随机变量与联合分布函数

1. 二维随机变量（向量）

(1) 联合分布函数

① 定义：

定义 设 (X,Y) 是二维随机变量，对于任意实数 x,y，二元函数：

$$F(x,y)=P\{(X\leqslant x)\cap(Y\leqslant y)\}\xlongequal{\text{记成}}P\{X\leqslant x,Y\leqslant y\}$$

称为二维随机变量 (X,Y) 的**分布函数**，或称为随机变量 X 和 Y 的**联合分布函数**.

依照上述解释，借助于图 3—3 容易算出随机点 (X,Y) 落在矩形域 $\{(x,y)\mid x_1<x\leqslant x_2,y_1<y\leqslant y_2\}$ 的概率为

$$\begin{aligned} &P\{x_1<X\leqslant x_2,y_1<Y\leqslant y_2\} \\ &=F(x_2,y_2)-F(x_2,y_1)+F(x_1,y_1)-F(x_1,y_2). \end{aligned}$$

② 性质：

1° $F(x,y)$ 是变量 x 和 y 的不减函数，即对于任意固定的 y，当 $x_2>x_1$ 时 $F(x_2,y)\geqslant F(x_1,y)$；对于任意固定的 x，当 $y_2>y_1$ 时 $F(x,y_2)\geqslant F(x,y_1)$.

2° $0\leqslant F(x,y)\leqslant 1$，且

对于任意固定的 y，$F(-\infty,y)=0$，

对于任意固定的 x，$F(x,-\infty)=0$，

$F(-\infty,-\infty)=0$，$F(\infty,\infty)=1$.

3° $F(x+0,y)=F(x,y)$，$F(x,y+0)=F(x,y)$，即 $F(x,y)$ 关于 x 右连续，关于 y 也右连续.

4° 对于任意 (x_1,y_1)，(x_2,y_2)，$x_1<x_2$，$y_1<y_2$，下述不等式成立：

$$F(x_2,y_2)-F(x_2,y_1)+F(x_1,y_1)-F(x_1,y_2)\geqslant 0.$$

(2) 边缘分布函数

若给定二维随机变量(X,Y)的联合分布函数$F(x,y)$，则它的两个分量即随机变量X、Y的分布函数$F_X(x)$、$F_Y(y)$也随之确定. 因为

$$\begin{aligned} F_X(x) &= P\{X \leqslant x\} \\ &= P\{X \leqslant x, Y < +\infty\} \\ &= \lim_{y \to +\infty} P\{X \leqslant x, Y < y\} \quad \text{(利用概率的单调性和连续性)} \\ &= \lim_{y \to +\infty} F(x,y) \\ &= F(x, +\infty), \end{aligned}$$

即
$$F_X(x) = F(x, +\infty). \tag{2.2.3}$$

同理
$$F_Y(y) = F(+\infty, y). \tag{2.2.4}$$

$F_X(x)$、$F_Y(y)$分别称为二维随机变量(X,Y)关于X、关于Y的**边缘分布函数**.

由上述讨论可知，$F_X(x)$、$F_Y(y)$可由$F(x,y)$惟一地确定，但是，反过来并不一定成立

2. n维随机变量（向量）

某些随机试验的结果需要同时用两个或两个以上的随机变量来描述。

定义 2.2.1 设E是随机试验，$\Omega=\{\omega\}$是E的样本空间，而$X_1(\omega), X_2(\omega), \cdots, X_n(\omega)$是定义在$\Omega$上的$n$个随机变量，则$n$维向量$(X_1(\omega), X_2(\omega), \cdots, X_n(\omega))$称为***n*** **维随机变量**或***n*** **维随机向量**. 通常把$(X_1(\omega), X_2(\omega), \cdots, X_n(\omega))$简记为$(X_1, X_2, \cdots, X_n)$.

对于任意n个实数$x_1, x_2, \cdots, x_n$，n元函数

$$F(x_1, x_2, \cdots, x_n) = P\{X_1 \leqslant x_1, X_2 \leqslant x_2, \cdots, X_n \leqslant x_n\}$$

称为n维随机变量$(X_1, X_2, \cdots, X_n)$的**分布函数**或随机变量$X_1, X_2, \cdots, X_n$的**联合分布函数**. 它具有类似于二维随机变量的分布函数的性质.

2.2.2 二维离散型随机变量

1. 定义

如果二维随机变量(X,Y)全部可能取到的值是有限对或可列无限多对，则称(X,Y)是**离散型的随机变量**.

设二维离散型随机变量(X,Y)所有可能取的值为(x_i, y_j)，$i,j=1,2,\cdots$，记

$P\{X=x_i, Y=y_j\}=p_{ij}$，$i,j=1,2,\cdots$，则由概率的定义有

$$p_{ij} \geqslant 0, \quad \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} p_{ij} = 1.$$

2. 联合分布律

我们称 $P\{X=x_i, Y=y_j\}=p_{ij}, i,j=1,2,\cdots$ 为二维离散型随机变量 (X,Y) 的**分布律**，或随机变量 X 和 Y 的**联合分布律**.

我们也能用表格来表示 X 和 Y 的联合分布律，如下表所示.

$Y \backslash X$	x_1	x_2	$\cdots$	x_i	$\cdots$
y_1	p_{11}	p_{21}	$\cdots$	p_{i1}	$\cdots$
y_2	p_{12}	p_{22}	$\cdots$	p_{i2}	$\cdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	
y_j	p_{1j}	p_{2j}	$\cdots$	p_{ij}	$\cdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	

已知联合分布律可以求出其联合分布函数，由分布函数也可以求出其联合分布律。

2.2.3 二维连续型随机变量

1. 定义：

与一维随机变量相似，对于二维随机变量 (X,Y) 的分布函数 $F(x,y)$，如果存在非负的函数 $f(x,y)$ 使对于任意 x,y 有

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

则称 (X,Y) 是**连续型的二维随机变量**，函数 $f(x,y)$ 称为二维随机变量 (X,Y) 的**概率密度**，或称为随机变量 X 和 Y 的**联合概率密度**.

2. 性质：

1° $f(x,y)\geqslant 0$.

2° $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\mathrm{d}x\mathrm{d}y = F(\infty,\infty)=1.$

3° 设 G 是 xOy 平面上的区域，点 (X,Y) 落在 G 内的概率为

$$P\{(X,Y)\in G\}=\iint_G f(x,y)\mathrm{d}x\mathrm{d}y.$$

4° 若 $f(x,y)$ 在点 (x,y) 连续，则有

$$\frac{\partial^2 F(x,y)}{\partial x \partial y}=f(x,y).$$

这表示若 $f(x,y)$ 在点 (x,y) 处连续，则当 $\Delta x,\Delta y$ 很小时

$$P\{x<X\leqslant x+\Delta x, y<Y\leqslant y+\Delta y\}\approx f(x,y)\Delta x\Delta y,$$

也就是点 (X,Y) 落在小长方形 $(x,x+\Delta x]\times(y,y+\Delta y]$ 内的概率近似地等于 $f(x,y)\Delta x\Delta y$.

在几何上 $z=f(x,y)$ 表示空间的一个曲面. 由性质 2° 知，介于它和 xOy 平面的空间区域的体积为 1. 由性质 3°，$P\{(X,Y)\in G\}$ 的值等于以 G 为底，以曲面 $z=f(x,y)$ 为顶面的柱体体积.

3. 常用连续型二维随机变量分布
   (1) 二维均匀分布 U(G)

设(X,Y)为二维随机变量，G是平面上的一个有界区域，其面积为$A(A\neq 0)$，若二维随机变量(X,Y)的概率密度为

$$f(x,y)=\begin{cases}\dfrac{1}{A}, & (x,y)\in G,\\ 0, & (x,y)\notin G,\end{cases} \tag{2.2.14}$$

则称(X,Y)在区域G上服从**二维均匀分布**.

分布概率与面积占比。

边平行于坐标轴的矩形域上均匀分布的边缘分布仍为均匀分布。

(2) 二维正态分布

若二维随机变量(X,Y)的概率密度为

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{\frac{-1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\} \tag{2.2.13}$$

其中$-\infty<\mu_1,\mu_2<+\infty,\sigma_1>0,\sigma_2>0,|\rho|<1$，则称$(X,Y)$服从**二维正态分布**$N(\mu_1,\sigma_1{}^2;\mu_2,\sigma_2{}^2;\rho)$.

重要结论：

a. ρ=0 => $f_X(x)f_Y(y)=f(x,y)$ => X，Y独立

b. X，Y独立 => 令$x=\mu_1,y=\mu_2$=> $\rho=0$

2.3 条件分布

2.3.1 二维离散型随机变量的条件分布

1. 对于二维随机变量(X,Y)，考虑当它的一个分量的取值给定时，另一个分量的概率分布，这种分布就是二维条件分布。

2. 二维离散型随机变量的条件分布

定义 2.3.1 设(X,Y)为二维离散型随机变量，其联合分布律以及(X,Y)关于X、关于Y的边缘分布律分别为$p_{ij},p_{i\cdot},p_{\cdot j},i,j=1,2,\cdots$，对于固定的$j$，若$P\{Y=y_j\}>0$，则

$$P\{X=x_i\mid Y=y_j\}=\frac{p_{ij}}{p_{\cdot j}},\ i=1,2,\cdots \tag{2.3.1}$$

为在$Y=y_j$的条件下随机变量X的**条件分布律**.

同理，对于固定的i，若$p_{i\cdot}=P\{X=x_i\}>0$，则称

$$P\{Y=y_j\mid X=x_i\}=\frac{p_{ij}}{p_{i\cdot}},\ j=1,2,\cdots \tag{2.3.2}$$

为在$X=x_i$的条件下随机变量Y的**条件分布律**.

由条件分布律就可以得到条件分布函数：给定条件$Y=y_j(P\{Y=y_j\}>0)$下，X的条件分布函数为

$$F_{X|Y}(x\mid y_j)=P\{X\leqslant x\mid Y=y_j\}=\sum_{x_i\leqslant x}\frac{p_{ij}}{p_{\cdot j}}, \tag{2.3.3}$$

类似地，给定条件$X=x_i(P\{X=x_i\}>0)$，Y的条件分布函数为

$$F_{Y|X}(y\mid x_i)=P\{Y\leqslant y\mid X=x_i\}=\sum_{y_j\leqslant y}\frac{p_{ij}}{p_{i\cdot}}. \tag{2.3.4}$$

2.3.2 二维连续型随机变量的条件分布

1. 二维连续型随机变量的条件分布

定义 2.3.2 设(X,Y)为二维连续型随机变量，其联合概率密度为$f(x,y)$，(X,Y)关于X,Y的边缘概率密度为$f_X(x)$，$f_Y(y)$. 若对于固定的y，$f_Y(y)>0$，则在给定$Y=y$时X的条件分布函数为

$$F_{X|Y}(x \mid y) = P\{X \leqslant x \mid Y = y\} = \int_{-\infty}^{x} f_{X|Y}(u \mid y)\mathrm{d}u, \tag{2.3.7}$$

其中$f_{X|Y}(x|y)=f(x,y)/f_Y(y)$称为在$Y=y$的条件下$X$的**条件概率密度**.

类似地，若对于固定的x，$f_X(x)>0$，则在给定$X=x$时Y的条件分布函数为

$$F_{Y|X}(y \mid x) = P\{Y \leqslant y \mid X = x\} = \int_{-\infty}^{y} f_{Y|X}(v \mid x)\mathrm{d}v, \tag{2.3.8}$$

其中$f_{Y|X}(y|x)=f(x,y)/f_X(x)$称为在$X=x$的条件下$Y$的**条件概率密度**.

2.4 随机变量的相互独立性

1. 随机变量的相互独立性

定义 2.4.1 设二维随机变量(X,Y)的联合分布函数为$F(x,y)$，X与Y的边缘分布函数分别为$F_X(x)$、$F_Y(y)$，如果对任意实数x,y，恒有

$$P\{X \leqslant x, Y \leqslant y\} = P\{X \leqslant x\}P\{Y \leqslant y\},$$

即

$$F(x,y) = F_X(x)F_Y(y), \tag{2.4.1}$$

则称随机变量X与Y是**相互独立**的.

当(X,Y)是二维离散型随机变量时，X与Y相互独立的充要条件为对(X,Y)的所有可能取值$(x_i,y_j)(i,j=1,2,\cdots)$都有

$$P\{X = x_i, Y = y_j\} = P\{X = x_i\}P\{Y = y_j\} \ (i,j = 1,2,\cdots),$$

即

$$p_{ij} = p_{i\cdot} \cdot p_{\cdot j} \quad (i,j = 1,2,\cdots). \tag{2.4.2}$$

当(X,Y)是二维连续型随机变量时，设$f(x,y)$及$f_X(x)$、$f_Y(y)$分别是(X,Y)的联合概率密度及边缘概率密度，则X与Y相互独立的充要条件为：对任意实数x,y，下式几乎处处成立

$$f(x,y) = f_X(x)f_Y(y) \tag{2.4.3}$$

这里“几乎处处”的含义是指平面上除去面积为0的点外，(2.4.3)式处处成立.

n维随机变量$(X_1,X_2,\cdots,X_n)$的联合分布函数定义为

$$F(x_1,x_2,\cdots,x_n) = P\{X_1 \leqslant x_1, X_2 \leqslant x_2, \cdots, X_n \leqslant x_n\},$$

若对任意实数$x_1,x_2,\cdots,x_n$，都有

$$F(x_1,x_2,\cdots,x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\cdots F_{X_n}(x_n), \tag{2.4.4}$$

则称随机变量$X_1,X_2,\cdots,X_n$是相互独立的.

对于n维离散型随机变量和n维连续型随机变量，也相应地有类似于(2.4.2)式和(2.4.3)式的相互独立的充要条件，这里就不再罗列.

相互独立概念还可以作进一步的推广.

设$X_1,X_2,\cdots,X_n,\cdots$为一串随机变量(或称为随机变量序列)，若对任意正整数$n>1$，都

有 $X_1,X_2,\cdots,X_n$ 相互独立，则称随机变量序列 $X_1,X_2,\cdots,X_n,\cdots$ 是相互独立的.

设 m 维随机变量 $(X_1,X_2,\cdots,X_m)$ 的分布函数为 $F_1(x_1,x_2,\cdots,x_m)$，n 维随机变量 $(Y_1,Y_2,\cdots,Y_n)$ 的分布函数为 $F_2(y_1,y_2,\cdots,y_n)$，$m+n$ 维随机变量 $(X_1,\cdots,X_m,Y_1,\cdots,Y_n)$ 的分布函数为 $F(x_1,\cdots,x_m,y_1,\cdots,y_n)$. 若对任意实数 $x_1,\cdots,x_m,y_1,\cdots,y_n$，都有

$$F(x_1,\cdots,x_m,y_1,\cdots,y_n)=F_1(x_1,\cdots,x_m)F_2(y_1,\cdots,y_n), \tag{2.4.5}$$

则称 $(X_1,\cdots,X_m)$ 与 $(Y_1\cdots,Y_n)$ 是相互独立的.

下面，我们不加证明地介绍在后续课程中有重要作用的两个有关独立性的定理.

定理 2.4.1 设 $(X_1,\cdots,X_m)$ 与 $(Y_1\cdots,Y_n)$ 相互独立，若 h,g 是多元连续函数，则 $h(X_1,\cdots,X_m)$ 与 $g(Y_1\cdots,Y_n)$ 也相互独立.

定理 2.4.2 若随机变量 $X_1,X_2,\cdots,X_n$ 相互独立，把它们分为不相交的 k 个组，每个组中所有变量由一个连续函数复合而生成一个新的随机变量，则这 k 个新的随机变量仍相互独立.

两个随机变量相互独立，又有相同的分布，不能说这两个随机变量相等。

2.5 随机变量函数的概率分布

研究相关随机变量之间的关系，由已知的某个随机变量的概率分布求出另一个与之有关的随机变量的概率分布。

定义 2.5.1 记随机变量 X 的一切可能值集合为 D，设 $g(x)$ 是定义在 D 上的连续函数或分段单调函数的实函数，若对于 X 的每一个可能值 $x\in D$，随机变量 Y 相应地取值 $y=g(x)$，则称 Y 为 X 的函数，记为 $Y=g(X)$.

类似地，可以定义 n 维随机变量 $(X_1,\cdots,X_n)$ 的函数 $Y=g(X_1,\cdots,X_n)$.

本节将讨论如何从一些随机变量的概率分布导出这些随机变量的函数的概率分布.

2.5.1 一维随机变量的函数的概率分布

1. 一维离散型随机变量函数的分布

一般情况下，确定 $Y=g(X)$ 的分布律的方法在原则上与上例是一样的：把 $Y=g(X)$ 可能取的不同值找出来，再把与 Y 的某个可能值相应的所有 X 值的概率加起来，即得 Y 取这个值的概率.

2. 一维连续型随机变量函数的分布

对于这个问题，我们一般是先求出 Y 的分布函数，

$$F_Y(y)=P\{Y\leqslant y\}=P\{g(X)\leqslant y\}=\int_{g(x)\leqslant y} f_X(x)\mathrm{d}x, \tag{2.5.1}$$

再对分布函数 $F_Y(y)$ 求导，得到 Y 的概率密度 $f_Y(y)$. 这里，计算的关键是给出(2.5.1)式中的积分区间.

定理 2.5.1 设连续型随机变量 X 的概率密度为 $f_X(x)$，若 $y=g(x)$ 为 $(-\infty,+\infty)$ 上严格单调的可导函数，则 $Y=g(X)$ 也是一个连续型随机变量，且其概率密度为

$$f_Y(y)=\begin{cases} f_X(g^{-1}(y))\left|\dfrac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y}\right|, & \alpha<y<\beta, \\ 0, & \text{其他}, \end{cases}$$

其中 $\alpha=\min\limits_{-\infty<x<+\infty} g(x)$，$\beta=\max\limits_{-\infty<x<+\infty} g(x)$.

2.5.2 二维随机变量的函数的概率分布

1. 二维离散型随机变量函数的分布

先将(X,Y)的联合分布律改为逐点取值的形式，再求出随机变量函数在每一点的值，将相同值的概率合并，得到随机变量函数的分布。

独立的二项分布与独立的泊松分布具有可加性。

2. 二维连续型随机变量函数的分布

设(X,Y)是二维连续型随机变量，$f(x,y)$是其概率密度，又$Z=g(X,Y)$是X与Y的函数($g(x,y)$为已知的连续函数)，且Z是连续型随机变量，求Z的概率密度$f_Z(z)$.

对于这个问题，我们一般是先求出Z的分布函数

$$F_Z(z)=P\{Z\leqslant z\}=P\{g(X,Y)\leqslant z\}=\iint_{g(x,y)\leqslant z} f(x,y)\mathrm{d}x\mathrm{d}y, \tag{2.5.3}$$

再对分布函数$F_Z(z)$求导，得到Z的概率密度$f_Z(z)$. 这里，计算的关键是给出(2.5.3)式的积分区域.

(1) 随机变量和的概率密度公式

$$f_Z(z)=\int_{-\infty}^{+\infty} f(x,z-x)\mathrm{d}x.\quad f_Z(z)=\int_{-\infty}^{+\infty} f(z-y,y)\mathrm{d}y.$$

$$f_Z(z)=\int_{-\infty}^{+\infty} f_X(x)f_Y(z-x)\mathrm{d}x=\int_{-\infty}^{+\infty} f_X(z-y)f_Y(y)\mathrm{d}y$$

若$X_i\sim N(\mu_i,\sigma_i^2)(i=1,2,\cdots,n)$且它们相互独立，则$Z=X_1+X_2+\cdots+X_n$仍服从正态分布，且有$Z\sim N(\mu_1+\mu_2+\cdots+\mu_n,\sigma_1^2+\sigma_2^2+\cdots+\sigma_n^2)$.

此外，n 个独立的正态随机变量的线性组合仍然是正态随机变量。

(2) 随机变量商的概率密度公式

由概率密度的定义，得随机变量商$Z=X/Y$的概率密度公式为

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

特别，当X与Y相互独立时，$Z=X/Y$的概率密度为

$$f_Z(z)=\int_{-\infty}^{+\infty} |y| f_X(yz)f_Y(y)\mathrm{d}y,$$

其中$f_X(x)$、$f_Y(y)$分别为(X,Y)关于X、Y的边缘概率密度.

(3) 随机变量的概率密度公式

若X与Y相互独立，则有

$$F_M(z)=F(z,z)=F_X(z)F_Y(z), \tag{2.5.9}$$

$$F_N(z)=-[F(z,z)-F_X(z)-F_Y(z)]=1-(1-F_X(z))(1-F_Y(z)). \tag{2.5.10}$$

上述结果不难推广到$n(n>2)$个相互独立的随机变量的情形.

设$X_1,X_2,\cdots,X_n$是n个相互独立的随机变量，$F_{X_i}(x_i)$为$X_i(i=1,2,\cdots,n)$的分布函数，若设$M=\max(X_1,X_2,\cdots,X_n)$，$N=\min(X_1,X_2,\cdots,X_n)$，则有

$$F_M(z)=F_{X_1}(z)F_{X_2}(z)\cdots F_{X_n}(z),$$

$$F_N(z)=1-(1-F_{X_1}(z))(1-F_{X_2}(z))\cdots(1-F_{X_n}(z)).$$

特别地，当$X_1,X_2,\cdots,X_n$是n个相互独立同分布的随机变量时，有

$$F_M(z)=(F_{X_1}(z))^n, \tag{2.5.11}$$

$$F_N(z)=1-(1-F_{X_1}(z))^n. \tag{2.5.12}$$

对上述M与N的分布函数求导，就可得到M与N的概率密度为

$$f_M(z)=n(F_{X_1}(z))^{n-1}f_{X_1}(z), \tag{2.5.13}$$

$$f_N(z)=n(1-F_{X_1}(z))^{n-1}f_{X_1}(z), \tag{2.5.14}$$

三、随机变量的数字特征

随机变量的分布函数能够完整地描述随机变量的概率性质，而随机变量的某些特征指标，如分布的中心位置、分散程度等，一般称为随机变量的数字特征。

- 随机变量的平均值——数学期望
- 取值平均偏离均值情况——方差
- 描述两个随机变量间的某种关系——协方差和相关系数

3.1 随机变量的数学期望

3.1.1 数学期望的概念

定义 3.1.1 设离散型随机变量 X 的分布律为 $P\{X=x_i\}=p_i, i=1,2,\cdots$. 若级数 $\sum_{i=1}^{\infty} x_i p_i$ 绝对收敛，则称 $\sum_{i=1}^{\infty} x_i p_i$ 为随机变量 X 的**数学期望**，或称为**理论均值**，记作 $E(X)$，即

$$E(X)=\sum_{i=1}^{\infty} x_i p_i. \tag{3.1.1}$$

对于连续型随机变量，可以用积分代替求和，从而得到相应的数学期望定义.

定义 3.1.2 设连续型随机变量 X 具有概率密度 $f(x)$，若反常积分 $\int_{-\infty}^{+\infty} x f(x)\mathrm{d}x$ 绝对收敛，则称 $\int_{-\infty}^{+\infty} x f(x)\mathrm{d}x$ 为随机变量 X 的数学期望，记为 $E(X)$，即

$$E(X)=\int_{-\infty}^{+\infty} x f(x)\mathrm{d}x. \tag{3.1.2}$$

一些常见分布的数学期望。

3.1.2 随机变量的函数的数学期望

定理 3.1.1 设 X 是随机变量，$y=g(x)$ 为连续函数，$Y=g(X)$ 是随机变量函数，则有：

(1) 当 X 为离散型随机变量,其分布律为 $P\{X=x_i\}=p_i, i=1,2,\cdots$ 时,若 $\sum_{i=1}^{\infty}|g(x_i)|p_i$ 绝对收敛,则 $E(Y)$ 存在,且

$$E(Y)=E(g(X))=\sum_{i=1}^{\infty}g(x_i)p_i. \tag{3.1.3}$$

(2)当 X 为连续型随机变量,其概率密度为 $f_X(x)$ 时,若反常积分 $\int_{-\infty}^{+\infty}|(g(x)|f(x)\mathrm{d}x$ 绝对收敛,则 $E(Y)$ 存在,且

$$E(Y)=E(g(X))=\int_{-\infty}^{+\infty}g(x)f(x)\mathrm{d}x. \tag{3.1.4}$$

对于二维随机变量函数类似地有如下定理.

定理 3.1.2 设 (X,Y) 是二维随机变量,$z=g(x,y)$ 为二维连续函数,$Z=g(X,Y)$ 是随机变量函数,则有:

(1) 当 (X,Y) 为二维离散型随机变量,其联合分布律为 $P\{X=x_i,Y=y_j\}=p_{ij}, i,j=1,2,\cdots$ 时,若级数 $\sum_{i=1}^{\infty}|g(x_i,y_j)|p_{ij}$ 绝对收敛,则 $E(Z)$ 存在,且

$$E(Z)=E(g(X,Y))=\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}g(x_i,y_j)p_{ij}. \tag{3.1.5}$$

(2)当 (X,Y) 为连续型随机变量,其联合概率密度为 $f(x,y)$ 时,如果反常积分 $\int_{-\infty}^{+\infty}|g(x,y)|f(x,y)\mathrm{d}x\mathrm{d}y$ 绝对收敛,则 $E(Z)$ 存在,且

$$E(Z)=E(g(X,Y))=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x,y)f(x,y)\mathrm{d}x\mathrm{d}y. \tag{3.1.6}$$

3.1.3 数学期望的性质

(1) $E(C)=C$;

(2) $E(CX)=CE(X)$;

(3) $E(X+Y)=E(X)+E(Y)$;

(4) 若 X 与 Y 相互独立,则

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

3.2 方差

3.2.1 方差和标准差的概念

定义 3.2.1 设 X 为随机变量,若 $E(X-E(X))^2$ 存在,则称 $E(X-E(X))^2$ 为随机变量 X 的方差,记为 $D(X)$ 或 $\mathrm{Var}(X)$,即

$$D(X)=E(X-E(X))^2. \tag{3.2.1}$$

称 $\sqrt{D(X)}$ 为 X 的标准差,记为 $\sigma(X)$,即 $\sigma(X)=\sqrt{D(X)}$.

由于 $\sigma(X)$ 的量纲与 X 的量纲相同,因而在工程技术领域中常用标准差 $\sigma(X)$. 记 $\mu=E(X)$,因为

$$(X-E(X))^2=(X-\mu)^2=X^2-2\mu X+\mu^2$$

根据数学期望的性质,有

$$E(X-E(X))^2=E(X-\mu)^2=E(X^2)-2\mu E(X)+\mu^2=E(X^2)-\mu^2,$$

所以

$$D(X)=E(X^2)-(E(X))^2. \tag{3.2.2}$$

3.2.2 方差的性质

(1) $D(C)=0$；

(2) $D(CX)=C^2D(X)$；

(3) $D(X\pm Y)=D(X)\pm 2E((X-E(X))(Y-E(Y))+D(Y)$；

特别地，若 X 与 Y 相互独立，则有

$$D(X\pm Y)=D(X)+D(Y).$$

(4) $D(X)=0$ 的充要条件是 $P\{X=E(X)\}=1$.

(5) 若 $X_1,X_2,\cdots,X_n$ 是相互独立的随机变量，则

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

3.3 协方差 相关系数和矩

3.3.1 协方差与相关系数

定义 3.3.1 设 (X,Y) 为二维随机变量，若 $E\{(X-E(X))(Y-E(Y))\}$ 存在，则称 $E\{(X-E(X))(Y-E(Y))\}$ 为 X 与 Y 的**协方差**，记为 $\mathrm{Cov}(X,Y)$，即

$$\mathrm{Cov}(X,Y)=E\{(X-E(X))(Y-E(Y))\}. \tag{3.3.1}$$

由上述定义可知，$\mathrm{Cov}(X,X)=E[(X-EX)^2]=D(X)$. 既然 $(X-EX)$ 与 $(X-EX)$ 之积的数学期望称为方差，现在把其中的一个 $(X-EX)$ 换成 $(Y-EY)$，由于其形式与方差类似，又是 X 与 Y 协同参与的结果，故称之为“协方差”.

把(3.3.1)式右端的各项展开，再利用数学期望的性质，可得到一个较实用的计算 $\mathrm{Cov}(X,Y)$ 的公式

$$\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y). \tag{3.3.2}$$

根据协方差的定义，不难验证协方差的下述性质.

(1) $\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$；

(2) $\mathrm{Cov}(aX,bY)=ab\mathrm{Cov}(X,Y)$，其中 a,b 为常数；

(3) $\mathrm{Cov}(X_1+X_2,Y)=\mathrm{Cov}(X_1,Y)+\mathrm{Cov}(X_2,Y)$.

定理 3.3.1 设 (X,Y) 为二维随机变量，

(1) 若 X 与 Y 相互独立，则

$$\mathrm{Cov}(X,Y)=0.$$

(3)若 $E(X^2)$、$E(Y^2)$ 存在，则有

$$E((XY)^2)\leqslant E(X^2)E(Y^2).$$

特别地，有

$$(\mathrm{Cov}(X,Y))^2\leqslant D(X)D(Y).$$

定义 3.3.2 称 $\dfrac{\mathrm{Cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$ 为 X 与 Y 的相关系数，记为 ρ_{XY}，即

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

相关系数具有下述性质：

(1) 若 X 与 Y 相互独立，则 $\rho_{XY}=0$；

(2) $|\rho_{XY}|\leqslant 1$；

(3) $|\rho_{XY}|=1$ 的充要条件是：存在常数 a,b 使得 $P\{Y=a+bX\}=1$.

3.3.2 矩

- $E(X^k)$——X 的 k 阶原点矩
- $E(X)$——X 的数学期望（一阶原点矩）
- $E(|X^k|)$——X 的 k 阶绝对原点矩
- $E((X-E(X))^k)$——X 的 k 阶中心矩
- $E((X-E(X))^2)=D(X)$——X 的方差（二阶中心矩）
- $E(X^kY^l)$——X，Y 的 k+l 阶混合原点矩
- $E((X-E(X))^k(Y-E(Y))^l)$——X，Y 的 k+l 阶混合中心距
- $E((X-E(X))(Y-E(Y)))$——X，Y 的协方差（二阶混合中心矩）

三、大数定律与中心极限定理

4.1 大数定律

✧ 设非负随机变量 X 的期望 $E(X)$ 存在，则对于任意实数 $\varepsilon>0$，有 $P(X\ge\varepsilon)\le\dfrac{E(X)}{\varepsilon}$。

✧ 马尔可夫不等式：设随机变量 X 的 k 阶绝对原点矩 $E(|X|^k)$ 存在，则对于任意实数

$\varepsilon>0$，有 $P(|X|\ge\varepsilon)\le\dfrac{E(|X|^k)}{\varepsilon^k}$。

✧ 切比雪夫不等式：设随机变量 X 的方差 $D(X)$ 存在，则对于任意实数 $\varepsilon>0$，有

$P(|X-E(X)|\ge\varepsilon)\le\dfrac{D(X)}{\varepsilon^2}$。（当 $\varepsilon^2\le D(X)$ 时，该不等式无实际意义）

✧ 设 X_1，X_2，$\cdots X_n,\cdots$ 为随机变量序列，a 为常数，若对任意 $\varepsilon>0$，有：

$\lim\limits_{n\to\infty}P\{|X_n-a|<\varepsilon\}=1$ 或 $\lim\limits_{n\to\infty}P\{|X_n-a|\ge\varepsilon\}=0$，则称 $\{X_n\}$ 依概率收敛于 a。

1. 伯努利大数定律

   设 η_n 是 n 重伯努利试验中 A 发生的次数，p 是事件 A 发生的概率，则对任意 $\varepsilon>0$，

   有：$\lim\limits_{n\to\infty}P\left\{\left|\dfrac{\eta_n}{n}-p\right|<\varepsilon\right\}=1$ 或 $\lim\limits_{n\to\infty}P\left\{\left|\dfrac{\eta_n}{n}-p\right|\ge\varepsilon\right\}=0$。

2. 切比雪夫大数定律

   设 X_1，X_2，$\cdots X_n,\cdots$ 是两两不相关的随机变量序列，分别存在数学期望 $E(X_i)$ 和方

差 $D(X_i)$，且方差一致有界，即存在某一常数 C，使 $D(X_i)<C$，对任意 $\varepsilon>0$，有：

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}E(X_i)\right|<\varepsilon\right\}=1 \text{ 或 } \lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}E(X_i)\right|\geq\varepsilon\right\}=0 。$$

推论：设 X_1，X_2，$\cdots X_n,\cdots$ 是相互独立的随机变量序列，且存在数学期望 $E(X_i)=\mu$，方差 $D(X_i)=\sigma^2$，则对任意 $\varepsilon>0$，有：

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right|<\varepsilon\right\}=1 \text{ 或 } \lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right|\geq\varepsilon\right\}=0 。$$

3. 辛钦大数定律（独立同分布随机变量的大数定律）

设 X_1，X_2，$\cdots X_n,\cdots$ 是独立同分布的随机变量序列，且数学期望 $E(X_i)=\mu$，则对任意 $\varepsilon>0$，有：$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right|<\varepsilon\right\}=1$ 或 $\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right|\geq\varepsilon\right\}=0$。

4.2 中心极限定理

1. 林德伯格-列维中心极限定理（独立同分布的中心极限定理）

设随机变量序列 X_1，X_2，$\cdots X_n,\cdots$ 独立同分布，且有数学期望和方差 $E(X_k)=\mu$，$D(X_k)=\sigma^2$，则有

$$\lim_{n\to\infty}P\left\{\frac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{n}\sigma}\leq x\right\}=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt=\Phi(x)。$$

标准化随机变量 $Y_n=\dfrac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{n}\sigma}$，总以标准正态分布 $N(0,1)$ 为其极限分布。

n 个独立随机变量的和 $\sum_{i=1}^{n}X_i=\sqrt{n}\sigma Y_n+n\mu$ 的极限分布是正态分布 $N(n\mu,n\sigma^2)$。

2. 德莫佛-拉普拉斯中心极限定理（二项分布以正态分布为极限分布）

设 η_n 是 n 重伯努利试验中 A 发生的次数，p 是事件 A 发生的概率，则对任意 $x\in R$，有：$\lim_{n\to\infty}P\left\{\frac{\eta_n-np}{\sqrt{np(1-p)}}\leq x\right\}=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt=\Phi(x)$

五、数理统计的基本概念

5.1 总体、样本与经验分布函数

1. 总体——研究对象全体元素组成的集合

 所研究的对象的某个(或某些)数量指标的全体,它是一个随机变量(或多维随机变量),记为 X。X 的分布函数和数字特征称为总体的分布函数和数字特征。

2. 个体——组成总体的每一个元素

 总体的每个数量指标,可看作随机变量 X 的某个取值,用 X_i 表示。

3. 样本——从总体中抽取的部分个体

 用 $(X_1, X_2, \cdots, X_n)$ 表示,n 为样本容量,称 $(x_1, x_2, \cdots, x_n)$ 为总体 X 的一个容量为 n 的样本观测值,或称样本的一个实现。

4. 样本空间——样本所有可能取值的集合

5. 简单随机样本

 (1) $X_1, X_2, \cdots, X_n$ 与 X 有相同的分布。

 (2) $X_1, X_2, \cdots, X_n$ 相互独立。

 一般地,对有限总体,放回抽样所得到的样本为简单随机样本,但使用不方便,常用不放回抽样代替,代替条件是 N/n≥10。

6. 分布函数

 (1) 设总体 X 的分布函数为 $F(x)$,则样本 $(X_1, X_2, \cdots, X_n)$ 的联合分布函数为:

 $$F_{总}(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} F(x_i)。$$

 (2) 若总体 X 是离散型随机变量,其分布律为 $P\{X = x_i\} = p(x_i)$,且样本 $(X_1, X_2, \cdots, X_n)$ 的观测值为 $(x_1, x_2, \cdots, x_n)$,则联合分布律为:

 $$P\{X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n\} = \prod_{i=1}^{n} p(x_i)。$$

 (3) 若总体 X 是连续型的随机变量,密度函数为 $f(x)$,则样本 $(X_1, X_2, \cdots, X_n)$ 的联合密度函数为:$f_{总}(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f(x_i)$。

5.1.2 经验分布函数

1. 样本分布:总体 X 是随机变量,其概率分布是客观存在的,常常用样本分布作为总体分布的近似,常用方法有:频数分布于频率分布(对离散型总体),频率直方图(对连续型总体),经验分布函数(对各类型总体)。

2. 经验分布函数

 设 $(x_1, x_2, \cdots, x_n)$ 是来自于总体 X 的样本值,将这些值按由小到大的顺序进行排列为

$x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$，对于任意的实数 x，定义函数 $F_n(x)$ 为总体 X 的经验分布函数。

$$F_n(x)=\begin{cases}0, x<x_{(1)} \\ k/n, x_{(k)} \le x < x_{(k+1)} \\ 1, x \ge x_{(n)}\end{cases}$$

3. 格里纹科定理
   当样本容量 n 充分大时，经验分布函数和总体分布函数最大的绝对偏差可以足够小，即当 n 充分大时，经验分布函数是总体分布函数的一个很好的近似。

   $P\{\sup|F_n(x)-F(x)| \to 0\}=1$

5.2 统计量及其数字特征

1. 设 $(X_1, X_2, \cdots, X_n)$ 是取自总体 X 的一个样本，$g(r_1, r_2, \cdots, r_n)$ 为一实值连续函数，且不含未知参数，则称随机变量 $g(X_1, X_2, \cdots, X_n)$ 为统计量，若 $(x_1, x_2, \cdots, x_n)$ 是一个样本值，称 $g(x_1, x_2, \cdots, x_n)$ 为统计量的一个样本值，统计量的分布称为抽样分布。

2. 常用的统计量

   设 $(X_1, X_2, \cdots, X_n)$ 是来自总体 X 的容量为 n 的样本，则有：

   (1) 样本均值：$\overline{X}=\dfrac{1}{n}\sum_{i=1}^{n} X_i$。

   (2) 样本方差：$S^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$。

   (3) 样本标准差：$S=\sqrt{\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2}$。

   (4) 样本的 k 阶原点矩：$A_k=\dfrac{1}{n}\sum_{i=1}^{n} X_i^k$。

   (5) 样本的 k 阶中心矩：$B_k=\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k$。

   $$\sum_{i=1}^{n}(X_i-\overline{X})^2=\sum_{i=1}^{n} X_i^{\,2}-n\overline{X}^2=n(A_2-\overline{X}^2)$$

   (6) 顺序统计量、样本极值与极差

   设 $(X_1, X_2, \cdots, X_n)$ 为样本，$(x_1, x_2, \cdots, x_n)$ 为样本值，将观测值按从小到大顺序排列，记为 $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$，定义 $X_{(k)}=x_{(k)}$，则有：

顺序统计量：$X_{(1)},X_{(2)},\cdots,X_{(n)}$。

样本极值：$X_{(n)}=\max\limits_{1\le k\le n}\{X_k\}$，$X_{(1)}=\min\limits_{1\le k\le n}\{X_k\}$。

样本极差：$D_n=X_{(n)}-X_{(1)}$。

(7) 样本中位数 M

$$M=\begin{cases}X_{(\frac{n+1}{2})},n\text{为奇数}\\ \frac{1}{2}(X_{(\frac{n}{2})}+X_{(\frac{n}{2}+1)}),n\text{为偶数}\end{cases}$$

(8) 样本 P 分位数

$M_p=X_{([(n+1)p])},0<p<1$

3. 定理：设总体 X 的数学期望 $E(X)=\mu$ 和方差 $D(X)=\sigma^2$ 存在，则有：

(1) $E(\overline{X})=\mu$，$D(\overline{X})=\dfrac{\sigma^2}{n}$，且当 $n\to\infty$ 时，$\overline{X}$ 依概率收敛于 μ。。

(2) 当 X 的四阶矩存在，$E(S^2)=\sigma^2$，S^2 依概率收敛于 σ^2。

4. 定理：设总体 X 的 k 阶矩 $\alpha_k=E(X^k)$ 存在，则有：

(1) $E(A_k)=\alpha_k$。

(2) 当 X 的 2k 阶矩存在，A_k 依概率收敛于 α_k。

5. 样本方差 S^2 与样本二阶中心矩 S_n^2

(1) $S_n^2=\dfrac{n-1}{n}S^2$。

(2) $E(S_n^2)=\dfrac{n-1}{n}\sigma^2$，$E(S^2)=\sigma^2$。

6. 数据的简单处理方法

5.3 抽样分布

1. 统计量的分布称为抽样分布。正态总体是最常见的总体，基于正态分布提出统计推断。除正态分布外，数理统计中常用的三大分布（连续型）：χ^2 分布、t 分布、F 分布。

2. U 分布——正态总体样本均值的分布

- 设 $(X_1,X_2,\cdots,X_n)$ 是正态总体 $X\sim N(\mu,\sigma^2)$ 的一个样本，$Y=\sum\limits_{i=1}^{n}c_iX_i$，则 Y 也服从

正态分布，且$E(Y)=\sum_{i=1}^{n} c_i\mu_i$， $D(Y)=\sum_{i=1}^{n} c_i^2\sigma_i^2$ 。

- 设$(X_1,X_2,\cdots,X_n)$是正态总体$X\sim N(\mu,\sigma^2)$的一个样本，样本均值也服从正态分布，即$\overline{X}\sim N(\mu,\frac{\sigma^2}{n})$， $U=\frac{\sqrt{n}\left(\overline{X}-\mu\right)}{\sigma}\sim N(0,1)$。
- 设$(X_1,X_2,\cdots,X_n)$和$(Y_1,Y_2,\cdots,Y_n)$是分别来自正态总体$N(\mu_1,\sigma_1^2)$和$N(\mu_2,\sigma_2^2)$的样本，且两个样本相互独立，则$U=\frac{\left(\overline{X}-\overline{Y}\right)-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}\sim N(0,1)$。

3. 概率分布的分位数（分位点）

- 设随机变量 X 的分布函数为$F(x)=P\{X\le x\}$，对于给定的$\alpha(0<\alpha<1)$，若存在使$P\{X>x_\alpha\}=1-F(x_\alpha)=\alpha$，则称$x_\alpha$为 X 分布的上侧$\alpha$分位数或上侧临界值。
- 若存在数λ_1和λ_2，使$P\{X\ge\lambda_1\}=P\{X\le\lambda_2\}=\frac{\alpha}{2}$，则称$\lambda_1$、$\lambda_2$为 X 分布的双侧$\alpha$分位数或双侧临界值，$\lambda_1=x_{\alpha/2}$、$\lambda_2=x_{1-\alpha/2}$。
- 当 X 的分布关于 y 轴对称时，若存在$x_{\alpha/2}$，使$P\{|X|\ge x_{\alpha/2}\}=\alpha$，则称$x_{\alpha/2}$为 X 分布的双侧$\alpha$分位数或双侧临界值。

4. χ^2分布

设总体$X\sim N(0,1)$，$(X_1,X_2,\cdots,X_n)$是 X 的一个样本，则称统计量$\chi^2=X_1^2+X_2^2\cdots+X_n^2$服从自由度为 n 的$\chi^2$分布，记作$\chi^2\sim\chi^2(n)$（自由度是指独立随机变量的个数）。$\chi^2(n)$分布的密度函数为（其图形随自由度的不同而不同）：

$$f(y)=\chi^2(y;n)=\begin{cases}\frac{1}{2^{n/2}\Gamma(n/2)}y^{n/2-1}e^{-y/2}, y\ge 0\\ 0, \quad y<0\end{cases}$$

- $\chi^2(1)$为标准正态分布，$\chi^2(2)$为参数为 1/2 的指数分布。
- $\chi^2(n)$分布的上侧α分位数：$P\{\chi^2(n)>\chi_\alpha^2(n)\}=\int_{\chi_\alpha^2(n)}^{\infty}f(y)dy=\alpha$。
- $\chi^2(n)$分布的双侧α分位数：$P\{\chi^2\le\chi_{1-\alpha/2}^2(n)\}=P\{\chi^2\ge\chi_{\alpha/2}^2(n)\}=\frac{\alpha}{2}$。
- $\chi^2(n)$分布的数学期望和方差：$E(\chi^2)=n$，$D(\chi^2)=2n$。

- $\chi^2(n)$ 分布的可加性 $Z_1 \sim \chi^2(n_1)$，$Z_2 \sim \chi^2(n_2)$，且 Z_1、Z_2 相互独立，则有 $Z_1 + Z_2 \sim \chi^2(n_1 + n_2)$，可以推广到多个随机变量的情形。
- 设随机变量 $(X_1, X_2, \cdots, X_n)$ 相互独立，且都服从标准正态分布，则随机变量 $\chi^2 = \sum_{i=1}^{n} X_i^2$，服从自由度为 n 的 χ^2 分布，即 $\chi^2 \sim \chi^2(n)$。
- 设 $(X_1, X_2, \cdots, X_n)$ 为取自正态总体 $X \sim N(\mu, \sigma^2)$ 的样本，则 $\dfrac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2} \sim \chi^2(n)$。
- 设 $(X_1, X_2, \cdots, X_n)$ 为取自正态总体 $X \sim N(\mu, \sigma^2)$ 的样本，则有：$\overline{X}$ 与 S^2 相互独立，$\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$。

5. t 分布

设随机变量 $X \sim N(0,1)$，$Y \sim \chi^2(n)$，且 X 与 Y 相互独立，则称统计量 $T = \dfrac{X}{\sqrt{\dfrac{Y}{n}}}$ 服从自由度为 n 的 t 分布或学生式分布，记作 $T \sim t(n)$。其密度函数为：

$$f(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\,\Gamma(\frac{n}{2})}(1 + \frac{x^2}{n})^{-\frac{n+1}{2}}。$$

- t 分布密度函数形状类似于标准正态分布，n 比较大时，t 分布近似于标准正态分布，但对较小的 n 值，t 分布与标准正态分布之间有较大差异，且 $P\{|T| \geq t_0\} = P\{|X| \geq t_0\}$。
- t 分布的数学期望和方差：$E(T) = 0$，$D(T) = \dfrac{n}{n-2}(n > 2)$。
- 设 $(X_1, X_2, \cdots, X_n)$ 为取自正态总体 $X \sim N(\mu, \sigma^2)$ 的样本，则有 $T = \dfrac{\overline{X} - \mu}{S / \sqrt{n}} \sim t(n-1)$。
- 设 $(X_1, X_2, \cdots, X_{n_1})$ 和 $(Y_1, Y_2, \cdots, Y_{n_2})$ 为取自正态总体 $N(\mu_1, \sigma^2)$，$N(\mu_2, \sigma^2)$ 的样本，且相互独立，则有：$T = \dfrac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{S_n \sqrt{1/n_1 + 1/n_2}} \sim t(n_1 + n_2 - 2)$，其中

$S_n = \sqrt{\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-1}}$ 。

- t 分布的上侧 α 分位数：$P\{T > t_\alpha(n)\} = \int_{t_\alpha(n)}^{\infty} f(t)dt = \alpha$ 。
- t 分布的双侧 α 分位数：$P\{|T| > t_{\alpha/2}(n)\} = \alpha$ 。

6. F 分布

设随机变量 $X \sim \chi^2(n_1)$，$Y \sim \chi^2(n_2)$，且 X 与 Y 相互独立，则称统计量 $F = \dfrac{X/n_1}{Y/n_2}$ 服从第一自由度为 n_1，第二自由度为 n_2 的 F 分布。记作 $F \sim F(n_1,n_2)$。其密度函数略（见书本）。$X \sim F(n_1,n_2)$，则 $\dfrac{1}{\mathrm{X}} \sim F(n_2,n_1)$。

- F 分布的上侧 α 分位数：$P\{F(n_1,n_2) > F_\alpha(n_1,n_2)\} = \int_{F_\alpha(n_1,n_2)}^{\infty} f(y)dy = \alpha$ 。
- F 分布的双侧 α 分位数：$P\{F \le F_{1-\alpha/2}(n_1,n_2)\} = P\{F \ge F_{\alpha/2}(n_1,n_2)\} = \dfrac{\alpha}{2}$ 。
- 由 $X \sim F(n_1,n_2)$，则 $\dfrac{1}{\mathrm{X}} \sim F(n_2,n_1)$，有：$F_\alpha(n_1,n_2) = \dfrac{1}{F_{1-\alpha}(n_2,n_1)}$ 。
- 设 n_1，S_1^2 为正态总体 $N(\mu_1,\sigma_1^2)$ 的样本容量和样本方差，n_2，S_2^2 为正态总体 $N(\mu_2,\sigma_2^2)$ 的样本容量和样本方差，且两个样本相互独立，则有：$\dfrac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1-1,n_2-1)$ 。

六、参数估计

6.1 参数的点估计——估计未知参数的值

1. 参数是刻画总体某概率特性的数量，当此数量未知时，从总体抽出一系列样本，用某种方法对这个未知参数进行估计就是参数估计。
2. 点估计的思想方法：设总体 X 的分布函数形式已知，但含有一个或多个未知参数：$\theta_1,\theta_2,\cdots,\theta_k$，设 $X_1,X_2,\cdots,X_n$ 为总体的一个样本，构造 k 个统计量：$\theta_1(X_1,X_2,\cdots,X_n)\cdots$ $\theta_k(X_1,X_2,\cdots,X_n)$（估计量），代入样本观测值得到 k 个数（估计值）：$\hat{\theta}_1(x_1,x_2,\cdots,x_n)\cdots$ $\hat{\theta}_k(x_1,x_2,\cdots,x_n)$。
3. 频率替换法

   利用事件 A 在 n 次试验中发生的频率 n_A/n 作为事件 A 发生概率 p 的估计量。

4. 矩估计法
   (1) 用样本的 k 阶原点矩作为总体的 k 阶矩的估计量，建立含有待估参数的方程（组），从而解出待估参数。定理：设总体 X 的 k 阶矩 $\alpha_k = E(X^k)$ 存在，则有：$E(A_k) = \alpha_k$；

当 X 的 2k 阶矩存在，A_k 依概率收敛于 α_k。

(2) 设待估计参数为 $\theta_1,\theta_2,\cdots,\theta_k$，设总体的 r 阶原点矩存在，记为 $E(X^r)=\alpha_r(\theta_1,\theta_2,\cdots,\theta_k)$，样本 $X_1,X_2,\cdots,X_n$ 的 r 阶原点矩为 $A_r=\frac{1}{n}\sum_{i=1}^{n}X_i^r$，令 $\alpha_r(\theta_1,\theta_2,\cdots,\theta_k)=\frac{1}{n}\sum_{i=1}^{n}X_i^r$，得到含未知参数的方程（组），解得 k 个统计量 $\hat{\theta}_1(X_1,X_2,\cdots,X_n)\cdots\hat{\theta}_k(X_1,X_2,\cdots,X_n)$（未知参数的矩估计量），代入样本观测值得 k 个数：$\hat{\theta}_1(x_1,x_2,\cdots,x_n)\cdots\hat{\theta}_k(x_1,x_2,\cdots,x_n)$（未知参数的矩估计值）。

(3) 一般地，不论总体服从什么分布，总体期望 μ 与方差 σ^2 存在时，则有 $\hat{\mu}=\overline{X}$、$\hat{\sigma}^2=A_2-\hat{\mu}^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2=B_2$。

5. 极大似然估计法

(1) 思想方法：一次试验就出现的事件有较大的概率。

(2) 一般地，设 X 为离散型随机变量，其分布律为 $P(X=x)=f(x,\theta)$，则样本 $X_1,X_2,\cdots,X_n$ 的联合概率分布为 $P(X_1=x_1,\cdots,X_n=x_n)=f(x_1,\theta)\cdots f(x_n,\theta)$，记作 $L(x_1,x_2,\cdots,x_n,\theta)$ 或 $L(\theta)$，称 $L(\theta)$ 为样本的似然函数。

✧ 若 X 连续，取 $f(x_i,\theta)$ 为 X_i 的密度函数，则样本的似然函数为 $L(\theta)=\prod_{i=1}^{n}f(x_i,\theta)$。

✧ 若有多个未知参数 $\theta_1,\theta_2,\cdots,\theta_k$，则可设 X 的密度（或分布）为 $f(x,\theta_1,\cdots,\theta_k)$，则样本似然函数为 $L(\theta_1,\theta_2,\cdots,\theta_k)=\prod_{i=1}^{n}f(x_i,\theta_1,\theta_2,\cdots,\theta_k)$。

(3) 极大似然法的思想：选择适当的 $\theta=\hat{\theta}$ 使 $L(\theta)$ 取最大值（令导数或取对数后导数为零得似然方程组），即 $L(x_1,x_2,\cdots,x_n,\hat{\theta})=\max\{f(x_1,\theta)f(x_2,\theta)\cdots f(x_n,\theta)\}$，称这样得到的 $\hat{\theta}=g(x_1,x_2,\cdots,x_n)$ 为参数 θ 的极大似然估计值，称统计量 $\hat{\theta}=g(X_1,X_2,\cdots,X_n)$ 参数 θ 的极大似然估计量。

(4) 极大似然法估计参数的步骤

① 写出似然函数（联合分布函数）：

$$L(x_1,x_2,\cdots,x_n;\theta_1,\theta_2,\cdots,\theta_k)=\prod_{i=1}^{n}f(x_i,\theta_1,\theta_2,\cdots,\theta_k)$$

② 求出 $\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_k$，使似然函数极大，即得到参数估计值 $\hat{\theta}_r=g(x_1,x_2,\cdots,x_n)$。

③ 可进一步得到估计量 $\hat{\theta}_r=g(X_1,X_2,\cdots,X_n)$。

(5) 极大似然估计的不变性

设 $\hat{\theta}$ 是 θ 的极大似然估计值，$u(\theta)$ 是 θ 的函数，且有单值反函数（单调），则 $\hat{u}=u(\hat{\theta})$ 是 $u(\theta)$ 的极大似然估计值。

6.2 估计量的评价标准

对于同一个未知参数，不同方法得到的估计量可能不同，则需要一定的标准来衡量某个估计量的优劣：无偏性（有无系统性误差）、有效性（波动性的大小）、一致性（当样本容量增大时估计值是否越来越精确）。

1. 无偏性

(1) 设 $\hat{\theta}(X_1,X_2,\cdots,X_n)$ 是未知参数 θ 的估计量，若 $E(\hat{\theta})=\theta$（估计值的期望与真值相等），则称 $\hat{\theta}$ 是 θ 的无偏估计量；若 $E(\hat{\theta})\neq\theta$，但 $\lim\limits_{n\to\infty}E(\hat{\theta})=\theta$，则称 $\hat{\theta}$ 是 θ 的渐近无偏估计量。

(2) 对任意总体，样本均值是总体期望的无偏估计量，样本方差是总体方差的无偏估计量，样本 k 阶原点矩是总体 k 阶原点矩的无偏估计量。

(3) 对任意总体，样本均值平方是总体期望平方的渐近无偏估计量，样本二阶中心矩是总体方差的渐近无偏估计量。

2. 有效性

(1) 设 $\hat{\theta}_1(X_1,X_2,\cdots,X_n)$ 和 $\hat{\theta}_2(X_1,X_2,\cdots,X_n)$ 都是对总体中某一未知参数 θ 的无偏估计量，若 $D(\hat{\theta}_1)<D(\hat{\theta}_2)$，则称估计量 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 更有效。

(2) 设 $(X_1,X_2,\cdots,X_n)$ 是来自总体 X 的样本，其中总体均值 $E(X)=\mu$ 与方差 $D(X)=\sigma^2$ 都存在，则样本均值 $\overline{X}$ 在 μ 的所有线性无偏估计量中是最有效的（称为最小方差线性无偏估计量）。

3. 一致性（相合性）

设 $\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$ 是总体参数 θ 的估计量，若对于任意的 $\theta\in\Theta$，当 $n\to\infty$ 时，$\hat{\theta}$ 依概率收敛于 θ，则称 $\hat{\theta}$ 为 θ 的一致（相合）估计量。若当 $n\to\infty$ 时，$\hat{\theta}$ 均方收敛于 θ，则称 $\hat{\theta}$ 为 θ 的均方相合估计量。

- 样本均值和样本方差分别是总体期望和总体方差的一致估计量，样本 k 阶原点矩既是总体 k 阶原点矩的一致估计量，又是其均方相合估计量。
- 设 $\hat{\theta}$ 是 θ 的无偏估计量，且 $\lim\limits_{n\to\infty} D(\hat{\theta})=0$，则 $\hat{\theta}$ 是 θ 的一致估计量。
- 矩法估计得到的估计量一般为一致估计量；在一定条件下，极大似然估计可具有一致性。

6.3 参数的区间估计——估计未知参数的取值范围

1. 双侧区间估计

   由不同样本求得的参数估计值不同，希望根据所给样本确定一个随机区间，使其包含未知参数真值的概率达到指定要求。

2. 置信区间

   设 θ 为待估参数，$\alpha(0<\alpha<1)$ 是一个给定的数，若能找到两个统计量 T_1、T_2，使得

   $P(T_1\le\theta\le T_2)=1-\alpha$，则称 $[T_1,T_2]$ 为 θ 置信水平为 $1-\alpha$ 的置信区间。

- 置信区间的长度 T_2-T_1 反映了估计精度，T_2-T_1 越小，估计精度越高。
- α 反映估计的可靠度，α 越小，$1-\alpha$ 即越大，估计的可靠度越高，但同时 T_2-T_1 往往增大，即估计精度降低。
- α 确定以后，置信区间的选取方法不唯一，常选择区间长度最小的一个。

3. 求置信区间的步骤

   (1) 寻找一个样本函数 $g(\theta,X_1,X_2,\cdots,X_n)$（称为枢轴量），使其含有待估参数，不含其他未知参数，且分布已知，且分布不依赖于待估参数（常由 θ 的点估计出发考虑）。

   (2) 给定置信度 $1-\alpha$，定出常数 a，b 使得 $P(a<g(\theta,X_1,X_2,\cdots,X_n)<b)=1-\alpha$。

   (3) 由 $a<g(\theta,X_1,X_2,\cdots,X_n)<b$，解得置信区间 (T_1,T_2)。

   (4) 代入样本观测值，得到具体的置信区间。

4. 一个正态总体参数的区间估计

   (1) 均值 μ 的置信区间

   ① σ^2 已知：$U=\dfrac{\sqrt{n}(\overline{X}-\mu)}{\sigma}\sim N(0,1)$

   ② σ^2 未知：$T=\dfrac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$

   (2) 方差 σ^2 的置信区间

   ① μ 已知：$\chi^2=\dfrac{1}{\sigma^2}\sum\limits_{i=1}^{n}\left(X_i-\mu\right)^2\sim\chi^2(n)$

   ② μ 未知：$\chi^2=\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$

5. 两个正态总体参数的区间估计

(1) 均值差 $\mu_1-\mu_2$ 的置信区间

① σ_1^2、σ_2^2 已知：$U=\dfrac{\left(\overline{X}-\overline{Y}\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\sim N(0,1)$

② $\sigma_1^2=\sigma_2^2=\sigma^2$ 未知：$T=\dfrac{\left(\overline{X}-\overline{Y}\right)-\left(\mu_1-\mu_2\right)}{S_w\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim t(n_1+n_2-2)$

$$S_w=\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$$

③ σ_1^2、σ_2^2 未知，且 $\sigma_1^2\neq\sigma_2^2$：$U=\dfrac{\left(\overline{X}-\overline{Y}\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\dfrac{S_{1n_1}^2}{n_1}+\dfrac{S_{2n_2}^2}{n_2}}}\sim N(0,1)$

④ σ_1^2、σ_2^2 未知，但 $n_1=n_2$，令 $Z_i=X_i-Y_i$，则 $Z\sim N(\mu_1-\mu_2,\sigma_1^2+\sigma_2^2)$

$S_Z^2=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}[(X_i-Y_i)-(\overline{X}-\overline{Y})]^2$，取 $T=\dfrac{\left(\overline{X}-\overline{Y}\right)-\left(\mu_1-\mu_2\right)}{S_Z/\sqrt{n}}\sim t(n-1)$

(2) 方差比 $\dfrac{\sigma_1^2}{\sigma_2^2}$ 的置信区间

① μ_1、μ_2 已知，$F=\dfrac{\dfrac{1}{n_1}\dfrac{\sum\limits_{i=1}^{n_1}(X_i-\mu_1)^2}{\sigma_1^2}}{\dfrac{1}{n_2}\dfrac{\sum\limits_{i=1}^{n_2}(Y_i-\mu_2)^2}{\sigma_2^2}}=\dfrac{\dfrac{n_2}{n_1}\dfrac{\sum\limits_{i=1}^{n_1}(X_i-\mu_1)^2}{\sum\limits_{i=1}^{n_2}(Y_i-\mu_2)^2}}{\dfrac{\sigma_1^2}{\sigma_2^2}}\sim F(n_1,n_2)$

② μ_1、μ_2 未知，$F=\dfrac{S_{1n_1}^2/S_{2n_2}^2}{\sigma_1^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$

七、假设检验

7.1 假设检验的基本概念

1. 在总体的分布函数完全未知或只知其形式、但不知其参数的情况下，为了推断总体的某些性质，对总体提出某项统计假设（假设），假设是否正确需要利用样本（采样数据）对假设的真假进行判断，即假设检验。

2. 当总体的分布类型已知，对总体分布中未知参数的检验称为参数假设检验，当总体分布类型未知，对总体分布类型或分布性质的检验称为非参数假设检验。

3.　参数假设检验：对总体分布的某个参数（通常是总体的某个数字特征）提出假设，用总体的样本值检验该假设是否成立。

7.1.1 假设检验的基本原理

要给出检验假设 H_0 的一个检验法，主要是在 H_0 成立的前提下找到一个适当的小概率事件，按照实际推断原理，在一次具体采样后如果这个小概率事件发生了，就拒绝原假设 H_0，如果这个小概率事件没有发生，就接受原假设 H_0。根据实际问题事先给定一个显著性水平α，事件发生的概率不超过α时，就认为是一个小概率事件。

7.1.2 两类错误

第一类错误（拒真错误）：原假设 H_0 为真，但拒绝了 H_0。

第二类错误（存伪错误）：原假设 H_0 不真，但接受了 H_0。

7.1.3 原假设和备择假设的不平等性

H_0 和 H_1 地位不平等，H_0 受特殊保护，因此接受了 H_0，并不一定说明 H_0 一定为真。因为有较大概率实际不真也误认为是真的。但是若在很小的α下仍拒绝 H_0，则有充分理由否定 H_0。接受 H_0 只说明目前样本提供的信息还不足以否定 H_0。

7.1.4 假设检验的一般步骤

1.　恰当地提出原假设 H_0 和备择假设 H_1。

2.　构造检验统计量 Z，并求出在 H_0 成立的前提下 Z 的概率分布，要求 Z 的分布不依赖于任何未知参数。

3.　确定拒绝域。在 H_0 成立时，以不利于 H_0（亦即有利于 H_1）的方式设定拒绝域的形式 W，再根据给定的水平α和 Z 的分布，由 $P\{W \mid H_0\text{为真}\} = \alpha$ 确定拒绝域 W。

4.　进行一次抽样，根据得到的样本值与上面确定的拒绝域 W，对 H_0 作出拒绝或接受的判断。

7.2 正态总体参数的假设检验

7.2.1 单个正态总体参数的假设检验

1.　正态总体 $N(\mu,\sigma^2)$ 中均值 μ 的检验

(1)　σ^2 已知——u 检验：$U = \dfrac{\sqrt{n}(\overline{X} - \mu_0)}{\sigma} \sim N(0,1)$

(2)　σ^2 未知——t 检验：$T = \dfrac{\overline{X} - \mu_0}{S / \sqrt{n}} \sim t(n-1)$

2. 正态总体$N(\mu,\sigma^2)$中方差σ^2的检验——χ^2检验检验

(1) μ已知：$\chi^2=\dfrac{1}{\sigma_0^2}\sum_{i=1}^{n}(X_i-\mu)^2\sim\chi^2(n)$

(2) μ未知：$\chi^2=\dfrac{(n-1)S^2}{\sigma_0^2}\sim\chi^2(n-1)$

7.2.2 两个正态总体参数的假设检验

1. 两总体均值差$\mu_1-\mu_2$的检验——t 检验

(1) σ_1^2、σ_2^2已知：$U=\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\sim N(0,1)$

(2) $\sigma_1^2=\sigma_2^2=\sigma^2$未知：$T=\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim t(n_1+n_2-2)$

$$S_w=\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$$

(3) σ_1^2、σ_2^2未知，且$\sigma_1^2\neq\sigma_2^2$：$U=\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_{1n_1}^2}{n_1}+\dfrac{S_{2n_2}^2}{n_2}}}\sim N(0,1)$

(4) σ_1^2、σ_2^2未知，但$n_1=n_2$，令$Z_i=X_i-Y_i$，则$Z\sim N(\mu_1-\mu_2,\sigma_1^2+\sigma_2^2)$

$S_Z^2=\dfrac{1}{n-1}\sum_{i=1}^{n}[(X_i-Y_i)-(\overline{X}-\overline{Y})]^2$，取$T=\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_Z/\sqrt{n}}\sim t(n-1)$

2. 两总体方差比$\dfrac{\sigma_1^2}{\sigma_2^2}$的检验——F 检验

(1) μ_1、μ_2已知，$F=\dfrac{\dfrac{1}{n_1}\dfrac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2}{\sigma_1^2}}{\dfrac{1}{n_2}\dfrac{\sum_{i=1}^{n_2}(Y_i-\mu_2)^2}{\sigma_2^2}}=\dfrac{\dfrac{n_2}{n_1}\dfrac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2}{\sum_{i=1}^{n_2}(Y_i-\mu_2)^2}}{\dfrac{\sigma_1^2}{\sigma_2^2}}\sim F(n_1,n_2)$

(2) μ_1、μ_2未知，$F=\dfrac{S_{1n_1}^2/S_{2n_2}^2}{\sigma_1^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$

3. 双边假设检验与单边假设检验

(1) 双边假设检验

$H_0:\theta=\theta_0$、$H_1:\theta\neq\theta_0$

(2) 单边假设检验

① $H_0:\theta=\theta_0$、$H_1:\theta>\theta_0$。

② $H_0:\theta=\theta_0$、$H_1:\theta<\theta_0$。

③ $H_0:\theta\le\theta_0$、$H_1:\theta>\theta_0$。

④ $H_0:\theta\ge\theta_0$、$H_1:\theta<\theta_0$。

7.2.3 大样本总体参数的假设检验