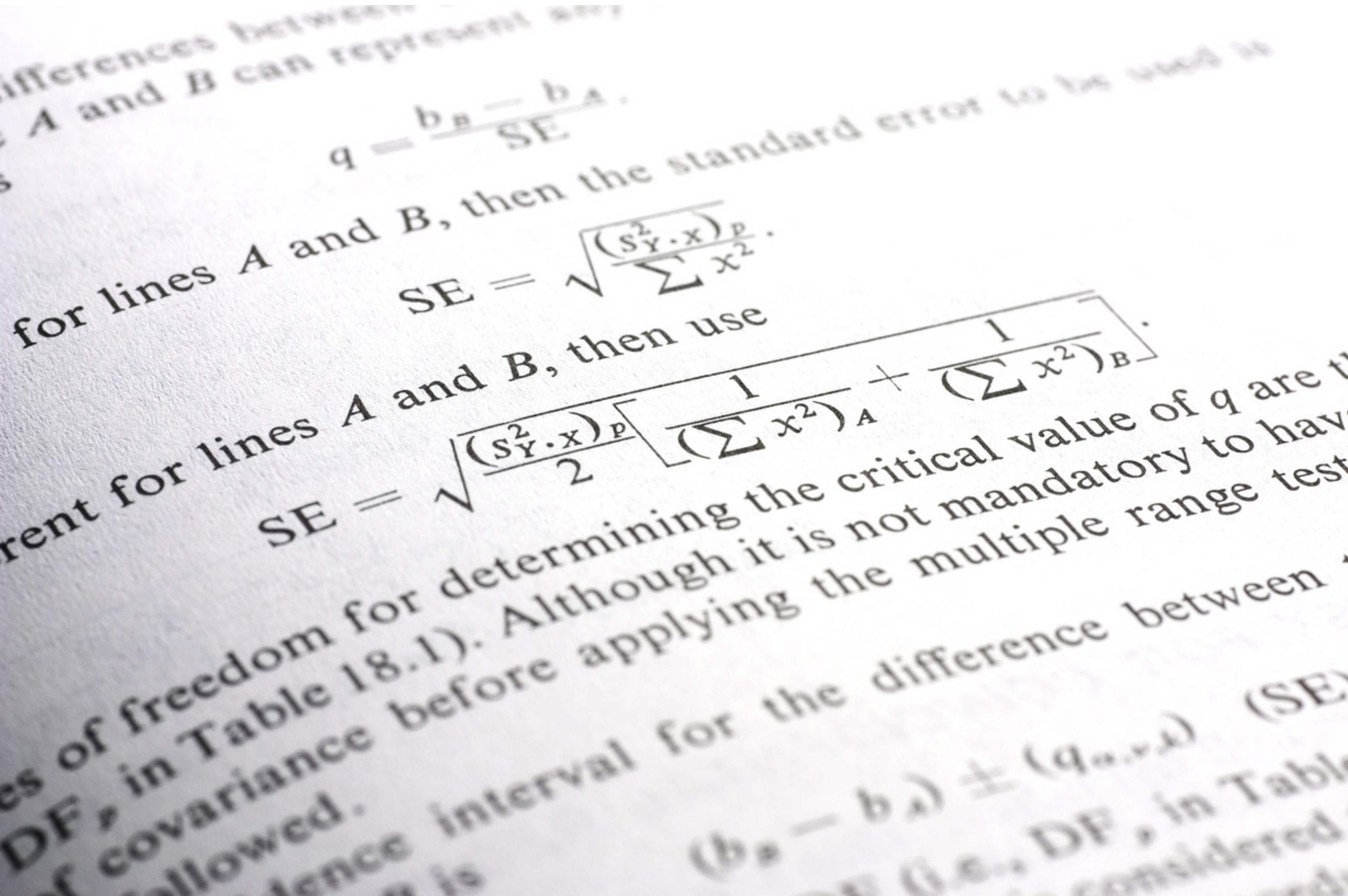
概率论与随机过程

2022秋

期末小助手

仲英学业辅导中心出品



前言

历时一年,概率论与数理统计小助手的编写终于完成.作为新手的我们对LaTeX并不熟悉,在编写过程中遇到了诸多棘手的问题,最终在我们的不懈努力下被一一解决.在这份小助手中,我们总结了原书中每章的知识点,并准备了相应的例题来辅助大家理解它们(其中原书第九章方差分析不在考试范围中).作为编者的我们非常感谢您能够选择使用这份小助手来辅助您的期末复习,同时也诚挚地希望我们精心编写的小助手可以切实地帮助到您.当然,优异成绩的背后少不了深刻的理解与充分的练习,我们也衷心祝愿您能在最终的考试中取得满意的成绩.在初版中,难免出现一些被我们忽视掉的纰漏,如有错误欢迎大家批评指正.

仲英学业辅导中心

2023年2月

目录

第一章　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

   (a) 可以在相同条件下重复进行.

   (b) 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.

   (c) 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 $S(\Omega)$.

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.

   (a) 基本事件: 一个样本点组成的单点集.

(b) 必然事件: 如$S(\Omega)$.

(c) 不可能事件: 如$\varnothing$.

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. $A-B=A\bar{B}$, $A-B=A-AB$.

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

1.2.1　概率的古典定义

1. 等可能概型(古典概型)

(a) 有限性: 试验的样本空间只包含有限个元素.

(b) 等可能性：试验中每个基本事件发生的可能性相同.

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

$$P\left(\{e_1\}\right)=\frac{1}{n}, i=1,2,\cdots,n.$$

2. 等可能事件中事件A概率的计算公式:

$$P\left(A\right)=\sum_{j=1}^{k}P\left(\{e_1\}\right)=\frac{k}{n}=\frac{A\text{中包含的基本事件数}}{S\text{中基本事件的总数}}.$$

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}, $\bar{A}=$ {摸出的球中没有红球}; 请计算A及$\bar{A}$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($r \le m$), 令$A_i=$ {摸出的r个球中有i个红球}, $i=0,1,2,\cdots,r$, 请计算A_i的概率, 并观察你计算的结果, 会得出什么数学公式?

$$C_n^r = \sum_{i=0}^{r} C_m^i C_{n-m}^{r-i}.$$

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.

$$P(A) = \frac{A\text{的几何度量}}{S\text{的几何度量}} = \frac{L(A)}{L(S)}.$$

   - 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. $\dfrac{n_A}{n}$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
   (a) 非负性: $0 \le f_n(A) \le 1$.
   (b) 规范性: $f_n(\Omega) = 1$.
   (c) 可列可加性: 若$A_k(k=1,2,\cdots,n)$互不相容, 则$f_n(A_1 \cup A_2 \cup \cdots A_n) = f_n(A_1) + f_n(A_2) + \cdots + f_n(A_n)$.

4. 频率不等同于概率. 由伯努利大数定律, 当n趋向于无穷大的时候, 频率$f_n(A)$在一定意义下接近于概率$P(A)$.

5. 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$\frac{m}{n}$, 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. 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$P(A)$满足下列三条公理, 则称$P(A)$为事件A的概率:

   (a) 非负性: $0 \leq f_n(A) \leq 1$.

   (b) 规范性: $f_n(\Omega) = 1$.

   (c) 可列可加性: 若$A_k(k = 1, 2, \cdots, n)$互不相容, 则$f_n(A_1 \cup A_2 \cup \cdots A_n) = f_n(A_1) + f_n(A_2) + \cdots + f_n(A_n)$.

1.2.4 概率的性质

1. $P(\varnothing) = 0$.

2. **(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:

$$f_n(A_1 \cup A_2 \cup \cdots A_n) = f_n(A_1) + f_n(A_2) + \cdots + f_n(A_n).$$

3. 设A, B是两个事件, 若$A \subseteq B$, 则有

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

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

   推论: 对于任意事件A, 有$P(B - A) = P(B) - P(AB)$.

4. 对于任一事件A, 有

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

5. **(逆事件的概率)** 对于任一事件A, 有

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

6. **(加法公式)** 对于任意两事件A, B, 有

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

推广:

$P\left(A_1 \cup A_2 \cup A_3\right)$

$= P\left(A_1\right) + P\left(A_2\right) + P\left(A_3\right) - P\left(A_1 A_2\right) - P\left(A_1 A_3\right) - P\left(A_2 A_3\right) + P\left(A_1 A_2 A_3\right).$

$P\left(A_1 \cup A_2 \cup \cdots \cup A_n\right)$

$$= \sum_{i=1}^{n} P(A_i) - \sum_{1 \le i < j \le n} P(A_i A_j) + \sum_{1 \le i < j < k \le n} P(A_i A_j A_k) + \cdots + (-1)^{n-1} P(A_1 A_2 \cdots A_n).$$

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:

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

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

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($P(A) > 0$)的条件下, 事件B发生的概率称为条件概率.

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

3. 乘法定理: 若$P(A)>0$, 则有

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

设A,B,C为事件, 且$P(AB)>0$, 则有

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

设$A_1,A_2,\cdots,A_n$为n个事件, $n\geq 2$, 且$P(A_1A_2\cdots A_{n-1})>0$, 则有

$$P(A_1A_2\cdots A_{n-1}A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1}).$$

1.3.2 全概率公式和贝叶斯公式

1. 样本空间的划分

定义 1.1. 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

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

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

则称$B_1,B_2,\cdots,B_n$为样本空间S的一个划分. 又称为互斥完备事件组.

2. 全概率公式

定理 1.2. 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

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

3. 贝叶斯公式(后验公式)

定理 1.3. 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

$$P(B_i|A)=\frac{P(A|B_i)P(B_i)}{\sum\limits_{j=1}^{n}P(A|B_j)P(B_j)},i=1,2,\cdots,n.$$

常用形式:

$$P(A)=P(A|B)P(B)+P(A|\bar{B})P(\bar{B});$$
$$P(B|A)=\frac{P(AB)}{P(A)}=\frac{P(A|B)P(B)}{P(A|B)P(B)+P(A|\bar{B})P(\bar{B})}.$$

1.4　随机事件的独立性

1.4.1　两个事件的独立性

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$P(AB)=P(A)P(B)$. 由此, 引出下述定义:

定义 1.4. *若两事件A, B满足*

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

则称A, B相互独立, 简称A, B独立.

需要注意的是:

- 概率为0或1的事件与任意事件相互独立(也可通过下一条证明);
- 当$P(A)P(B)\neq 0$时,

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

- 在实际问题中, 大多并不用这一定义式判断两事件A, B是否独立, 而是从试验的具体条件以及事件的本质去判断它们不应有关联, 从而是独立的, 然后就可以计算积事件的概率.

1.4.2　多个事件的独立性

由事件独立性的概念与性质, 可推广到多个事件独立性的情形:

1. 多个事件独立性的概念

   定义 1.5. *设A_1, A_2, $\cdots$, A_n是n个事件, 若对任意$k(1<k\leq n)$, 任意$1\leq i_1<i_2<\cdots<i_k\leq n$,*

   $$P\left(A_{i_1}A_{i_2}\cdots A_{i_k}\right)=P\left(A_{i_1}\right)P\left(A_{i_2}\right)\cdots P\left(A_{i_k}\right)$$

   都成立, 则称事件$A_1, A_2, \cdots, A_n$相互独立.

注意: 上述定义式中, 包含的等式总和为:

$$C_n^2+C_n^3+\cdots+C_n^n=\sum_{k=0}^{n}C_n^k-C_n^1-C_n^0=2^n-n-1$$

这表明:

- 若$A_1,A_2,\cdots,A_n$相互独立, 那么其中的任意$k(1<k\leq n)$个事件也相互独立. 由于等式数目(体现了判断条件是否充足)不够, 该命题在一般情况下($2^n-n-1\neq k$)不能反推.
- 特别地, 当$k=2$时, 它们之中任意两个事件都独立(称为$A_1,A_2,\cdots,A_n$两两独立). 但是, 为$A_1,A_2,\cdots,A_n$不能保证为$A_1,A_2,\cdots,A_n$相互独立.

2. 事件独立性的定理、性质和推论

定理 1.6. *对于两事件A和B, 若四对事件$\{A,B\},\{A,\bar{B}\},\{\bar{A},B\},\{\bar{A},\bar{B}\}$, 则另外三对也是相互独立的.*

类似地, 也可推广到n个相互独立事件的情形:

定理 1.7. *设n个事件$A_1,A_2,\cdots,A_n$相互独立, 则把其中任意$m(1\leq m\leq n)$个事件相应地换成它们的对立事件, 所得的n个事件仍相互独立.*

该定理的证明可用数学归纳法完成: 对所含对立事件个数进行归纳.

若n个事件相互独立, 则有$P\left(\bigcup\limits_{i=1}^{n}A_i\right)=1-P\left(\bigcap\limits_{i=1}^{n}\bar{A}_i\right)$.

1.5　经典例题

例 1.1. *设A、B是两个随机事件, $P\left(A|B\right)=0.4$, $P\left(B|A\right)=0.4$, $P\left(\bar{A}|\bar{B}\right)=0.7$, 求$P\left(A+B\right)$.*

解: 变形得$P\left(\bar{A}|\bar{B}\right)=0.7=\dfrac{P\left(\bar{A}\bar{B}\right)}{1-P(B)}=\dfrac{1-P(A+B))}{1-P(B)}$,

因为$P\left(A|B\right)=0.4$, $P\left(B|A\right)=0.4$,

所以$P(A)=P(B)$, $P(AB)=0.4P(A)$,

$P\left(\bar{A}|\bar{B}\right)=0.7=\dfrac{1-P(A)-P(B)+P(AB)}{1-P(B)}=\dfrac{1-P(A+B)}{1-P(B)}=\dfrac{1-2P(A)+0.4P(A)}{1-P(A)}$,

解得$P(A)=P(B)=\dfrac{1}{3}$, $P(AB)=\dfrac{1}{15}$,

于是$P(A+B)=P(A)+P(B)-P(AB)=\dfrac{8}{15}$.

例 1.2. $P(A)=0.6$, $P\left(A\bar{B}\right)=0.2$, $P\left(\bar{A}B\right)=0.3$, 求$P\left(A+\bar{B}|\bar{A}\right)$.

解: 由$P\left(A\bar{B}\right)=P(A-B)=P(A)-P(AB)=0.2$, $P(A)=0.6$,

可得$P\left(AB\right)=0.4$,

再由$P\left(\bar{A}B\right)=P(B-A)=P(B)-P(AB)=0.3$,

可得$P(B)=0.7$,

所以 $P\left(A+\bar{B}|\bar{A}\right)=\dfrac{P\left(\bar{A}\left(A+\bar{B}\right)\right)}{P\left(\bar{A}\right)}=\dfrac{P\left(\bar{A}\bar{B}\right)}{P\left(\bar{A}\right)}=\dfrac{1-P(A+B)}{1-P(A)}=\dfrac{1-P(A)-P(B)+P(AB)}{1-P(A)}=\dfrac{1}{4}$.

例 1.3. 设事件A, B, C两两独立, 则事件A, B, C相互独立的充要条件是(　　).

A. A与BC相互独立

B. AB与$A+C$相互独立

C. AB与AC相互独立

D. $A+B$与$A+C$相互独立

解: 在A, B, C两两独立的情况下,

A, B, C相互独立 $\Longleftrightarrow P(ABC)=P(A)P(B)P(C)\Longleftrightarrow P(A)P(B)P(C)=P(A)P(BC)$.

所以正确答案为A.

例 1.4. 将编号为1, 2, 3的三本书随意排列在书架上, 求至少有一本书从左到右排列的序号与它的编号相同的概率.

解: 设 $A_i=$第i本书正好在第i个位置,

$B=$至少有一本书从左到右排列的序号与它的编号相同,

则$B=A_1+A_2+A_3$,

且$P\left(A_i\right)=\dfrac{2!}{3!}(i=1,2,3)$, $P\left(A_iA_j\right)=\dfrac{1!}{3!}=\dfrac{1}{6}(i,j=1,2,3,i\neq j)$,

$P\left(A_1A_2A_3\right)=\dfrac{1!}{3!}=\dfrac{1}{6}$,

故$P(B)=P(A_1)+P(A_2)+P(A_3)-P(A_1A_2)-P(A_1A_3)-P(A_2A_3)+P(A_1A_2A_3)=\dfrac{2}{3}$.

第二章　随机变量及其概率分布

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　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. 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$\{\omega | X(\omega) \le x\}$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。

   定义 2.2. 设X是一个随机变量，x是任意实数，函数

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

   称为X的**分布函数**。

   对于任意实数x_1, x_2 $(x_1 < x_2)$, 有

   $$P\{x_1 < X \le x_2\} = P\{X \le x_2\} - P\{X \le x_1\} = F(x_2) - F(x_1)$$

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.

性质 2.3 (分布函数的基本性质). *(a)* $F(x)$是一个不减函数。对于任意实数x_1,x_2 $(x_1<x_2)$, 有

$$F(x_2)-F(x_1)=P\left\{x_1<X\leq x_2\right\}\geq 0.$$

(b) $0\leqslant F(x)\leqslant 1$, 且

$$F(-\infty)=\lim_{x\to\infty}F(x)=0,$$
$$F(\infty)=\lim_{x\to\infty}F(x)=1.$$

(c) $F(x+0)=F(x)$, 即 $F(x)$是右连续的.

(d) 一般，设离散型随机变量X的分布律为

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

由概率的可列可加性得X的分布函数为

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

即

$$F(x)=\sum_{x_k\leq x}p_*.$$

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

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

由概率的定义, p_k满足如下两个条件：

1° $p_k \geqslant 0, k = 1, 2, \cdots$;

2° $\sum_{k=1}^{\infty} p_k = 1$

重点：每点概率大于0，总概率为1（归一化）。

2. 离散型随机变量的分布律

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

一般考试答题，离散型随机变量都要写成如下表格形式

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（重要，期中、期末考试常考，牢记分布型）

(a) 0-1分布（较为简单，分清哪个取值为p）

设随机变量X只能取0与1两个值，它的分布律是:

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

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

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

(b) 伯努利试验、二项分布(重要，牢记)

设试验E只有两个可能结果：A及$\overline{A}$，则称E为伯努利试验。设$P(A) = p(0 < p < 1)$，此时设$P(A) = 1 - p$. 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

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

这种指定的方式共有$\begin{pmatrix} n \\ k \end{pmatrix}$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 $\begin{pmatrix} n \\ k \end{pmatrix} p^k(1-p)^{n-k}$, 记$q=1-p$, 即有

$$P\{X=k\}=\begin{pmatrix} n \\ k \end{pmatrix} p^k q^{n-k}, k=0,1,2,\cdots,n.$$

(c) 泊松分布(重要，牢记)

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

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

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

定理 2.4 (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

$$\lim_{n\to\infty}\begin{pmatrix} n \\ k \end{pmatrix} p_n^k\left(1-p_n\right)^{n-k}=\frac{\lambda^k \mathrm{e}^{-\lambda}}{k!}.$$

$$\begin{pmatrix} n \\ k \end{pmatrix} p^k(1-p)^{n-k}\approx\frac{\lambda^k \mathrm{e}^{-\lambda}}{k!} \quad (\text{其中 } \lambda=np\ ).$$

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.

(d) 其他重要分布（牢记分布律，考试可能会出现）

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

$$P\{\mathrm{Y}=\mathrm{n}\}=C_{n-1}^{r-1}p^r(1-p)^{n-r}, \mathrm{n}=\mathrm{r},\mathrm{r}+1\ldots\mathrm{N}$$

- Z服从参数为n和p的二项分布$Z\sim B(n,p)$,

$$P\{\mathrm{Z}=\mathrm{k}\}=C_n^k p^k(1-p)^{n-k}, \mathrm{k}=0,1,2\ldots\mathrm{n}$$

- 超几何分布$X \sim H(n,N,M)$,$P\{\mathrm{X=k}\}=\frac{c_M^k c_{N-M}^{n-k}}{c_N^n}$, $\mathrm{k}=0,1,2\ldots\min(\mathrm{n},\mathrm{M})$

2.1.3　连续型随机变量

1. 连续型随机变量及其概率密度

   (a) 定义（熟悉积分可积）

   一般, 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

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

   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. 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.

   (b) 性质

   i. $f(x)\geqslant 0$;

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

   iii. 对于任意实数$x_1,x_2\,(x_1\leqslant x_2)$,

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

   iv. 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$\Delta x\rightarrow 0$, 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

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

   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

(a) 均匀分布

分布律（虽然简单，但是在积分时要明确上下限）

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

称X在(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

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

(b) 指数分布

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

其中$\theta > 0$, 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

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

(c) 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（这一点要注意，可能应用条件会在应用题里面出现）。

i. 定义（虽然概率密度函数复杂，但是要记住，而且要熟练掌握相应的积分）

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

其中$\mu,\sigma(\sigma>0)$为常数, 记为$X\sim N\left(\mu,\sigma^2\right)$.

ii. 性质

- 曲线关$x=\mu$对称. 这表明对于任意$h>0$有

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

- 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, 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$\sigma$越小时图形变得越尖, 因而$X$落($\mu$附近的概率越大.

iii. 标准正态分布

- 定义

$$\mu=0,\sigma=1,X\sim N(0,1)$$

$$\Phi(-x)=1-\Phi(x)$$

- 转化（重要，牢记）

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

- 3σ法则

  设$X \sim N\left(\mu, \sigma^2\right)$, 由$\Phi(x)$的函数表还能得到

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

  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$3\sigma$法则.

- 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, 我们引人上α分位点的定义. $X \sim N(0,1)$, 若z_a满足条件

$$P\left\{X>z_a\right\}=\alpha, 0<\alpha<1,$$

  则称点z_{a}为标准正态分布的**上α分位点**. 下面列出了几个常用的z_a的值.

α	0.001	0.005	0.01	0.025	0.05	0.10
z_0	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 $z_{1-\alpha}=-z_\alpha$.

iv. 随机变量的函数的分布（熟记，可以简化计算）

定理 2.5. *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*

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

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

由已知随机变量X的概率分布, 可求得其函数的概率分布.

2.2　随机变量的函数及其概率分布

2.2.1　随机变量的函数的概念

定义 2.6. 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$Y = g\left(X_1, \cdots, X_n\right)$.

2.2.2　离散型随机变量的函数的概率分布

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.

2.2.3　连续型随机变量的函数的概率分布

对于这个问题,我们一般是先求出Y的分布函数,

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

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

定理 2.7. 设连续型随机变量X的概率密度为$f_X(x), y = g(x)$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

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

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

第三章　随机向量及其概率分布

本章在前两章的基础上, 结合向量, 引入了随机向量的概念以描述信息量更大的实验结果. 本章先给出随机向量的定义. 再从二维向量入手, 研究其概率分布、条件分布、相互独立性, 以及由随机向量构建的函数的概率密度计算.
[重点] 二维随机向量及其概率分布、条件分布、各维随机变量间的相互独立性、随机向量的函数的概率分布.

3.1　n维随机向量

3.1.1　随机向量的定义

定义 3.1 (随机向量). 设 Ω 是随机实验 E 的样本空间, ω 为样本点, 而 $X_1(\omega)$,$X_2(\omega)$, $\cdots$, $X_n(\omega)$ 是定义在 Ω 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维随机向量(或n维随机变量).

[注意]

1. 每一个分量都是一个随机变量.
2. 所有分量包含的ω 是同一个 ω .

因为二维的情况简单且便于在平面内想象其几何意义, 下面的讨论都将以二维变量为例. 更高维数的推广并不困难, 同学们可以自行探究.

3.1.2　分布函数与边缘分布函数

1. **分布函数**

   设(X, Y)是二维随机向量, 对任意实数x, y, 二元函数

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

称为二维随机向量(X,Y)的分布函数, 或X与Y的联合分布函数.

性质：

(a) $F(x,y)$对每个变量是单调增函数.

(b) $F(x,y)$对每个变量右连续.

(c) $F(-\infty,y)=F(x,-\infty)=F(-\infty,-\infty)=0$, $F(+\infty,+\infty)=1$.

(d) 对于任意两$(x_1,y_1),(x_2,y_2)$，若$x_1\le x_2$, $y_1\le y_2$, 则

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

$\boldsymbol{F(x,y)}$ 满足(b)(c)(d) $\Leftrightarrow$ $\boldsymbol{F(x,y)}$ 是某个二维随机向量的分布函数

2. **边缘分布函数**

$$F_X(x)=F(x,+\infty)$$

称$F_X(x)$为二维随机向量(X,Y)关于X的边缘分布函数.

同理,

$$F_Y(y)=F(+\infty,y)$$

称$F_Y(y)$为二维随机向量(X,Y)关于Y的边缘分布函数

3.1.3 二维随机向量的分类

1. **离散型随机变量** (X,Y)的所有可能取值是有限对(或可数对) .

   称$P\{X=x_i,Y=y_j\}=p_{ij}(i,j=1,2,\cdots)$为$(X,Y)$的分布律, 或$X$与$Y$的联合分布律.

2. **连续型随机变量** 对任意区域$A\subset \boldsymbol{R^2}$, 点(X,Y)落在该区域的概率可通过二重积分求得.

$$\boldsymbol{P\{(X,Y)\in A\}=\iint\limits_{A} f(x,y)dxdy.}$$

结合前面对分布函数、边缘分布函数的定义, 请同学们自行推导以上两种随机变量的分布函数与边缘分布函数. 推导完后, 可对照课本53~54页式(3.1.7) ~(3.1.9) (离散型随机变量) 与56页式(3.1.10) ~(3.1.12) 验证自己的结果.

连续型随机变量的两个实例

1. 二维正态分布

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

(其中, $-\infty<x<+\infty$,$-\infty<y<+\infty$)

(该公式并不要求记忆, 考试时会给出.)

推论 3.2. (X,Y)服从二维正态分布时, 其每一个变量均服从正态分布.

推论 3.3. *在仅改变参数ρ的情况下, 积分可发现X,Y服从的边缘概率密度不变. 由此可见,* ***各分量的边缘概率密度综合起来, 并不一定能唯一确定随机向量整体的概率密度****.*

2. 二维均匀分布

设G为平面上一个有界区域, 则服从二维均匀分布的二维随机向量概率密度为:

$$f(x,y)=\begin{cases}\dfrac{1}{A}, & (x,y)\in G\\ 0, & (x,y)\notin G\end{cases}\tag{3.1}$$

3.2　条件分布

多个随机变量之间存在相互影响与相互依赖关系. 为了研究这个问题, 我们引入条件分布这一概念. 其意在研究n维向量中, 当一部分分量值已知时, 其余分量的条件概率分布.
本节定义以1.3的条件概率公式为基础.
同3.1, 以下讨论都将以二维变量为例. 更高维数的推广仍请同学们自行探究.

3.2.1　离散型随机变量的条件分布律

定义 3.4. 设(X,Y)是二维离散型随机向量. 若对于固定的j, $P\{Y=y_j\}>0$, 则称

$$P\{X=x_i|Y=y_j\}=\frac{p_{ij}}{p_j}.$$

为在$Y=y_j$的条件下随机变量X的条件分布律.

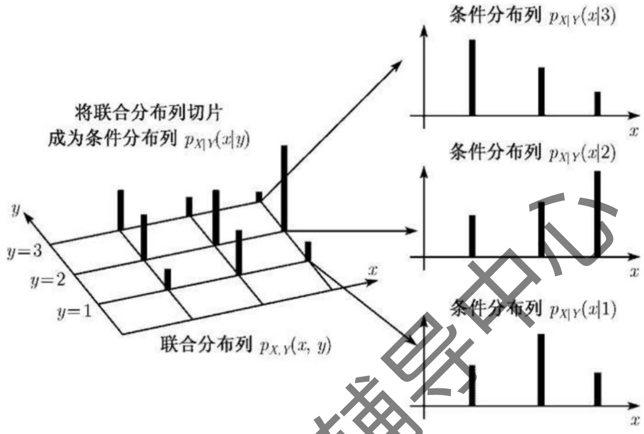


图 2.13　计算条件分布列 $p_{X|Y}(x|y)$ 的图示说明. 对每一个y, 可以将 $p_{X|Y}(x|y)$ 看成联合分布列 $p_{X,Y}(x,y)$ 沿 $Y=y$ 的一个切片，并且归一化后使得

$$\sum_x p_{X|Y}(x|y)=1$$

3.2.2　连续型随机变量的条件分布律

利用极限手段使条件概率计算公式中的分母非零, 可以得到如下定义:

定义 3.5. 设(X,Y)是二维离散型随机向量, 其概率密度为$f(x,y)$，(X,Y)关于Y的边缘概率密度为$f_y(y)$. 若对于固定的y，$f_y(y)>0$, 则称

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

为在$Y=y_j$的条件下, X的条件概率密度.

另, 称

$$F_{X|Y}(x|y)=\int_{-\infty}^{y}f_{X|Y}(u|y)du.$$

为在$Y=y_j$的条件下, X的条件分布函数.

[注意]区分条件概率密度与边缘概率密度概念的不同.

3.3　随机变量的相互独立性

本节内容由1.3节随机事件的相互独立性引申而来, 定义公式由$P(AB) = P(A)P(B)$引出.

3.3.1　二维随机变量相互独立的定义

定义 3.6. 对二维随机向量(X, Y), 若对任意实数x，y恒有

$$P\{X \le x, Y \le y\} = P\{X \le x\}P\{Y \le y\}.$$

则称随机变量X与Y是相互独立的.

该定义同时等价于

$$F(x, y) = F_x(x)F_y(y) \tag{3.2}$$

对二维离散型随机变量,(1.2)等价于

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

对所有(X, Y)可能取值成立.
对二维连续型随机变量, (1.2) 等价于

$$\boldsymbol{f(x, y) = f_x(x)f_y(y)}.$$

几乎处处成立.

3.3.2　相互独立在高维随机变量中的推广

上述相互独立性可以推广到多个随机变量之间, 单维随机变量也可被推广至多维随机变量.

1. **$\boldsymbol{n(n > 2)}$个随机变量之间**

   若满足

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

   则称$X_1, X_2, \cdots, X_n$相互独立.

2. 多维随机向量之间

若满足

$$F(x_1,x_2,\cdots,x_m,y_1,y_2,...,y_n)=F_1(x_1,x_2,\cdots,x_m)F_2(y_1,y_2,\cdots,y_n).$$

则称$(X_1,X_2,\cdots,X_m)$与$(Y_1,Y_2,\cdots,Y_n)$相互独立.

定理 3.7. 设$(X_1,X_2,\cdots,X_m)$与$(Y_1,Y_2,\cdots,Y_n)$相互独立. 若$h(x_1,x_2,\cdots,x_m),g(y_1,y_2,\cdots,y_n)$是**连续函数**, 则$\boldsymbol{h(X_1,X_2,\cdots,X_m)}$**与**$\boldsymbol{g(Y_1,Y_2,\cdots,Y_n)}$**也相互独立**.

定理 3.8. 若随机变量$X_1,X_2,\cdots,X_m$相互独立, 将其分为**不相交的**k组,每组中所有变量由一个**连续函数**复合生成一个新的随机变量, 则这k个随机变量相互独立.

3.4 随机向量的分布函数及其概率分布

本节讨论高维随机数据的**压缩**（或**降维**）.

3.4.1 随机向量的函数的定义

1. 单个随机变量

定义 3.9. 若$g(x_1,x_2,\cdots,x_n)$为定义在D上的n元函数, Y为随机变量. 当$(X_1,X_2,\cdots,X_n)$取可能值$(x_1,x_2,\cdots,x_n)\in D$时, Y取对应值$g(x_1,x_2,\cdots,x_n)$. 称$\boldsymbol{Y}$**为随机向量**$\boldsymbol{(X_1,X_2,\cdots,X_n)}$**的函数**, 记为$\boldsymbol{Y=g(X_1,X_2,\cdots,X_n)}$

2. 推广至多维随机变量

定义 3.10. 若$g_j(x_1,x_2,...,x_n)(j=1,2,...,k)$为定义在$D$上的$n$元函数, $Y_j(j=1,2,...,k)$为随机变量, 其中**正整数**$\boldsymbol{k\le n}$. 当$(X_1,X_2,...,X_n)$取可能值$(x_1,x_2,...,x_n)\in D$时, Y_j取对应值$g_j(x_1,x_2,...,x_n)$. 称$\boldsymbol{(Y_1,Y_2,...,Y_k)}$**为随机向量**$\boldsymbol{(X_1,X_2,...,X_n)}$**的函数**, 记为$\boldsymbol{Y=g(X_1,X_2,...,X_n)}$

3.4.2 二维随机向量的函数的概率分布

1. 离散型随机向量

确定离散型随机向量的函数的概率较简单, 将新随机变量表示的事件转化为老随机变量表示的等价时间即可.

2. **连续型随机变量**[**本章难点**]

设二维连续型随机向量(X,Y)的概率密度为$f(x,y)$, $(U,V)=(g_1(X,Y),g_2(X,Y))$为随机向量$(X,Y)$的函数, $X=h_1(U,V)$, $Y=h_2(U,V)$为其逆变换, 雅可比行列

式为$J=\dfrac{\partial(x,y)}{\partial(u,v)}=\begin{vmatrix}\dfrac{\partial h_1}{u} & \dfrac{\partial h_1}{v}\\ \dfrac{\partial h_2}{u} & \dfrac{\partial h_2}{v}\end{vmatrix}$. 则概率密度之间有变换关系:

$$f_{U,V}(u,v)=\begin{cases}\boldsymbol{f(h_1(u,v),h_2(u,v))|J|}\text{fi} & (u,v)\in D\\ 0\text{fi} & \text{其他}\end{cases}\tag{3.3}$$

（1）$\boldsymbol{X+Y}$**的概率密度**

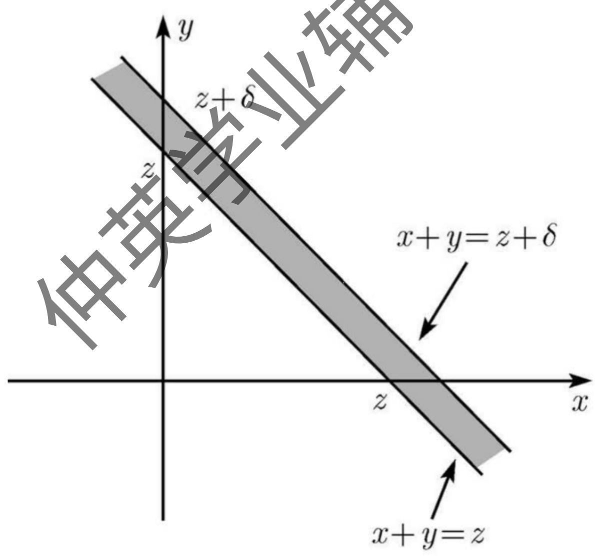


图 4.8　连续随机变量情形下卷积公式的说明(对比图4.7). 对非常小的 $\delta>0$, 图中带形区域所代表的事件发生的概率就是 $\mathrm{P}(z\le X+Y\le z+\delta)\approx f_Z(z)\delta$. 于是,

$$\begin{aligned}f_Z(z)\delta &= \mathrm{P}(z\le X+Y\le z+\delta)\\ &=\int_{-\infty}^{\infty}\int_{z-x}^{z-x+\delta}f_X(x)f_Y(y)\mathrm{d}y\mathrm{d}x\\ &\approx\int_{-\infty}^{\infty}f_X(x)f_Y(z-x)\delta\mathrm{d}x.\end{aligned}$$

去掉上式左右两边的 δ 即得所求公式

计算公式:

$$\boldsymbol{f_{X+Y}(t)=\int_{-\infty}^{+\infty}f(x,t-x)dx=\int_{-\infty}^{+\infty}f(t-y,y)dy}.$$

特别的, 当X与Y相互独立时, $f(x,y)=f_X(x)f_Y(y)$, 公式可进一步简化为:

$$\boldsymbol{f_{X+Y}(t)=\int_{-\infty}^{+\infty}f_X(x)f_Y(t-x)dx=\int_{-\infty}^{+\infty}f_X(t-y)f_Y(y)dy}.$$

上式也称独立变量值和的**密度卷积公式**.

(该知识点请给予足够重视! 建议反复推导课本P69的例3.12两种解法至熟练以掌握该公式的应用方法.)

推论 3.11. ***利用上述公式进行推导可以发现, n个相互独立的正态随机变量的线性组合仍为正态随机变量.***

（2）$\boldsymbol{\frac{X}{Y}}$**的概率密度**

计算公式:

$$\boldsymbol{f_{X/Y}(t)=\int_{-\infty}^{+\infty}|y|f(yt,y)dy}.$$

当X与Y相互独立时, 同（1）理, 有

$$\boldsymbol{f_{X/Y}(t)=\int_{-\infty}^{+\infty}|y|f_X(yt)f_Y(y)dy}.$$

（3）$\mathbf{max}\boldsymbol{\{X,Y\}}$, $\mathbf{min}\boldsymbol{\{X,Y\}}$**的分布**

计算公式（设$M=max\{X_1,X_2,\cdots,X_n\}$，$N=min\{X_1,X_2,\cdots,X_n\}$:

$$\boldsymbol{F_M(z)=F_{X_1}(z)F_{X_2}(z)...F_{X_n}(z)} \tag{3.4}$$

$$\boldsymbol{F_M(z)=1-[1-F_{X_1}(z)][1-F_{X_2}(z)]...[1-F_{X_n}(z)]} \tag{3.5}$$

【注意】（1.4）（1.5）两式同样适用于离散型随机变量.

（熟记以上三个函数的相关公式为基本要求. 此外, 编者还强烈建议熟悉推导过程, 有助于加深对本章内容的理解. ）

3.5　经典例题

不论从理解知识点的角度还是从应试的角度, **课本例题有很高的参考价值**. 下面的例题源自其他教材, 仅作为补充资料, 请同学们仍以课本例题与作业为重. 往年考试题也是很好的备考资料.

例 3.1. *(布丰的抛针实验)这是一个著名的例子, 几何概率由此发源, 所讨论的问题是对随机放置的对象的几何性质的分析. 在平面上画了若干条平行线, 相互之间的距离为d. 现在往平面上随机地抛掷一根针, 针的长度为l. 问: 针与直线相交的概率有多大? (我们假定$l < d$, 这样针不会同时与两条直线同时相交.)*

解: 设X为针的中点李德最近的那一条直线的垂直距离, θ表示针与平行直线之间的夹角. 理想情况下, (X,θ)的联合概率密度为矩形集合$\{(x,\theta)|0 \le x \le d/2, 0 \le \theta \le \pi/2\}$上的联合均匀概率密度函数. 因此:

$$f(x,y) = \begin{cases} 4/(\pi d), & \text{若} x \in [0,d/2] \text{且} \theta \in \pi/2 \\ 0, & \text{其他}, \notin G \end{cases} \tag{3.6}$$

可以分析得到, 针与平行直线相交的充要条件为:

$$X \le \frac{1}{2} sin\theta$$

其相应的概率为

$$\begin{aligned} P(X \le (l/2)sin\theta) &= \int_{x \le (l/2)sin\theta} \int f_{X,\theta}(x,\theta) dx d\theta \\ &= \frac{4}{\pi d} \int_0^{\pi/2} \int_0^{(l/2)sin\theta} dx d\theta \\ &= \frac{4}{\pi d} \int_0^{\pi/2} \frac{l}{2} sin\theta d\theta \\ &= \frac{2l}{\pi d} (-cos\theta)\Big|_0^{\pi/2} \\ &= \frac{2l}{\pi d} \end{aligned}$$

可以看到, 该概率值与π有关. 大量重复抛针实验, 我们可以对π进行经验值估计.

3.6　结语

受制于篇幅与本章编者能力, 本章内容到此结束. 对概率论有兴趣的同学, 推荐进一步阅

读Dimitri P. Bertsekas和John N. Tsitsiklis编写的《概率导论》一书, 该教材讲解清晰生动, 且相较于我们使用的教材涵盖了更多内容. 学校使用的教材第3章内容在该书2-4章中均有分布, 小助手的编写也参考了这本书中的插图.

第四章　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

4.1　数学期望

4.1.1　数学期望的定义

定义 4.1 (离散型随机变量的数学期望)**.** 设离散型随机变量X的分布律为 $P\{X=x_i\}=p_i, i=1,2,\cdots$. 若级数$\sum\limits_{i=1}^{\infty}x_ip_i$绝对收敛, 则称$\sum\limits_{i=1}^{\infty}x_ip_i$为随机变量$X$的**数学期望**, 或称为**理论均值**, 记作$E(X)$, 即

$$E(X)=\sum_{i=1}^{\infty}x_ip_i.$$

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

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

例 4.1. 设随机变量X的分布律为$P\left\{X=\dfrac{(-1)^k2^k}{k}\right\}=\dfrac{1}{2^k},k=1,2\ldots$,证明随机变量$X$的数学期望不存在.

证明: $\displaystyle\sum_{i=1}^{\infty}|x_i|\,p_i=\sum_{i=1}^{\infty}\left|\frac{(-1)^k2^k}{k}\right|\frac{1}{2^k}=\sum_{i=1}^{\infty}\frac{1}{k}$发散,故数学期望不存在.

例 4.2. *Cauchy* 分布的密度函数为$f(\mathrm{x})=\frac{1}{\pi(1+x^2)},-\infty<\mathrm{x}<+\infty$, 证明其数学期望不存在.

证明:

$$\int_{-\infty}^{+\infty}|x|f(x)dx=\int_{-\infty}^{0}\frac{-x}{\pi\left(1+x^2\right)}dx+\int_{0}^{+\infty}\frac{x}{\pi\left(1+x^2\right)}dx=2\int_{0}^{+\infty}\frac{x}{\pi\left(1+x^2\right)}dx=\frac{1}{\pi}\ln\left(1+x^2\right)\Big|_{0}^{+\infty}$$

发散, 故数学期望不存在.

[注 1] 定义数学期望是为了表示随机变量的平均值, 该平均值在离散型随机变量的数学期望中是无穷级数的和, 它不应该因级数的并项或重排而改变, 即级数$\sum\limits_{i=1}^{\infty}x_ip_i$无论以何种方式并项或重排, 都能收敛到同一数值, 而绝对收敛的级数可以满足此要求, 这就是要求级数绝对收敛的理由. 同理, 当X为连续型随机变量时, 这就是定义要求反常积分绝对收敛的根据.

[注 2] 随机变量的数学期望是一个常数, 不再是一个随机变量, 因为随机因素已经被加权平均了.

[注 3] 连续型随机变量的数学期望定义式中的积分, 在力学上可解释为单位质量的棒形刚体的重心, 这里$f(x)$为在截面x处有单位质量的刚体的密度, 数学期望$E(X)$是由X的概率分布所确定的一个常数. 当随机变量X服从某一分布时, 也把$E(X)$称为该分布的数学期望.

4.1.2 随机变量的函数的数学期望

定理 4.3. 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 $P\left\{X=x_i\right\}=p_i, i=1,2,\cdots$时, 若$\sum\limits_{i=1}^{\infty}\left|g\left(x_i\right)\right|p_i$绝对收敛, 则$E(Y)$存在, 且

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

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

对于二维随机变量函数, 类似地, 有如下定理.

定理 4.4. 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$\displaystyle\sum_{i=1}^{\infty}|g(x_i,y_j)|p_{ij}$绝对收敛, 则$E(Z)$存在, 且

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

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

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

4.1.3 数学期望的性质

性质 4.5. 数学期望具有如下性质

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

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

把性质 (2) 与性质 (3)结合起来, 则有

$$E(C_1X+C_2Y)=C_1E(X)+C_2E(Y).$$

一般地, 数学期望具有如下的线性性质

$$E\left[\sum_{k=1}^{n} C_k X_k\right]=\sum_{k=1}^{n} C_k E\left(X_k\right).$$

性质 (4) 还可以推广到$n(n>2)$个相互独立的随机变量的情形, 即若$X_1, X_2, \cdots, X_n$是n个相互独立的随机变量, 则

$$E\left(X_1 X_2 \cdots X_n\right)=E\left(X_1\right) E\left(X_2\right) \cdots E\left(X_n\right).$$

4.2 方差

4.2.1 方差与标准差

定义 4.6 (方差与标准差)**.** *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*

$$D(X)=E(X-E(X))^2.$$

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

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

所以

$$D(X)=E\left(X^2\right)-(E(X))^2.$$

4.2.2 方差的性质

性质 4.7. *1. $D(C)=0$;*

2. $D(CX)=C^2 D(X)$;

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

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

4.3　协方差与相关系数、矩

4.3.1　协方差与相关系数

定义 4.8 (协方差)**.** 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

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

由上述定义可知, $\mathrm{Cov}(X,X)=E\left[(X-EX)^2\right]=D(X)$. 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”. 把上式右端的各项展开, 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

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

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

定理 4.9. 设(X,Y)为二维随机变量,

1. 若X与Y相互独立, 则

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

2. 若$E\left(X^2\right)$、$E\left(Y^2\right)$存在, 则有

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

特别地, 有

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

定义 4.10. 称$\frac{\mathrm{Cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$为$X$与$Y$的相关系数, 记为$\rho_{XY}$, 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$P\{Y=a+bX\}=1$.

定义 4.11. 若$\rho_{XY}=0$, 则称X与Y不相关.

因为$\rho_{XY}=0$等价于$\mathrm{Cov}(X,Y)=0$, 所以, 当协方差$\mathrm{Cov}(X,Y)=0$时, 同样称X与Y不相关.

不难看出, 当X与Y不相关时, 就有$D(X\pm Y)=D(X)+D(Y)$. 因此,方差的性质 (3) 成立的条件可由“若X与Y相互独立” 改为“若X与Y不相关”. 事实上, 这是用较弱的条件取代较强的条件. 因而, 其结论更具一般性.

例 4.3. $X\sim N(1,3^2)$, $Y\sim N(0,4^2)$, X和Y的相关系数$\rho_{XY}=-\frac{1}{2}$, 设$Z=\frac{X}{3}+\frac{Y}{2}$, 求$X$和$Z$的相关系数.

解: $\mathrm{Cov}(X,Z)=\mathrm{Cov}\left(X,\frac{X}{3}+\frac{Y}{2}\right)=\mathrm{Cov}\left(X,\frac{X}{3}\right)+\mathrm{Cov}\left(X,\frac{Y}{2}\right)=\frac{1}{3}D(X)+\frac{1}{2}\mathrm{Cov}(X,Y)=\frac{1}{3}D(X)+\frac{1}{2}\rho_{XY}\sigma_X\sigma_Y=\frac{1}{3}\times 9+\frac{1}{2}\times\left(-\frac{1}{2}\right)\times 3\times 4=0.$

4.3.2 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;
- $E\left\{[X-E(X)]^k\right\}$ ——X的k阶中心矩;
- $E\left\{[X-E(X)]^2\right\}=D(X)$——$X$ 的方差（二阶中心矩）
- $E(X^kY^l)$——X,Y的$k+l$阶混合原点矩;

- $E\left\{[X-E(X)]^k[Y-E(Y)]^l\right\}$——$X,Y$的$k+l$阶混合中心矩;
- $E((X-E(X))(Y-E(Y)))$——X,Y的协方差（二阶混合中心矩）.

第五章　大数定律与中心极限定理

5.1　依概率收敛与切比雪夫不等式

命题 5.1. 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$P(X \geq \varepsilon) \leq \dfrac{E(X)}{\varepsilon}$.

定理 5.2 (马尔可夫不等式)**.** 设随机变量X的k阶绝对原点矩$E\left(|X|^k\right)$存在, 则对于任意实数$\varepsilon > 0$, 有$P(|X| \geq \varepsilon) \leq \dfrac{E\left(|X|^k\right)}{\varepsilon^k}$.

定理 5.3 (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$P(|X - E(X)| \geq \varepsilon) \leq \dfrac{D(X)}{\varepsilon^2}$ (当$\varepsilon^2 \leq D(X)$时, 该不等式无实际意义).

定义 5.4 (依概率收敛)**.** 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:

$$\lim_{n\to\infty} P\left\{|X_n - a| < \varepsilon\right\} = 1 \quad \text{或} \lim_{n\to\infty} P\left\{|X_n - a| \geq \varepsilon\right\} = 0,$$

则称$\{X_n\}$依概率收敛于a。

5.2　大数定律

定理 5.5 (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:

$$\lim_{n\to\infty} P\left\{\left|\frac{\eta_n}{n} - p\right| < \varepsilon\right\} = 1 \quad \text{或} \lim_{n\to\infty} P\left\{\left|\frac{\eta_n}{n} - p\right| \geq \varepsilon\right\} = 0.$$

定理 5.6 (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$E\left(X_i\right)$和方差$D\left(X_i\right)$, 且方差一致有界, 即存在某一常数$C$, 使$D\left(X_i\right) < C$,

对任意$\varepsilon>0$, 有:

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

推论 5.7. 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$E\left(X_i\right)=\mu$, 方差$D\left(X_i\right)=\sigma^2$, 则对任意$\varepsilon>0$, 有:

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

定理 5.8 (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$E\left(X_i\right)=\mu$, 则对任意$\varepsilon>0$, 有:

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

5.3　中心极限定理

定理 5.9 (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 $E\left(X_k\right)=\mu,D\left(X_k\right)=\sigma^2$, 则有

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

推论 5.10. 标准化变量$Y_n=\dfrac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{n}\sigma}$ 总以标准正态分布$N(0,1)$为其极限分布.

推论 5.11. n个独立随机变量的和$\displaystyle\sum_{i=1}^{n}X_i=\sqrt{n}\sigma Y_n+n\mu$ 的极限分布是正态分布$N\left(n\mu,n\sigma^2\right)$.

定理 5.12 (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:

$$\lim_{n\to\infty}P\left\{\frac{\eta_n-np}{\sqrt{np(1-p)}}\leq x\right\}=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt=\Phi(x).$$

第六章　数理统计学的基本概念

6.1　总体与样本

6.1.1　总体

定义 6.1 (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.

定义 6.2 (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.

6.1.2　样本

定义 6.3 (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.
样本空间——样本所有可能取值的集合

定义 6.4 (简单随机样本). 若$(X_1, X_2, \cdots, X_n)$是从总体X抽取的样本, 而且满足:

- $X_1, X_2, \cdots, X_n$与X有相同的分布.
- $X_1, X_2, \cdots, X_n$相互独立.

则称$(X_1, X_2, \cdots, X_n)$为简单随机样本.

命题 6.5 (样本的联合分布函数).

- 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:

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

- 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:

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

- 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:

$$f_{总}(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f(x_i).$$

6.1.3 样本分布

命题 6.6 (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)*

定义 6.7 (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$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$, 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.*

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

定理 6.8 (格里绞科定理). *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.*

6.2 统计量及其数字特征

6.2.1 统计量的基本概念

- 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$(x_1, x_2, \cdots, x_n)$是样本 $(X_1, X_2, \cdots, X_n)$的一个观测值, 则称$g(x_1, x_2, \cdots, x_n)$为统计量的一个观测值.

- 统计量的分布称为抽样分布.

6.2.2 常用统计量

- 样本均值: $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$.

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

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

- 样本的k阶原点矩: $A_k = \dfrac{1}{n}\sum_{i=1}^{n} X_i^k$.

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

  特别地, 当$k=2$时, $\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \sum_{i=1}^{n} X_i^2 - n\overline{X}^2 = n\left(A_2 - \overline{X}^2\right)$.

- 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: $X_{(n)} = \max_{1\le k\le n}\{X_k\}, X_{(1)} = \min_{1\le k\le n}\{X_k\}$.
  - 样本极差: $D_n = X_{(n)} - X_{(1)}$.

定理 6.9. *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. $E\left(S^2\right)=\sigma^2$.

3. 若X的四阶矩存在, S^2依概率收敛于σ^2.

定理 6.10. 设总体X的k阶原点矩$\alpha=E\left(X^k\right)(k\geq 1)$存在, 则有:

$E\left(A_k\right)=\alpha_k$.

当X的$2k$阶原点矩存在, A_k依概率收敛于α_k。.

6.3 抽样分布

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.

6.3.1 三个重要分布

定义 6.11 (χ^2分布)**.** 若随机变量Z具有概率密度:

$$\chi^2(x;n)=\begin{cases}\dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}, & x>0\\ 0, & x\leqslant 0\end{cases}$$

则称Z服从自由度为n的χ^2分布, 记作$Z\sim\chi^2(n)$. $\chi^2(n)$分布的概率密度曲线(需大致了解曲线随自由度n的变化):

命题 6.12 (χ^2分布的性质)**.**

- $\chi^2(2)$为参数为$\frac{1}{2}$的指数分布.
- $\chi^2(n)$分布的数学期望和方差: $E\left(\chi^2\right)=n, D\left(\chi^2\right)=2n$.
- $\chi^2(n)$分布的可加性: $Z_1\sim\chi^2\left(n_1\right)$, $Z_2\sim\chi^2\left(n_2\right)$, 且$Z_1,Z_2$相互独立, 则$Z_1+Z_2\sim\chi^2\left(n_1+n_2\right)$, 可推广到多个随机变量相加的情形.
- 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$N(0,1)$, 则随机变量$\chi^2=X_1^2+X_2^2+\cdots+X_n^2$服从自由度为$n$的$\chi^2$分布, 记作$\chi^2\sim\chi^2(n)$. 特别地: $X_i\sim N(0,1)$, $X_i^2\sim\chi^2(1)$.

定义 6.13 (t分布)**.** 若随机变量T具有概率密度:

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

则称T服从自由度为n的t分布, 记为$T\sim t(n)$.

命题 6.14 (t分布的性质)**.**

- t分布密度函数关于$x=0$对称, 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
- t分布的数学期望和方差: $E(T)=0, D(T)=\frac{n}{n-2}\quad(n>2)$.

定义 6.15 (F分布)**.** 若随机变量F具有概率密度

$$f\left(x;n_1,n_2\right)=\begin{cases}\dfrac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}\left(\frac{n_1}{n_2}\right)\left(\frac{n_1}{n_2}x\right)^{\frac{n_1}{2}-1}\left(1+\frac{n_1}{n_2}x\right)^{-\frac{n_1+n_2}{2}}, & x>0,\\ 0, & x\leqslant 0\end{cases}$$

则称F服从自由度为(n_1,n_2)的F分布, 记作$F\sim F\left(n_1,n_2\right)$.

命题 6.16 (F分布的性质)**.**

1. 设随机变量$X_1\sim\chi^2\left(n_1\right)$, $X_2\sim\chi^2\left(n_2\right)$, 且$X$与$Y$相互独立, 则$F=\frac{X_1/n_1}{X_2/n_2}\sim F\left(n_1,n_2\right)$
2. 若$X\sim F\left(n_1,n_2\right)$, 则$\frac{1}{X}\sim F\left(n_2,n_1\right)$.

6.3.2 分位数

定义 6.17 (分位数)**.** 设随机变量X的分布函数为$F(x)=P\{X\leq x\}$, 对于给定的$\alpha(0<\alpha<1)$, 若存在x_α 使$P\left\{X>x_\alpha\right\}=1-F\left(x_\alpha\right)=\alpha$, 则称$x_\alpha$为$X$分布的上侧$\alpha$分位数. 若存在$x_p$, 使$P\left\{X\leqslant x_p\right\}=1-F\left(x_p\right)=p$, 则称$x_p$为$X$分布的下侧侧$p$分位数.

由定义可见, 上侧α分位数就是下侧$1-\alpha$分位数.

6.3.3 正态总体的抽样分布

定理 6.18. 设$X_1, X_2, \cdots, X_n$是取自$N(\mu, \sigma^2)$的样本, $\overline{X}$为样本均值, S^2为样本的方差, 则

1. $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$;
2. $\dfrac{(n-1)S^2}{\sigma^2} = \dfrac{1}{\sigma^2}\sum\limits_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \sim \chi^2(n-1)$;
3. $\overline{X}$与S^2相互独立.

定理 6.19. 设$X_1, X_2, \cdots, X_n$是取自$N(\mu, \sigma^2)$的样本, $\overline{X}$为样本均值, S^2为样本的方差, 则$T = \dfrac{\sqrt{n}(\overline{X} - \mu)}{S} \sim t(n-1)$.

定理 6.20. 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$N(\mu_1, \sigma_1^2)$和$N(\mu_2, \sigma_2^2)$的样本且相互独立, 则有

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{S_n\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2), \quad 其中 S_n = \sqrt{\frac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 1}}.$$

定理 6.21. 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

$$\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1).$$

6.4 经典例题

例 6.1. 设总体X的概率密度为$f(x) = \begin{cases} c|x|, & |x| < 1 \\ 0, & |x| \geq 1 \end{cases}$, $(X_1, X_2, \cdots, X_{50})$为总体的样本, 求:

(1) $\overline{X}$的数学期望与方差

(2) $E(S^2)$

(3) $P(|\overline{X}| > 0.02)$

解:

(1) 由定理可知$E(\overline{X}) = E(X) = \displaystyle\int_{-1}^{1} x|x|dx = 0$, $D(\overline{X}) = \frac{1}{50}D(X) = \frac{1}{50}E\left(X^2\right) = \frac{1}{50} \times 2\int_0^1 x^2|x|dx = \frac{1}{100}$;

(2) 由定理可知$E\left(S^2\right) = D(X) = E\left(X^2\right) = \frac{1}{2}$;

(3) 由中心极限定理可知$\overline{X} \sim N\left(0, \frac{1}{100}\right)$, 则有

$$\begin{aligned} P(|\overline{X}| > 0.02) =& 1 - P(|\overline{X}| \leq 0.02) \\ =& 2\left(1 - \Phi\left(\frac{0.02 - 0}{0.1}\right)\right). \\ =& 2(1 - \Phi(0.2)) \\ =& 0.8414 \end{aligned}$$

例 6.2. 设总体$X \sim N\left(0, \sigma^2\right)$, 样本方差的均值$E\left(S^2\right) = 4$, 求$D\left(\frac{X^2}{\sigma^3}\right)$.

解:

$\because E\left(S^2\right) = 1, \quad \therefore \sigma^2 = E\left(S^2\right) = 4$

$\because X \sim N\left(0, \sigma^2\right), \quad \therefore \dfrac{X}{\sigma} \sim N(0,1), \dfrac{X^2}{\sigma^2} \sim \chi^2(1)$

则$D\left(\dfrac{X^2}{\sigma^2}\right) = 2$, $D\left(\dfrac{X^2}{\sigma^3}\right) = \dfrac{1}{\sigma^2}D\left(\dfrac{X^2}{\sigma^2}\right) = \dfrac{1}{2}$.

例 6.3. 设$X_1, X_2, \cdots, X_n$是取自$N\left(\mu, \sigma^2\right)$的样本, S^2为样本的方差, 求$2S^2$的期望和方差.

解:

由定理知:$\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1), E\left(S^2\right) = \sigma^2 \therefore D\left[\dfrac{(n-1)S^2}{\sigma^2}\right] = 2(n-1)$, $D\left(S^2\right) = \dfrac{2\sigma^4}{n-1}$

故$E\left(2S^2\right) = 2\sigma^2, D\left(2S^2\right) = 4 \times \dfrac{2\sigma^4}{n-1} = \dfrac{8\sigma^4}{n-1}$.

例 6.4. 设$(X_1, X_2, \cdots, X_n)$是来自总体$N\left(\mu, \sigma^2\right)$的样本$(n > 1)$, 下述的两个结论是否正确? 说明理由.

$$(1) \quad \frac{(\overline{X} - \mu)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}}\sqrt{n(n-1)} \sim t(n-1);$$

$$(2) \quad \frac{n(\overline{X} - \mu)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \mu\right)^2}} \sim t(n).$$

解:

(1)正确.

$\because X \sim N\left(\mu, \sigma^2\right)$, $\quad \therefore U = \dfrac{\sqrt{n}(\overline{X}-\mu)}{\sigma} \sim N(0,1)$, $\quad W = \dfrac{(n-1)S^2}{\sigma^2} = \dfrac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{\sigma^2} \sim \chi^2(n-1)$

$\therefore \overline{X}$与S^2独立(定理),从而U与W独立.

则 $\dfrac{U}{\sqrt{W/(n-1)}} = \dfrac{(\overline{X}-\mu)}{\sqrt{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}}\sqrt{n(n-1)} \sim t(n-1)$.

(2)不正确.

同理: $U = \dfrac{\sqrt{n}(\overline{X}-\mu)}{\sigma} \sim N(0,1)$, 又$X_i \sim N\left(\mu, \sigma^2\right)$, $\therefore Y_i = \frac{X_i-\mu}{\sigma} \sim N(0,1)$, 且$Y_1, Y_2, \cdots, Y_n$独立.

故$W = \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} = \sum_{i=1}^{n} Y_i^2 \sim \chi^2(n)$,

但是U与W不独立, 所以不能就此断言$\frac{U}{\sqrt{W/n}} \sim t(n)$.

例 6.5. 设X服从参数为λ的指数分布, $(X_1, X_2, \cdots, X_n)$是来自总体X的样本, 则$2\lambda\sum_{i=1}^{n} X_i$服从什么分布?

解: 设$X \sim \exp(\lambda)$, 则$Y = 2\lambda X$的概率密度函数为

$$f_Y(y) = f_X\left(\frac{y}{2\lambda}\right)\frac{1}{2\lambda} = \begin{cases} \frac{1}{2}e^{-\frac{y}{2}} & y > 0 \\ 0 & y \le 0 \end{cases}$$

即$Y = 2\lambda X \sim \exp\left(\frac{1}{2}\right) = \chi^2(2)$(重要结论可得)

则由$(X_1, X_2, \cdots, X_n)$是来自总体X的样本可知 $2\lambda X_i \sim \chi^2(2)$, $\quad i = 1, 2, \cdots, n$ 　且相互独立,

因此由χ^2分布的可加性可知$2\lambda\sum_{i=1}^{n} X_i \sim \chi^2(2n)$.

例 6.6. 设$X \sim t(n)$,$Y \sim F(1,n)$, 给定$\alpha(0 < \alpha < 0.5)$, 常数c满足$P(X > c) = \alpha$, 则$P\left(Y > c^2\right) =$ ______.

解:

此题需运用结论: 若$X \sim t(n)$, 则$X^2 \sim F(1,n)$.

证明过程如下: 设$U \sim N(0,1), W \sim \chi^2(n)$且$U$与$W$相互独立, 记$X = \frac{U}{\sqrt{W/n}}$ 则$X \sim t(n)$且$X^2 = \frac{U^2/1}{W/n}$.

又$U^2 \sim \chi^2(1), W \sim \chi^2(n)$, 且$U$与$W$相互独立, $\quad \therefore X^2 \sim F(1,n)$.

得证.

$$\begin{aligned}P\left\{Y>c^2\right\}&=P\left\{X^2>c^2\right\}\\&=P\{(X>c)\cup(X<-c)\}\\&=P\{X>c\}+P\{X<-c\}=2\alpha\end{aligned}$$

例 6.7. 设$X_1,X_2,\cdots,X_9$和$Y_1,Y_2,\cdots,Y_5$分别是从正态总体 $N(0,4)$和$N(8,9)$取出的相互独立的两组简单随机样本, 记$\overline{Y}=\frac{1}{5}\sum_{j=1}^5 Y_j$, 则$\dfrac{\sum_{j=1}^9 X_j}{\sqrt{\sum_{j=1}^5\left(Y_j-\overline{Y}\right)^2}}$服从什么分布?

解:

易知$\overline{X}=\dfrac{1}{9}\displaystyle\sum_{j=1}^9 X_j\sim N\left(0,\dfrac{4}{9}\right)$, 则$U=\dfrac{\overline{X}}{2/3}=\dfrac{1}{6}\displaystyle\sum_{j=1}^9 X_j\sim N(0,1)$.

又$V=\dfrac{1}{9}\displaystyle\sum_{j=1}^5\left(Y_j-\overline{Y}\right)^2\sim\chi^2(4)$, $\quad\therefore\dfrac{U}{\sqrt{V/4}}=\dfrac{\sum_{j=1}^9 X_j}{\sqrt{\sum_{j=1}^5\left(Y_j-\overline{Y}\right)^2}}\sim t(4)$

例 6.8. 设X_1,X_2,X_3,X_4是来自$N\left(0,2^2\right)$的样本, 则当a,b满足什么条件时, 统计量 $a\left(X_1-2X_2\right)^2+b\left(3X_3-4X_4\right)^2$服从$\chi^2$分布?

解:

显然$X_1-2X_2\sim N(0,20), 3X_3-4X_4\sim N(0,100)$, 且他们之间相互独立,

则$\frac{1}{\sqrt{20}}\left(X_1-2X_2\right)\sim N(0,1)$, $\frac{1}{10}\left(3X_3-4X_4\right)\sim N(0,1)$.

$\therefore\frac{1}{20}\left(X_1-2X_2\right)^2\sim\chi^2(1), \frac{1}{100}\left(3X_3-4X_4\right)^2\sim\chi^2(1)$

$\therefore\frac{1}{20}\left(X_1-2X_2\right)^2+\frac{1}{100}\left(3X_3-4X_4\right)^2\sim\chi^2(2)\ \therefore a=0.05, b=0.01.$

例 6.9. 设总体$X\sim N\left(0,\sigma^2\right)$, $X_1,X_2,\cdots,X_5$ 为来自该总体的一个容量为5的简单随机样本, 则当$c=$ 　　时, $Z=\dfrac{C\left(X_1^2+X_2^2\right)}{X_3^2+X_4^2+X_5^2}$服从$F$分布.

解:

易知$\dfrac{X_i}{\sigma}\sim N(0,1), \dfrac{X_i^2}{\sigma^2}\sim\chi^2(1), i=1,2,\cdots,5.\quad\therefore\dfrac{X_1^2+X_2^2}{\sigma^2}\sim\chi^2(2), \dfrac{X_3^2+X_4^2+X_5^2}{\sigma^2}\sim\chi^2(3)$, 且他们之间独立.

$$\therefore\frac{\dfrac{X_1^2+X_2^2}{\sigma^2}/2}{\dfrac{X_3^2+X_4^2+X_5^2}{\sigma^2}/3}=\frac{3}{2}\frac{X_1^2+X_2^2}{X_3^2+X_4^2+X_5^2}\sim F(2,3)\quad\therefore C=\tfrac{3}{2}.$$

例 6.10. 设随机变量$X\sim F(6,6), a>0$满足$P(X>a)=0.05$, 则$P\left(X>\dfrac{1}{a}\right)=$ 　　.

解:

$\because X\sim F(6,6)\quad\therefore\dfrac{1}{X}\sim F(6,6)$

$\therefore P\left(\frac{1}{X}>a\right)=P(X>a)=0.05$

$\therefore P\left(X>\frac{1}{a}\right)=P\left(\frac{1}{X}<a\right)=1-P\left(\frac{1}{X}>a\right)=0.95$

例 6.11. 设X_1,X_2,X_3,X_4是来自$N\left(0,\sigma^2\right)$的一个样本

(1) 试证: $\left(X_1+X_2\right)^2$与$\left(X_1-X_2\right)^2$ 相互独立;

(2) 试求统计量$Y=\dfrac{2\left(X_3+X_4\right)^2}{\left(X_1+X_2\right)^2+\left(X_1-X_2\right)^2}$的分布.

解:

(1) 记X_1,X_2的样本均值$\overline{X}=\dfrac{X_1+X_2}{2}$, 样本方差$S^2=\sum_{i=1}^{2}\left(X_i-\overline{X}\right)^2$ 且$\overline{X}$与S^2.

而$\left(X_1+X_2\right)^2=4\left(\dfrac{X_1+X_2}{2}\right)^2=4\overline{X}^2,\left(X_1-X_2\right)^2=2\left(X_1-\dfrac{X_1+X_2}{2}\right)^2+2\left(X_2-\dfrac{X_1+X_2}{2}\right)^2$

$2S^2$.

$\therefore\left(X_1+X_2\right)^2$ ffi $\left(X_1-X_2\right)^2$相互独立.

(2) 由已知, $X_1\pm X_2\sim N\left(0,2\sigma^2\right)$, $X_3+X_4\sim N\left(0.2\sigma^3\right)$

则$\dfrac{X_3+X_4}{\sqrt{2}\sigma}\sim N(0,1)$, $\dfrac{\left(X_3+X_4\right)^2}{2\sigma^2}\sim\chi^2(1)$ $\dfrac{X_1\pm X_2}{\sqrt{2}\sigma}\sim N(0,1)$, $\dfrac{\left(X_1\pm X_2\right)^2}{2\sigma^2}\sim$
$\chi^2(1)$

又由$\left(X_1+X_2\right)^2$ ffi $\left(X_1-X_2\right)^2$相互独立可知 $\dfrac{\left(X_1+X_2\right)^2}{2\sigma^2}+\dfrac{\left(X_1-X_2\right)^2}{2\sigma^2}\sim\chi^2(2)$

且$\dfrac{\left(X_3+X_4\right)^2}{2\sigma^2}$, 与$\dfrac{\left(X_1+X_2\right)^2}{2\sigma^2}+\dfrac{\left(X_1-X_2\right)^2}{2\sigma^2}$独立

第七章　参数估计

参数估计是已知总体的分布(或近似分布)的类型, 但其中可能含有某些未知参数, 我们需要由样本来推断总体的这些未知参数. 本章的主要内容主要包括点估计、点估计优良性的评判标准以及总体均值和方差的估计内容等. 本章在期末考试中主要是以大题的形式进行考察, 矩估计和最大似然估计中二选一考察一个大题, 区间估计考察一个大题, 对于区间估计的复习应区分单侧置信区间和双侧置信区间的概念, 注重公式推导的方法. 估计量的评选标准则是作为上述大题中, 一般不会单独考察.

7.1　点估计

7.1.1　矩估计

1. 首先区分两个概念估计量与估计值.

   设总体的分布已知, $X_1, X_2, \cdots, X_n$为总体的一个样本, $\theta_1, \theta_2, \cdots, \theta_n$为总体的未知参数.$\theta_i(X_1, X_2, \cdots, X_n)$为估计量(本质上是一个随机变量)；而$\theta_i(x_1, x_2, \cdots, x_n)$是估计值。

2. 2.矩估计的一般步骤为:

   (a) 确定列出矩法方程的个数. 有几个未知参数就列几个方程;

   (b) 列出矩阵方程. 以两个未知参数为例:

   $$\begin{cases} E(X) = g_1(\theta_1, \theta_2) \\ E(X^2) = g_2(\theta_1, \theta_2) \end{cases}$$

   (c) 解方程组, 求出θ_1和θ_1的值.

例 7.1. 设总体$X \sim U[a,b]$, a,b未知; $x_1, x_2, \cdots, x_n$为X的样本, 求a,b的矩估计量.

解:

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

$$E\left(X^2\right) = DX + (EX)^2 = \frac{(a+b)^2}{12} + \left(\frac{a+b}{2}\right)^2$$

$$令 \frac{a+b}{2} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\frac{(a+b)^2}{12} + \left(\frac{a+b}{2}\right)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

$$由上述两式可得:$$

$$\frac{(a-b)^2}{12} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = S_n^2$$

$$所以 b - a = 2\sqrt{3}S_n \quad a + b = 2\bar{X}$$

$$解得: \hat{a} = \bar{X} - \sqrt{3}S_n \quad \hat{b} = \bar{X} + \sqrt{3}S_n$$

.

例 7.2. 设随机变量X的概率密度函数为

$$f(x) = \begin{cases} \sqrt{\theta}x^{\sqrt{\theta}} - 1, & 0 \le x \le 1 \\ 0, & 其他 \end{cases}$$

求未知参数θ的值.

解:

$$设 E(X) = \int_0^1 \sqrt{\theta}x^{\sqrt{\theta}} - 1dx = \frac{\sqrt{\theta}}{\sqrt{\theta}+1}$$

$$令 E(X) = \frac{\sqrt{\theta}}{\sqrt{\theta}+1} = \bar{X}$$

$$解得: \bar{\theta} = \left(\frac{\bar{X}}{1-\bar{X}}\right)^2$$

7.1.2 最大似然估计

1. 极大似然估计的基本原理:

当样本取值为$X_1=x_1,X_2=x_2,\cdots,X_n=x_n$,$P(X_1=x_1,X_2=x_2,\cdots,X_n=x_n,\theta)$可以看做参数$\theta$的函数,$\theta$的极大似然估计就是在参数空间$\Theta$中寻找一个参数$\theta_0$. 在参数取值为$\theta_0$时, 使得$\{X_1=x_1,X_2=x_2,\cdots,X_n=x_n\}$发生的概率是最大的.

2. 离散情况的解题步骤

(a) X为离散型随机变量, 其分布率为$P(X=x)=p(x,\theta)$

(b) 列出似然函数.

$$\begin{aligned}&P(X_1=x_1,X_2=x_2,\cdots,X_n=x_n)\\&=P(X_1=x_1)P(X_2=x_2)\cdots P(X_n=x_n)\\&=\prod_{i=1}^{n}f(x_i;\theta)\end{aligned}$$

(c) 选择θ_0使$L(\theta)=\prod_{i=1}^{n}f(x_i;\theta)$达到最大. 等价问题就是对最大似然函数求导, 令导函数等于0即可.为了方便求解我们有时也把$ln((L))$作为最大似然函数. (对数函数具有单调性, 两个函数到函数的零点是相同的, 这么处理完全是为了方便求解).

例 7.3. 设$X\sim B(1,p)$; $X_1,X_2,\cdots,X_n$是来X的样本, 试求参数p的极大似然估计量.

解:设$x_1,x_2,\cdots,x_n$是一个样本值. X的分布律为:

$$P(X=x)=p^x(1-p)^{1-x},x=0,1$$

故似然函数为

$$ln(L(p))=\left(\sum_{i=1}^{n}x_i\right)lnp+\left(n-\sum_{i=1}^{n}x_i\right)ln(1-p)$$

令

$$\frac{d}{dp}ln(L(p))=\frac{\sum_{i=1}^{n}x_i}{p}-\frac{n-\sum_{i=1}^{n}x_i}{1-p}=0$$

解得p的极大似然估计值为: $\hat{p}=\dfrac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}$;

p的极大似然估计量为: $\hat{p}=\dfrac{1}{n}\sum_{i=1}^{n}x_i=\bar{X}$

3. 连续情况以及多变量时的解题步骤

(a) 连续情况时的似然函数是将概率密度函数相乘即可得到.

(b) 多变量时分别对每个变量求偏导即可.

例 7.4. *设总体,$X \sim U[a,b]$,a,b未知; $x_1, x_2, \cdots, x_n$为X的样本值, 求a,b的极大似然估计量.*

解: 设$x_{(1)} = min(x_1, x_2, \cdots, x_n)$, $x_{(n)} = max(x_1, x_2, \cdots, x_n)$, X的概率密度函数为

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

因为$a \le x_1, x_2, \cdots, x_n \le b$等价于$a \le x_1, x_2, \cdots, x_n \le b$,

$$L(a,b) = \begin{cases} \frac{1}{b-a^n}, & a \le x_{(1)}, b \le x_{(n)} \\ 0, & \text{其他} \end{cases}$$

注意: 由于

$$\begin{cases} \frac{\partial L}{\partial a} = \frac{n}{b-a^{n+1}} > 0 \\ \frac{\partial L}{\partial b} = -\frac{n}{b-a^{n+1}} < 0 \end{cases}$$

所以直接从似然函数本身考虑求最大值:

因为$\frac{\partial L}{\partial a} = \frac{n}{b-a^{n+1}} > 0$, 所以$L(\theta)$关于$a$递增, 而$a \le x_{(1)}$, 所以取

$$a = x_{(1)} = min(x_1, x_2, \cdots, x_n)$$

同理可取

$$b = x_{(n)} = max(x_1, x_2, \cdots, x_n)$$

故a,b的极大似然估计值为:

$$\hat{a} = x_{(1)} = min(x_i), \quad \hat{b} = x_{(n)} = max(x_i)$$

a,b的极大似然估计量为:

$$\hat{a} = X_{(1)} = min(X_i), \quad \hat{b} = X_{(n)} = max(X_i)$$

7.2 估计量的评选标准

7.2.1 无偏性

1. 再次强调一遍估计量是一个随机变量, 它具有统计特性, 我们可以对估计量求均值和方差.

2. 既然估计量是随机变量, 那么他的估计值跟我们取样有关, 一次取样的估计值往往不能反应参数真实值.　无偏性就是刻画估计量均值与其真实值是否相同, 即$E\left(\hat{\theta}\right)=\theta$.

3. 有时候$E\left(\hat{\theta}\right)$的值还与我们每次取样的数目有关，当$E\left(\hat{\theta}\right)-\theta\neq 0$,但$\lim\limits_{n\to\infty}E\left(\hat{\theta}\right)-\theta=0$时, 则称为渐进无偏估计量.

例 7.5. 证明总体X方差σ^2未知,σ^2的估计量$S_n^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$是有偏的.

证: 已有:

$$S_n^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}^2, E\left(X_i\right)=E\left(X\right)=E\left(\bar{X}\right)\mu, D\left(X_i\right)=D\left(X\right)=\sigma^2, D\left(\bar{X}\right)=\frac{\sigma^2}{n}$$

又由

$$E\left(X_i^2\right)=D\left(X_i\right)+\left(E_i\right)^2=\sigma^2+\mu^2$$

所以:

$$E\left(S_n^2\right)=\frac{1}{n}\sum_{i=1}^{n}\left[D\left(X_i+\left(EX_i\right)^2\right)\right]-\left[D\left(\bar{X}\right)+\left(E\bar{X}\right)^2\right]$$

由此可得:

$$E\left(S_n^2\right)=\left(\sigma^2+\mu^2\right)-\left(\frac{\sigma^2}{n}+\mu^2\right)=\frac{n-1}{n}\sigma^2$$

因此S_n^2是σ^2的有偏估计量，也是渐进无偏的估计量。

7.2.2　有效性

一个参数的估计量往往并不唯一, 有效性就是对估计量求方差, 谁的方差小, 就说谁更有效. 设$\hat{\theta}_1$和$\hat{\theta}_2$都是参数θ的无偏估计量, 若$D\left(\hat{\theta}_1\right)<D\left(\hat{\theta}_1\right)$, 则称$\hat{\theta}_1$较$\hat{\theta}_2$有效.

例 7.6. $X\sim N\left(0,1\right)$,$\left(X_1,X_2\right)$是X的样本, 设:

$$\begin{aligned}\hat{\mu}_1&=\frac{2}{3}X_1+\frac{1}{3}X_2\\ \hat{\mu}_2&=\frac{1}{4}X_1+\frac{1}{4}X_2\\ \hat{\mu}_3&=\frac{1}{2}X_1+\frac{1}{2}X_2\end{aligned}$$

则$\hat{\mu}_1,\hat{\mu}_2,\hat{\mu}_3$都是$\mu$的无偏估计量, 由于$D\left(\hat{\mu}_1\right)=\frac{5}{9}$, $D\left(\hat{\mu}_2\right)=\frac{5}{8}$, $D\left(\hat{\mu}_3\right)=\frac{1}{2}$, 所以$\hat{\mu}_3$最有效.

例 7.7. 设总体X的密度函数为

$$p(x;\theta)=\begin{cases}\frac{1}{\theta}e^{-\frac{x}{\theta}} & ,x>0\\ 0 & ,x\leq 0\end{cases}$$

$(X_1,X_2,\cdots,X_n)$为X的一个样本. 设$Z=min\{X_1,X_2,\cdots,X_n\}$试证当$n>1$时, θ的无偏估计量$\bar{X}$较nZ有效. 此处先给出$Z_1=min\{X_1,X_2,\cdots,X_n\}$和$Z_2=max\{X_1,X_2,\cdots,X_n\}$分布函数,可以作为结论记下来:

设随机变量$X_1,X_2,\cdots,X_n$相互独立, 其分布函数分别为$F_{X_i}(x_i),(i=1,2,\cdots,n)$ 则$M=max(X_1,X_2,\cdots,X_n)$和$N=min(X_1,X_2,\cdots,X_n)$的分布函数分别为:

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

和:

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

特别地, 当$X_1,X_2,\cdots,X_n$相互独立且具有相同分布函数$F(x)$时, 有:

$$F_M(z)=[F(z)]^n,\quad F_N(z)=1-[1-F(z)]^n$$

所以本题可以得到如下证明:

由于$X\sim Exp\left(\frac{1}{\theta}\right)$故

$$E\left(\bar{X}\right)=E(X)=\theta,\quad D\left(\bar{X}\right)=\frac{\theta^2}{n},\quad Z\sim Exp\left(\frac{n}{\theta}\right)$$

所以可得:

$$E(Z)=\frac{\theta}{n},\quad D(Z)=\frac{\theta^2}{n^2}$$

故由$D(nZ)=\theta^2$, 当$n>1$时, $D(nZ)>D\left(\bar{X}\right)$ 故θ的无偏估计量$\bar{X}$较nZ有效

7.2.3 相合性(一致性)

这部分不作为考试的重点, 仅在这里给出概念.

定义 7.1. 设$(X_1,X_2,\cdots,X_n)$是总体X的样本, $\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$是总体参数$\theta$的估计量. 若$\hat{\theta}$依概率收敛到$\theta$, 即$\lim\limits_{n\to\infty}P\left(\left|\theta-\hat{\theta}\right|\geq\varepsilon\right)=0$, 则称$\hat{\theta}$是$\theta$的相合估计量(一致估计量). 若$\hat{\theta}$满足$\lim\limits_{n\to\infty}E\left(\theta-\hat{\theta}\right)^2=0$, 则称$\hat{\theta}$是$\theta$的均方相合估计量.

7.3　区间估计

有时求出未知参数的点估计值比较苛刻, 能够得出未知参数的大致范围也是可行的.未知参数是未知的、确定的, 根据所给样本得出一个随机区间, 是其包含(套住)参数真值的概率达到指定的要求就是区间估计.

7.3.1　双侧区间估计

$(X_1, X_2, \cdots, X_n)$是总体X的样本, θ是总体的未知参数, 对于给定值$\alpha\,(0<\alpha<1)$, 若统计量$\hat{\theta}_1\,(X_1, X_2, \cdots, X_n)$和$\hat{\theta}_2\,(X_1, X_2, \cdots, X_n)$满足:

$$P\left\{\hat{\theta}_1\,(X_1, X_2, \cdots, X_n) < \theta < \hat{\theta}_2\,(X_1, X_2, \cdots, X_n)\right\} = 1-\alpha$$

则称随机区间$\left(\hat{\theta}_1, \hat{\theta}_2\right)$为$\theta$的置信度为$1-\alpha$的置信区间, $\hat{\theta}_1$和$\hat{\theta}_2$分别称为置信下限和置信上限.

7.3.2　单侧区间估计

$(X_1, X_2, \cdots, X_n)$是总体X的样本, θ是总体的未知参数, 对于给定值$\alpha\,(0<\alpha<1)$,

1. 若统计量$\hat{\theta}_1\,(X_1, X_2, \cdots, X_n)$满足:

   $$P\left\{\theta > \hat{\theta}_1\,(X_1, X_2, \cdots, X_n)\right\} = 1-\alpha$$

   则称随机区间$\left(\hat{\theta}_1, +\infty\right)$为$\theta$的置信度为$1-\alpha$的单侧置信区间, $\hat{\theta}_1$和称为单侧置信下限.

2. 若统计量$\hat{\theta}_2\,(X_1, X_2, \cdots, X_n)$满足:

   $$P\left\{\theta < \hat{\theta}_2\,(X_1, X_2, \cdots, X_n)\right\} = 1-\alpha$$

   则称随机区间$\left(-\infty, \hat{\theta}_2\right)$为$\theta$的置信度为$1-\alpha$的单侧置信区间, $\hat{\theta}_2$称为单侧置信上限.

7.3.3　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、分布不依赖于待估参数.

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

3. 由$a<g\left(\theta;X_1,X_2,\cdots,X_n\right)<b$, 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.

7.3.4 一个正态总体参数的区间估计

1. 均值μ的置信区间

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

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

2. 方差σ^2的置信区间

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

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

7.3.5 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

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

   (b) σ_1^2,σ_2^2未知: $T=\dfrac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_1-\mu_2\right)}{S_W\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t\left(n_1+n_2-2\right)$

   (其中$S_W=\sqrt{\frac{(n_1-1)S_{1n_1}^2+(n_2-1)S_{2n_2}^2}{n_1+n_2-2}}$)

2. 方差比$\frac{\sigma_1^2}{\sigma_2^2}$的置信区间

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

   (b) μ_1,μ_2未知: $F=\dfrac{\sigma_2^2S_{1n_1}^2}{\sigma_1^2S_{2n_2}^2}\sim F\left(n_1-1,n_2-1\right)$

例 7.8. 某厂生产一批滚珠, 其直径X服从$N\left(\mu,\sigma^2\right)$, 现从某天的产品中随机抽取*6*件, 测得直径为$15.1,14.8,15.2,14.9,14.6,15.1$, 求解以下问题:

1. 若$\sigma^2=0.06$，求μ的置信度为95%的置信区间；

2. 若σ^2未知，求μ的置信度为95%的置信区间；

3. 求方差σ^2的置信度为95%的置信区间.

$(U_0.025=1.96;\bar{x}=14.95;n=6;s=0.226;s^2=0.051$
$t_0.025\left(5\right)=2.5706;\chi_0^2.025\left(5\right)=12.833;\chi_0^2.975=0.831)$

1. 若$\sigma^2=0.06$，求μ的置信度为95%的置信区间

   解：由

$$U=\frac{\sqrt{n}\left(\bar{X}-\mu\right)}{\sigma}\sim N\left(0,1\right)$$

   可得置信区间为：

$$\left(\bar{X}-\frac{\sigma}{\sqrt{n}}u_{a/2},\bar{X}+\frac{\sigma}{\sqrt{n}}u_{a/2}\right)=\left(14.75,15.15\right)$$

2. 若σ^2未知，求μ的置信度为95%的置信区间

   解：由

$$T=\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t\left(n-1\right)$$

   可得置信区间为：

$$\left(\bar{X}-\frac{S}{\sqrt{n}}t_{a/2}\left(n-1\right),\bar{X}+\frac{S}{\sqrt{n}}t_{a/2}\left(n-1\right)\right)=\left(14.71,15.187\right)$$

3. 求方差σ^2的置信度为95%的置信区间

   解：由

$$\chi^2=\frac{\left(n-1\right)S^2}{\sigma^2}\sim\chi^2\left(n-1\right)$$

   可得置信区间为：

$$\left(\frac{\left(n-1\right)S^2}{\chi^2_{a/2}\left(n-1\right)},\frac{\left(n-1\right)S^2}{\chi^2_{1-a/2}\left(n-1\right)}\right)=\left(0.0199,0.3069\right)$$

第八章　假设检验

8.1　假设检验的基本概念

- 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.

- 当总体的分布类型已知, 对总体分布中未知参数的假设称为参数假设, 用样本检验这类假设称为参数假设检验或称参数检验.

- 当总体分布类型未知, 对总体分布类型或分布性质的假设称为分布假设, 用样本检验这类假设称为分布假设检验.

8.1.1　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, “一个小概率事件在一次实验中几乎是不可能发生的”, 因此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. 因此可以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.
需要指出, 假设检验方法是建立在实际推断原理的基础上的, 但是仅利用它是不够的，还应当与一些直观分析相结合.

8.1.2　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. 记为$P\{拒绝H_0|H_0为真\} = \alpha$

- 第二类错误(存伪错误): 原假设不真H_0, 但接受了H_0. 记为$P\{接受H_0|H_0不真\} = \beta$

我们通常是控制犯第一类错误, 使它不超过某个给定的值α, 这种对犯第一类错误的概率加以控制而不考虑第二类错误的检验问题, 称为显著性检验问题. 若要求犯错误概率不超过α, 则可称这个检验法为显著性水平为α的检验法, 其中α称为检验的显著性水平.

8.1.3 原假设和备择假设的不平等性

原假设和备择假H_1设地位不平等, H_0受特殊保护. 接受了H_0, 并不一定说明H_0一定为真, 因为有概率H_0实际不真但被误认为是真的. 若在显著性水平α较小时仍拒绝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.

8.1.4 假设检验的一般步骤

临界值检验法

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$P\{W|H_0为真\} = \alpha$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.

p值检验法

在假设检验中,利用样本值做出的拒绝原假设的最小显著性水平称为p值, 其检验准则为:

- 若p值不大于检验水平α, 则拒绝原假设.
- 若p值大于检验水平α, 则接受原假设.

临界值检验法和p值检验法是等价的.

8.1.5　双侧和单侧假设检验

1. 双侧假设检验的一般形式为:

$$H_0 : \theta = \theta_0, \quad H_1 : \theta \neq \theta_0$$

其中备择假设H_1的参数区域在原假设H_0参数区域的两边, 所以其拒绝域有人在其接受域的两边, 这样的假设也称为双边假设。

2. 单侧假设检验的一般形式为:

    (a) $H_0 : \theta = \theta_0, \quad H_1 : \theta > \theta_0$

    (b) $H_0 : \theta = \theta_0, \quad H_1 : \theta < \theta_0$

    (c) $H_0 : \theta \leq \theta_0, \quad H_1 : \theta > \theta_0$

    (d) $H_0 : \theta \geq \theta_0, \quad H_1 : \theta < \theta_0$

其中备择假设H_1的参数区域在原假设H_0参数区域的一边, 所以其拒绝域有人在其接受域的一侧, 这样的假设也称为单边假设。

例 8.1. *假设$X \sim N(0,1)$, 其中方差σ^2已知,对μ进行显著性水平为α的显著性检验.*

解: 构造检验统计量$U = \dfrac{\sqrt{n}\left(\bar{X} - \mu_0\right)}{\sigma} \sim N(0,1)$

1. 对于双边假设检验, 假设为$H_0 : \theta = \theta_0, \quad H_1 : \theta \neq \theta_0$.

因为$\bar{X}$是μ的无偏估计量, 所以当H_0为真时, $|u|$不应太大, 当H_1成立时, $|u|$有偏大的趋势, 故拒绝域形式为

$$|u| = |\frac{\sqrt{n}\,(\bar{x} - \mu_0)}{\sigma_0}| \geq k\,(k\text{待定})$$

查标准正态分布表可得$k = u_{\alpha/2}$,使

$$P\left\{|U| \geq u_{\alpha/2}\right\} = \alpha$$

即得拒绝域为

$$|u| \geq u_{\alpha/2}$$

2. 对于单边假设检验, 假设为$H_0: \theta = \theta_0, \quad H_1: \theta > \theta_0$.

这意味着均值μ不会小于μ_0, 我们同样由$U \sim N(0,1)$出发, 可以得到拒绝域形式为

$$u = \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma_0} \geq k \,(k\text{待定})$$

查标准正态分布表可得$k = u_\alpha$,使

$$P\{U \geq u_\alpha\} = \alpha$$

即得拒绝域为

$$u \geq u_\alpha$$

由类似的方法可以得到, 对于上文中四个单边假设:

(a)、(c)的拒绝域为$u \geq u_\alpha$

(b)、(d)的拒绝域为$u \leq -u_\alpha$

8.2　正态总体参数的假设检验

在实际问题中, 常常有理由假定总体服从正态分布. 本节中, 当检验统计量服从正态分布、χ^2分布、t分布、F分布时, 本别称为u检验法、χ^2检验法、t检验法、F检验法.假设检验的具体原理已由8.1说明, 此处仅总结正太总体参数的假设检验中的不同情况下的统计量选择以及不同条件下的收敛域. 也可以参考见书上第191页的表格.

1. 方差σ_0^2已知时, 对μ进行检验

使用的检验统计量及分布为:

$$U = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma_0} \sim N(0,1)$$

假设与拒绝域的对应关系如下:

H_0	H_1	拒绝域
$\mu = \mu_0$	$\mu \neq \mu_0$	$\left\lvert \frac{\sqrt{n}(\bar{x}-\mu_0)}{\sigma_0} \right\rvert \geq u_{\alpha/2}$
$\mu \leq \mu_0$(或$\mu = \mu_0$)	$\mu > \mu_0$	$\frac{\sqrt{n}(\bar{x}-\mu_0)}{\sigma_0} \geq u_\alpha$
$\mu \geq \mu_0$(或$\mu = \mu_0$)	$\mu < \mu_0$	$\frac{\sqrt{n}(\bar{x}-\mu_0)}{\sigma_0} \leq -u_\alpha$

2. 方差σ^2未知时, 对μ进行检验

使用的检验统计量及分布为:

$$T=\frac{\sqrt{n}\left(\bar{X}-\mu_0\right)}{S}\sim t\left(n-1\right)$$

假设与拒绝域的对应关系如下:

H_0	H_1	拒绝域
$\mu=\mu_0$	$\mu\neq\mu_0$	$\left\lvert\frac{\sqrt{n}(\bar{x}-\mu_0)}{s}\right\rvert\geq t_{\alpha/2}\left(n-1\right)$
$\mu\leq\mu_0$(或$\mu=\mu_0$)	$\mu>\mu_0$	$\frac{\sqrt{n}(\bar{x}-\mu_0)}{s}\geq t_{\alpha}\left(n-1\right)$
$\mu\geq\mu_0$(或$\mu=\mu_0$)	$\mu<\mu_0$	$\frac{\sqrt{n}(\bar{x}-\mu_0)}{s}\leq -t_{\alpha}\left(n-1\right)$

3. 均值μ未知时, 对σ^2进行检验

使用的检验统计量及分布为:

$$\chi^2=\frac{\left(n-1\right)S^2}{\sigma_0^2}\sim\chi^2\left(n-1\right)$$

假设与拒绝域的对应关系如下:

H_0	H_1	拒绝域
$\sigma^2=\sigma_0^2$	$\sigma^2\neq\sigma_0^2$	$\frac{(n-1)s^2}{\sigma_0^2}\leq\chi^2_{1-\alpha/2}\left(n-1\right)$或$\frac{(n-1)s^2}{\sigma_0^2}\geq\chi^2_{\alpha/2}\left(n-1\right)$
$\sigma^2\leq\sigma_0^2$(或$\sigma^2=\sigma_0^2$)	$\sigma^2>\sigma_0^2$	$\frac{(n-1)s^2}{\sigma_0^2}\geq\chi^2_{\alpha}\left(n-1\right)$
$\sigma^2\geq\sigma_0^2$(或$\sigma^2=\sigma_0^2$)	$\sigma^2<\sigma_0^2$	$\frac{(n-1)s^2}{\sigma_0^2}\leq\chi^2_{1-\alpha}\left(n-1\right)$

4. 方差σ_1^2,σ_2^2未知但$\sigma_1^2=\sigma_2^2$时, 对$\mu_1-\mu_2$进行检验

使用的检验统计量及分布为:

$$T=\frac{\left(\bar{X}-\bar{Y}\right)-c}{S_W\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t\left(n_1+n_2-2\right)$$

$$\text{其中}S_W^2=\frac{\left(n_1-1\right)S_1^2n_1+\left(n_2-1\right)S_2^2n_2}{n_1+n_2-2}$$

假设与拒绝域的对应关系如下:

H_0	H_1	拒绝域
$\mu_1-\mu_2=c$	$\mu_1-\mu_2\neq c$	$\left\lvert\frac{(\bar{x}-\bar{y})-c}{s_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right\rvert\geq t_{\alpha/2}\left(n_1+n_2-2\right)$
$\mu_1-\mu_2\leq c$(或$\mu_1-\mu_2=c$)	$\mu_1-\mu_2>c$	$\frac{(\bar{x}-\bar{y})-c}{s_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\geq t_{\alpha}\left(n_1+n_2-2\right)$
$\mu_1-\mu_2\geq c$(或$\mu_1-\mu_2=c$)	$\mu_1-\mu_2<c$	$\frac{(\bar{x}-\bar{y})-c}{s_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\leq -t_{\alpha}\left(n_1+n_2-2\right)$

5. 均值μ_1, mu_2未知, 对$\frac{\sigma_1^2}{\sigma_2^2}$进行检验

使用的检验统计量及分布为:

$$F=\frac{S_1^2 n_1}{c S_1^2 n_1} \sim F\left(n_1-1, n_2-1\right)$$

假设与拒绝域的对应关系如下:

H_0	H_1	拒绝域
$\frac{\sigma_1^2}{\sigma_2^2}=c$	$\frac{\sigma_1^2}{\sigma_2^2} \neq c$	$\frac{S_1^2 n_1}{c S_1^2 n_1} \leq F_{1-\alpha/2}\left(n_1-1, n_2-2\right)$ 或$\frac{S_1^2 n_1}{c S_1^2 n_1} \geq F_{\alpha/2}\left(n_1-1, n_2-2\right)$
$\frac{\sigma_1^2}{\sigma_2^2} \leq c$(或$\frac{\sigma_1^2}{\sigma_2^2}=c$)	$\frac{\sigma_1^2}{\sigma_2^2}>c$	$\frac{S_1^2 n_1}{c S_1^2 n_1} \geq F_{\alpha}\left(n_1-1, n_2-2\right)$
$\frac{\sigma_1^2}{\sigma_2^2} \geq c$(或$\frac{\sigma_1^2}{\sigma_2^2}=c$)	$\frac{\sigma_1^2}{\sigma_2^2}<c$	$\frac{S_1^2 n_1}{c S_1^2 n_1} \leq F_{1-\alpha}\left(n_1-1, n_2-2\right)$

在熟记内容的同时,还需要对前面章节中不同分布的性质和分布之间的关系有足够的了解, 便于加深理解并在解题时熟练运用.(当然此处不一定是考试的重点,具体要看任课老师最终的要求).

8.3 非正态总体的假设检验

8.3.1 非正态总体的大样本检验

在8.2中最重要也是最困难的步骤是确定检验统计量Z以及其分布. 当Z的确切分布难以找到或过于复杂不便应用时, 可以考虑使用Z的近似分布, 特别是在样本容量较大时, 由用中心极限定理, 用Z的渐进分布作为Z的近似分布. 同时应注意, 这时由给定的显著性水平α得到的检验法, 其显著性水平近似为α, 而非确切为α.

例 8.2. 设总体X的分布是任意的，其均值$E(X)=\mu$和方差$D(X)=\sigma_0^2$存在，且$\sigma_0^2>0$已知. $(X_1,X_2,\cdots,X_n)$是来自总体X的样本，n比较大，要检验假设：

$$H_0:\mu=\mu_0,\quad H_1:\mu\neq\mu_0\quad (\mu_0\text{为已知数})$$

解：采用

$$U=\frac{\sqrt{n}\left(\bar{X}-\mu_0\right)}{\sigma_0}$$

为检验统计量. 因为$\bar{X}$是μ的无偏估计量，所以，当H_0为真时，$|u|$应偏小，当H_1为真时，$|u|$应偏大，因此拒绝域形式应为：

$$|u|=\left|\frac{\sqrt{n}\left(\bar{x}-\mu_0\right)}{\sigma_0}\right|\geq k(k\text{待定})$$

要由给定的显著性水平α确定k，就要在H_0为真的条件下求得U的分布，因为总体X的分布形式未知，所以无法求出U的确切分布. 但是，由中心极限定理可知，当H_0为真且n很大时，U近似地服从标准正态分布$N(0,1)$，因此，当n很大时，对给定的α，查标准正态分布表得$u_{\alpha/2}$，使

$$P\left\{|u|\geq u_{\alpha/2}\right\}\approx\alpha$$

即得到拒绝域

$$|u|=\left|\frac{\sqrt{n}\left(\bar{x}-\mu_0\right)}{\sigma_0}\right|\geq u_{\alpha/2}$$

由一次具体抽样后所得的样本值$(x_1,x_2,\cdots,x_n)$计算出U的观测值u，若$|u|\geq u_{\alpha/2}$，则拒绝H_0，否则接受H_0.
此大样本检验法的p值为$p=P\{|U|\geq|u|\}\approx P\{|Z|\geq|u|\}$，其中$Z\sim N(0,1)$. u是检验统计量U的观测值.

8.3.2 非正态总体的小样本检验

上面已经提到，在样本容量较大时，可以使用大样本检验的方法进行非正态总体的假设检验. 而当非正态总体样本容量较小时，使用大样本检验会出现一些误差. 此时我们可以考虑依据总体符合的分布(非正态分布)自行构造合适的统计量来完成假设检验.

例 8.3. 某厂家声称他们生产的电子元件平均寿命不低于*2000*小时，现随机抽取*10*个这种电子元件，测得其寿命(单位：小时)为：*717, 1996, 2532, 4213, 1574, 5437, 925, 3654, 6273, 2150,* 在检验水平下能否相信厂家的说法？

解: 设电子元件的寿命$X \sim Exp(\lambda)$, λ未知, 则要检验

$$H_0 : \frac{1}{\lambda} \geq \frac{1}{\lambda_0} = 2000, \quad H_1 : \frac{1}{\lambda} < \frac{1}{\lambda_0}$$

或

$$H_0 : \lambda \leq \lambda_0 = \frac{1}{2000}, \quad H_1 : \lambda > \lambda_0$$

因为检验的是总体的期望, 因此从样本均值出发构造统计量.

由于$2n\lambda\bar{X} = 2\lambda\sum_{i=1}^{n} X_i \sim \chi^2(2n)$(参加书上习题6第25题), 因此选取检验统计量为$\chi^2 = 2n\lambda_0\bar{X}$, 且当$\lambda = \lambda_0$时, $\chi^2 \sim \chi^2(2n)$, 拒绝域为$\chi^2 \leq \chi^2_{1-\alpha}(2n)$.

计算得, $\chi^2 = 29.471 > \chi^2_{0.95}(20) = 10.851$, 故接受原假设, 即认为电子元件的平均寿命超过2000小时.

此检验的p值为$p = P\{\chi^2 \leq 29.471\}\, 0.921 > 0.05$, 因此接受原假设.

8.4　成对数据的假设检验

如果样本是来自同一总体的不同观测, 它们成对出现, 那么称这样的数据为成对数据. 对于成对数据的假设检验问题, 我们往往将问题转化为单个正态总体的假设检验问题进行处理.

设有n对相互独立的样本$(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$, 令$Z_i = Y_i - X_i$, $i = 1, 2, \cdots, n$, 则$Z_1, Z_2, \cdots, Z_n$相互独立, 设$Z_1, Z_2, \cdots, Z_n$是来自正态总体$N \sim (\mu, \sigma)$的样本,则可以按照8.2中的方式对Z进行检验. 如在上述问题中可检验:

$$H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0;$$

$$H_0 : \mu \leq \mu_0, \quad H_1 : \mu > \mu_0;$$

$$H_0 : \mu \geq \mu_0, \quad H_1 : \mu < \mu_0.$$

尽管在8.2的两个正态总体的假设检验中要求两个总体相互独立, 相对应的, 所抽样本也应相互独立. 而实际应用中这样的要求不一定可以满足. 例如我们想检验一种要的效果时, 需要检验每个人在服药前后数据的变化. 而每一个人服药前后的两个数据并不相互独立, 且由于个人体质的差异, 将这些人服药前后的数据看做来自同一个正态总体也是不合理的, 而本节中的方法则规避了上述问题.

例 8.4. 某培训班声称经过培训, 学员的掷实心球成绩可以提高1米以上. 为检验他们的说法是否符合实际, 随机选出9人记录他们培训前后的成绩如下(单位: m):

培训前成绩x_i	*9.63*	*7.61*	*6.28*	*8.32*	*5.40*	*5.82*	*6.89*	*8.17*	*5.78*
培训后成绩y_i	*10.24*	*8.75*	*7.05*	*8.91*	*6.51*	*6.58*	*8.02*	*9.08*	*6.32*
差值$y_i - x_i$	*0.61*	*1.14*	*0.77*	*0.59*	*1.11*	*0.76*	*1.13*	*0.91*	*0.54*

设差值数据来自正态总体，在检验水平0.05下，是否可以认为该培训班的说法是对的？
解: 设差值数据来自正态总体$Z \sim N(\mu, \sigma^2)$，则要求检验

$$H_0 : \mu \geq 1, \quad H_1 : \mu < 1$$

选取检验统计量$T = \dfrac{\bar{Z} - 1}{S_Z/\sqrt{9}}$，在$\mu = 1$时，$T \sim t(8)$，该检验的拒绝域为$t \leq -t_{0.05}(8)$
计算得$\bar{z} = 0.84, S_z = 0.242, t = -1.98 < -t_{0.05}(8) = -1.8595$，故拒绝原假设，即认为培训后平均成绩提高不到1米.
可以计算此检验的p值为$p = P\{T \leq -1.98\} = 0.0415 < 0.05$，因此拒绝原假设.

8.5　分布假设检验

8.5.1　分布拟合检验与皮尔逊定理

在前面各节的参数假设检验中，往往要求了解样本的总体分布. 而实际问题中往往仅有对总体的一些观测得到的数据资料(即样本值), 而对总体的分布形式一无所知. 面对上述情况, 我们总希望可以判断出总体属于哪种分布. 通常的办法是, 我们先依据一些关于总体的先验知识去假定一个分布, 再利用假设检验的方法来判断总体的分布是否与前面假定的分布一致.
上述假设检验问题可以表示为: 设$(X_1, X_2, \cdots, X_n)$是来自总体X的样本, 由此要检验假设

$$H_0 : X\text{的分布函数为}F(x), \quad H_1 : X\text{的分布函数不是}F(x)$$

这里$F(x)$是一个已知的分布函数(可包含未知参数). 该问题也可以等价为: 衡量理论分布$F(x)$对样本$(X_1, X_2, \cdots, X_n)$的拟合效果. 因此通常称这类假设检验为分布拟合检验. 此处仅介绍常用的χ^2拟合检验法, 首先需要了解一个重要定理.

定理 8.1 (皮耳逊定理). *设一个随机试验的r个结果$A_1, A_2, \cdots, A_r$构成互斥完备事件群，在一次试验中它们发生的概率分别为$p_1, p_2, \cdots, p_r$，其中$p_i > 0\,(i = 1, 2, \cdots, r)$，且$\sum\limits_{i=1}^{r} p_i =$*

1, 以m_i表示n次独立重复实验中A_i发生的次数，那么当$n \to \infty$时, 随机变量

$$\chi^2 = \sum_{i=1}^{r} \frac{m_i - np_i{}^2}{np_i}$$

的分布收敛于自由度为$n-1$的χ^2分布.

8.5.2　χ^2拟合检验法

1. 理论分布完全已知的情形 设$(X_1, X_2, \cdots, X_n)$是来自总体X的样本, $F(x)$是一个完全已知分布的函数, 由此要检验假设

$$H_0: X\text{的分布函数为}F(x), \quad H_1: X\text{的分布函数不是}F(x)$$

这种情况下χ^2拟合检验法的基本思想和步骤为:

(a) 把总体X的一切可能值的集合Ω划分成有限个子集$\Omega_1, \Omega_2, \cdots, \Omega_r$,记

$$A_i = \{X \in \Omega\} (i = 1, 2, \cdots r)$$

则$A_1, A_2, \cdots, A_r$构成完备事件群.

(b) 在原假设H_0为真的前提下, 计算事件A_i的概率

$$P(A_i) = p_i \quad (i = 1, 2, \cdots, r)$$

从而在n次观察中事件A_i发生的理论频数为$np_i\,(i = 1, 2, \cdots, r)$

(c) 考虑统计量

$$\chi^2 = \sum_{i=1}^{r} \frac{m_i - np_i{}^2}{np_i}$$

其中χ^2可以理解为样本的实际频数m_i对理论频数np_i偏差$m_i - np_i$的加权平方和, 它的大小刻画了理论分布$F(x)$与样本值$(X_1, X_2, \cdots, X_n)$的拟合程度, 其值越小, 说明$F(x)$与$(X_1, X_2, \cdots, X_n)$拟合的越好.因此, χ^2可以作为检验统计量, 其拒绝域为

$$\chi^2 \geq k\,(k\text{待定})$$

(d) 由皮尔逊定理, 当H_0为真且n充分大时,χ^2近似地服从自由度为$n-1$的χ^2分布, 所以对给定的显著性水平α,查分布表可得$\chi^2_\alpha(r-1)$, 使

$$P\left\{\chi^2 \geq \chi^2_\alpha(r-1)\right\} \approxeq \alpha$$

于是得拒绝域

$$\chi^2=\sum_{i=1}^{r}\frac{m_i-np_i{}^2}{np_i}\geq\chi^2_\alpha\left(r-1\right)$$

(e) 此检验法的p值为

$$p=P\left\{\chi^2\geq\chi_0^2\right\}\approxeq P\left\{Z\geq\chi_0^2\right\}$$

其中χ_0^2是检验统计量χ^2的观测值,$Z\sim\chi^2\left(r-1\right)$

运用上述方法检验总体分布时, 要求:

- 样本容量较大, 通常取$n\geq 50$
- 划分可能值时, 要求各组的理论频数$np_i\geq 5$
- 此外, 分布拟合检验时, 备择假设可以不必写出来

2. 理论分布含未知参数的情况在分布拟合检验中, 如果只知道理论分布$F\left(x\right)$的形式，而其中含有未知参数$\theta_1,\theta_2,\cdots,\theta_l$, 即$F=F\left(x;\theta_1,\theta_2,\cdots,\theta_l\right)$. 则要检验的假设可以表示为

$H_0:X$的分布函数为$F\left(x;\theta_1,\theta_2,\cdots,\theta_l\right)$, $H_1:X$的分布函数不是$F\left(x;\theta_1,\theta_2,\cdots,\theta_l\right)$

这种情况下可以先利用样本$(X_1,X_2,\cdots,X_n)$求出参数$(\theta_1,\theta_2,\cdots,\theta_l)$的最大似然估计值$\hat\theta_1,\hat\theta_2,\cdots,\hat\theta_l$, 并带入分布函数$F$中, 得到分布完全已知的分布函数$F\left(x;\hat\theta_1,\hat\theta_2,\cdots,\hat\theta_l\right)$, 由此即可继续按照1中的方法进行假设检验.

需要注意的是

(a) 在计算$p_i=P\left(A_i\right)$时, 要按$F\left(x;\hat\theta_1,\hat\theta_2,\cdots,\hat\theta_l\right)$计算其估计值$\hat p_i\left(i=1,2,\cdots,r\right)$

(b) 采用的统计量为

$$\chi^2=\sum_{i=1}^{r}\frac{\left(m_i-n\hat p_i\right)^2}{n\hat p_i}$$

当$n\to\infty$时, 其极限分布为自由度为$r-l-1$的χ^2分布(对皮耳逊定理的修正).

于是对于给定的显著性水平α, 得拒绝域为

$$\chi^2=\sum_{i=1}^{r}\frac{\left(m_i-n\hat p_i\right)^2}{n\hat p_i}\geq\chi^2_\alpha\left(r-l-1\right)$$

此检验法的p值为

$$p=P\left\{\chi^2\geq\chi_0^2\right\}\approxeq P\left\{Z\geq\chi_0^2\right\}$$

其中χ_0^2是检验统计量χ^2的观测值,$Z\sim\chi^2\left(r-l-1\right)$.不同于1中的情况.

第九章　回归分析

[1]

9.1　一元线性回归模型

设x为可控变量, Y为与之相关的随机变量. 当自变量x取确定值时, Y有一确定的条件分布与之对应. 如果Y的数学期望存在, 那么其取值随x的取值而定, 因而它是x的函数, 记为$\mu(x)$, 即

$$\mu(x) = E(Y|x)$$

通常称$\mu(x)$为Y关于x的回归函数.
回归分析的基本任务是利用试验数据来估计回归函数.
设$\mu(x) = ax + b$为线性函数, 此时估计$\mu(x)$的问题称为一元线性回归问题.

定义 9.1. *设x为可控变量, Y是依赖于x的随机变量, 假定*

$$\begin{cases} Y = a + bx + \varepsilon \\ \varepsilon \backsim N(0, \sigma^2) \end{cases}$$

其中参数a, b都不依赖于x, 上式称为一元线性回归模型.

由于a, b均未知, 需要我们从收集到的数据$(x_i, y_i)\,(i = 1, 2, \cdots, n)$出发进行估计. 在收集数据时, 我们一般要求观测独立地进行, 即假定$Y_1, Y_2, \cdots, Y_n$相互独立. 综合上述诸项假定, 我们可以给出最简单, 常用的一元线性回归的统计模型的样本模型

$$\begin{cases} Y_i = a + bx_i + \varepsilon_i \\ \varepsilon_i\text{独立同分布, 其分布为}N\left(0, \sigma^2\right), i = 1, 2, \cdots, n \end{cases}$$

[1]实际上, 考试范围是一元线性回归分析

由样本观测值$(x_i, y_i)\,(i = 1, 2, \cdots, n)$可以获得$a, b$的估计$\hat{a}, \hat{b}$, 称

$$\hat{y} = \hat{a} + \hat{b}x$$

为y关于x的经验回归函数, 简称为回归方程, 其图形称为回归直线.
给定$x = x_0$后, 称$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$为回归值(在不同场合也称其为拟合值、预测值).

9.2　参数的估计(OLSE)

采用最小二乘法估计参数a, b, 令

$$Q\left(a, b\right) = \sum_{i=1}^{n} \left(y_i - a - bx_i\right)^2,$$

$$\text{subject to } Q\left(\hat{a}, \hat{b}\right) = \min_{a,b} Q\left(a, b\right)$$

由于$Q \geqslant 0$, 且对a, b的导数存在, 因此最小二乘估计可以通过求偏导数得到

$$\begin{cases} \dfrac{\partial Q}{\partial a} = -2\sum_{i=1}^{n}\left(y_i - a - bx_i\right) = 0 \\ \dfrac{\partial Q}{\partial b} = -2\sum_{i=1}^{n}\left(y_i - a - bx_i\right)x_i = 0 \end{cases}$$

这组方程被称为**正规方程组**, 经验证可得系数矩阵行列式不为0, 经过整理可得估测参数为

$$\hat{b} = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

$$\hat{a} = \overline{y} - \hat{b}\overline{x}$$

第十章　随机过程

10.1　随机过程的相关概念与符号

定义 10.1 (随机过程)**.** 设对于每一个参数$t \in T, X(t,\omega)$是一随机变量, 称随机变量族$X_t = X(t,\omega), t \in T$为一随机过程. 其中$T \subset \mathbb{R}$, 称为指标集.
用映射来表示X_T,

$$X(t,\omega) : T \times \Omega \to \mathbb{R}$$

固定$t \in T, X(t,\cdot)$是定义在样本空间Ω上的函数, 即为一个随机变量. 对于$\omega \in \Omega, X(\cdot,\omega)$, 其中$t$在$T$中顺序变化, 是参数$t \in T$的一般函数, 通常称之为样本函数, 或者随机过程的一个实现或者轨道.
为了简明起见,

$$X_T \overset{def}{=} \{X(t), t \in T\}$$

X_T可以取复数, $\mathbb{R}^n$等, 可能取值的全体所构成集合称为状态空间, 记作S.

10.2　有限维分布函数

定义 10.2 (有限维分布函数族)**.** 对于任意有限个$t_1, t_2, \cdots, t_n \in T$, 定义过程$X_T$的$n$维分布函数$F_{t_1,t_2,\cdots,t_n}(x_1, x_2, \cdots, x_n)$ 为

$$F_{t_1,t_2,\cdots,t_n}(x_1, x_2, \cdots, x_n) \overset{def}{=} P\{X(t_1) \le x_1, X(t_2) \le x_2, \cdots, X(t_n) \le x_n\}$$

随机过程X_T的全部一维分布函数, 所有二维分布函数$\cdots$的全体称为X_T的有限维分布函数族.

其基本性质:

- 对称性

$$F_{t_1,t_2,\cdots,t_n}(x_1,x_2,\cdots,x_n)=F_{t_{j_1},t_{j_2},\cdots,t_{j_n}}(x_{j_1},x_{j_2},\cdots,x_{j_n})$$

- 相容性

$$F_{t_1,t_2,\cdots,t_m}(x_1,x_2,\cdots,x_m)=F_{t_1,t_2,\cdots,t_m,t_{m+1},\cdots,t_n}(x_1,x_2,\cdots,x_m,\infty,\cdots,\infty)$$

例 10.1. 设随机相位正弦波为 $X(t)=a\cos(t+\Theta)(-\infty<t<+\infty)$, 其中 a 是正常数, Θ 是在区间 $(0,2\pi)$ 上服从均匀分布的随机变量.

(1) 当 Θ 取值 $\theta=\frac{\pi}{4},\frac{\pi}{2},\pi$ 时, 相应的样本函数是什么?

(2) 求 $X(t)$ 在 $t=\frac{\pi}{4}$ 时的一维概率密度.

解:

(1) 当 Θ 取值 $\theta=\frac{\pi}{4},\frac{\pi}{2},\pi$ 时, 相应的样本函数分别是

$$x_1(t)=a\cos\left(t+\frac{\pi}{4}\right)(-\infty<t<+\infty);$$
$$x_2(t)=a\cos\left(t+\frac{\pi}{2}\right)=-a\sin t(-\infty<t<+\infty);$$
$$x_3(t)=a\cos(t+\pi)=-a\cos t(-\infty<t<+\infty).$$

(2) $X(t)$ 在 $t=\frac{\pi}{4}$ 时, $X\left(\frac{\pi}{4}\right)=a\cos\left(\frac{\pi}{4}+\Theta\right)$, Θ 的概率密度为

$$f(\theta)=\begin{cases}\frac{1}{2\pi}, & 0<\theta<2\pi\\ 0, & \text{其他}\end{cases}$$

令 $X=X\left(\frac{\pi}{4}\right)=a\cos\left(\frac{\pi}{4}+\Theta\right)$, X 的分布函数为

$$F(x)=P\{X\leqslant x\}=P\left\{a\cos\left(\frac{\pi}{4}+\Theta\right)\leqslant x\right\}=\int_{a\cos\left(\frac{\pi}{4}+\theta\right)\leqslant x}f(\theta)\mathrm{d}\theta,$$

当 $x<-a$ 时, $F(x)=P\{\phi\}=0$;

当 $x\geqslant a$ 时, $F(x)=P\{S\}=1$;

当 $-a\leqslant x<\frac{\sqrt{2}}{2}a$ 时, $\arccos\frac{x}{a}\leqslant\frac{\pi}{4}+\theta\leqslant 2\pi-\arccos\frac{x}{a}$;

当 $\frac{\sqrt{2}}{2}a<x\leqslant a$ 时, $\arccos\frac{x}{a}\leqslant\frac{\pi}{4}+\theta\leqslant 2\pi-\arccos\frac{x}{a}$ 或 $2\pi+\arccos\frac{x}{a}\leqslant\frac{\pi}{4}+\theta\leqslant 4\pi-\arccos\frac{x}{a}$.

于是 $X=X\left(\frac{\pi}{4}\right)=a\cos\left(\frac{\pi}{4}+\Theta\right)$ 的概率密度为

$$f_1\left(x;\frac{\pi}{4}\right)=\frac{\mathrm{d}}{\mathrm{d}x}F(x)=\begin{cases}\frac{1}{\pi}\frac{1}{\sqrt{a^2-x^2}}, & |x|<a\\ 0, & |x|\geqslant a\end{cases}.$$

例 10.2. 设随机过程 $Y(t)=X\sin\omega t$, 式中 ω 是常数, X 是服从 $N(\mu,\sigma^2)$ 分布的随机变量, 求 $Y(t)$ 的一维概率密度.

解:

当 $\sin\omega t\neq 0$ 时, $Y(t)$ 是正态分布随机变量 X 的线性函数, 所以 $Y(t)$ 服从正态分布, 且 $Y(t)\sim N\left(\sin\omega t\mu,\sin^2\omega t\sigma^2\right)$, 故 $Y(t)$ 的一维概率密度为

$$f(y;t)=\frac{1}{\sqrt{2\pi}\sigma|\sin\omega t|}\exp\left\{-\frac{(y-\mu\sin\omega t)^2}{2\sigma^2\sin^2\omega t}\right\}(-\infty<y<+\infty).$$

当$\boldsymbol{\sin t=0, Y(t)\equiv 0}$.

10.3 数字特征

- 均值函数(假设右端均存在)

$$m(t)\overset{def}{=}E(X(t))$$

- 方差函数

$$D(t)\overset{def}{=}E\{(X(t)-m(t))^2\}$$

由此定义标准差函数:

$$\sigma(t)\overset{def}{=}\sqrt{D(t)}$$

- 自相关函数

$$R_X(t_1,t_2)\overset{def}{=}E[X(t_1)X(t_2)]$$

- 互相关函数

$$R_{XY}(t_1,t_2)\overset{def}{=}E[X(t_1)Y(t_2)]$$

- 自协方差函数

$$C_X(t_1,t_2)\overset{def}{=}Cov[X(t_1)X(t_2)]$$

- 互协方差函数

$$\begin{aligned}C_{XY}(t_1,t_2)&\overset{def}{=}Cov[X(t_1)Y(t_2)]\\&=E\{[X(t_1)-m_X(t1)][Y(t_2)-m_Y(t_2)]\}\end{aligned}$$

例 10.3. 设随机过程 $Y(t)=\mathrm{e}^{-tX}(t\in(-\infty,+\infty))$, 其中 X 是在 $(0,1)$ 上服从均匀分布的随机变量, 求 $Y(t)$ 的均值函数、自相关函数.

解:

由题设条件, 知 X 的概率密度为

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

$Y(t)$ 的均值函数为 $\mu_Y(t)=E[Y(t)]=E\left[\mathrm{e}^{-tX}\right]=\int_{-\infty}^{+\infty}\mathrm{e}^{-tx}f(x)\mathrm{d}x$

$$=\int_0^1\mathrm{e}^{-tx}\,\mathrm{d}x=\frac{1-\mathrm{e}^{-t}}{t}$$

$Y(t)$ 的自相关函数为 $R_Y(t_1,t_2)=E\left[Y(t_1)Y(t_2)\right]=E\left[\mathrm{e}^{-t_1x}\mathrm{e}^{-t_2x}\right]$

$$=\int_{-\infty}^{+\infty}\mathrm{e}^{-(t_1+t_2)x}f(x)\mathrm{d}x$$

$$=\int_0^1\mathrm{e}^{-(t_1+t_2)x}\,\mathrm{d}x=\frac{1-\mathrm{e}^{-(t_1+t_2)}}{t_1+t_2}$$

$t=0$时单独讨论.

10.4　不相关与相互独立

设$X_T=X_(t),t\in T,Y_T=Y(t),t\in T$是两个定义在同一个样本空间Ω上且具有同一个参数集$T$的随机过程.

定义 10.3 (不相关)**.** 若对任意的$t_1,t_2\in T$，有

$$C_{XY}(t_1,t_2)\equiv 0$$

则称X_T与Y_T是不相关的.

定义 10.4 (相互独立)**.** 若对任意n,m以及任意的$t_1,t_2,\cdots,t_n\in T,t_1',t_2',\cdots,t_m'\in T$, $(X(t_1),X(t_2),\cdots,X_(t_n)),(Y(t_1'),Y(t_2'),\cdots,Y_(t_m'))$相互独立, 则称则称$X_T$与$Y_T$是相互独立的.

10.5　随机过程分类

- 按照参数集和状态空间分类

  按照随机过程X_T的参数集T与状态空间S的可数与否, 将随机过程分为四类

  当T可列时, 也称X_T为离散参数随机过程或者随机序列, 有时称作时间序列. 当T是$\boldsymbol{R}$中有限或者无限区间时, 称X_T为连续参数随机过程.

- 按照过程性质特点分类

  - 二阶矩过程

    定义 10.5 (二阶矩过程). *设$X_T = X(t), t \in T$, 若对任意的$t \in T, E[X^2(t)]$都存在, 则称X_T为二阶矩过程.*

  - 正态过程

    定义 10.6 (正态过程). *若X_T的任何一个有限维分布都是多维正态分布, 亦即对任意的$n \geq 1$和任意n个不同的$t_1, t_2, \cdots, t_n \in T, n$维随机变量$(X_1(t), X_2(t), \cdots, X_n(t))$的概率密度为*

$$
\begin{aligned}
& f_{t_1,t_2,\cdots,t_n}(x_1, x_2, \cdots, x_n) \\
&= (2\pi)^{-\frac{n}{2}}|\boldsymbol{C}|^{-\frac{1}{2}} exp\{-\frac{1}{2}[(\boldsymbol{x}-\boldsymbol{m})^T \boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{m})]\}
\end{aligned}
$$

    其中

$$\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^T, \boldsymbol{m} = (m_X(t_1), m_X(t_2), \cdots, m_X(t_n))^T$$

$$\boldsymbol{C} = \begin{bmatrix} C_X(t_1,t_1) & C_X(t_1,t_2) & \cdots & C_X(t_1,t_n) \\ C_X(t_2,t_1) & C_X(t_2,t_2) & \cdots & C_X(t_2,t_n) \\ \vdots & \vdots & \ddots & \vdots \\ C_X(t_n,t_1) & C_X(t_n,t_2) & \cdots & C_X(t_n,t_n) \end{bmatrix}.$$

    为对称正定矩阵，则称X_T为正态过程.

  - 正交增量过程

定义 10.7 (正交增量过程). 设X_T为二阶矩过程, 若对任意的$0 \le t_1 < t_2 \le t_3 < t_4$,恒有

$$E[X(t_2) - X(t_1)][X(t_4) - X(t_3)] = 0$$

则称X_T为正交增量过程.

命题 10.8. 对于正交增量过程X_T, 如果规定$X(0) = 0$, 那么, 当$t < s$时, 有

$$E[X(s)X(t)] = E\{[X(t)-X(0)][X(s)-X(t)+X(t)]\} = E[X^2(t)] \overset{def}{=} g(t) = g(s \wedge t)$$

– 独立增量过程

定义 10.9 (独立增量过程). 若对任意有限个$0 \le t_1 < t_2 < \cdots < t_n$, 增量$X(t_2) - X(t_1), X(t_3) - X(t_2), \cdots, X(t_n) - X(t_{n-1})$相互独立，则称$X_T$为独立增量过程.

– 马尔可夫过程

定义 10.10 (马尔可夫过程). 参数集$T \subset [0, +\infty)$, 若对任意$n \ge 1$, 任意的n个时刻$0 \le t_1 < t_2 < \cdots < t_n$以及任意的$s > 0 (s + t_n \in T)$,恒有

$$P\{X(t_n + s) \le x | X(t_1) = x_1, X(t_2) = x_2, \cdots, X(t_n) = x_n\}$$

$$= P\{X(t_n + s) \le x | X(t_n) = x_n\}, x_1, x_2, \cdots, x_n \in \mathbb{R}$$

则称X_T为马尔可夫过程，简称马氏过程.

10.6　泊松过程

定义 10.11 (计数过程). 以$N(t)$表示在时间$(0, t]$内随机事件A发生的次数，则称随机过程$\{N(t), t \ge 0\}$为计数过程

命题 10.12.

- $N(t)$取值非负整数
- 若$0 \le s < t$, 则$N(s) \le N(t)$
- 若$0 \le s < t$, 则$N(t) - N(s)$表示在时间$(s, t]$内随机事件A发生的次数

定义 10.13 (泊松过程)**.** 设$N(t), t \geq 0$是计数过程, 若它满足条件:

- $N(0) = 0$
- 是独立增量过程
- 对任意的$s \geq 0, t > 0$,

  $$P\{N(t+s) - N(s) = k\} = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0, 1, 2, \cdots$$

  则称$N(t)$是强度为λ的泊松过程.

定义 10.14 (泊松过程)**.** 设$N(t), t \geq 0$是计数过程, 若它满足条件:

- $N(0) = 0$
- 是平稳独立增量过程
- $\lim_{h \to 0} \frac{P(N(h)=1)}{h} = \lambda;$
- $\lim_{h \to 0} \frac{P(N(h) \geq 2)}{h} = 0.$

则称$N(t)$是强度为λ的泊松过程.

定理 10.15. 上述两种定义等价.

定理 10.16. 计数过程是泊松过程的充要条件是$\{T_n, n \geq 1\}$是独立且参数同为λ的指数分布.

定理证明要点:

$$\{S_n > t\} = \{N(t) < n\}, S_n - S_{n-1} = T_n$$

命题 10.17. *content...*

例 10.4. 设到达火车站的客流符合参数为λ的泊松流$N(t)$, 且已知火车在t时刻到达, 求在$[0.t]$时刻到达车站的乘客等候时间总和的期望.

解:

设第 i 个顾客到达火车站的时刻为 S_i, 则 $[0, t]$ 到达车站的顾客等待时间总和为

$$S(t) = \sum_{i=1}^{N(t)} (t - S_i).$$

因

$$\begin{aligned}E(S(t)\mid N(t)=n)&=E\left\{\sum_{i=1}^{N(t)}(t-S_i)\mid N(t)=n\right\}\\&=E\left[\sum_{i=1}^{n}(t-S_i)\mid N(t)=n\right]=nt-E\left[\sum_{i=1}^{n}S_i\mid N(t)=n\right],\end{aligned}$$

仍记 $\{Y_i,1\le i\le n\}$ 为 $[0,t]$ 上独立同均匀分布的 $r.v.$.（随机变量）则

$$\begin{aligned}E\left\{\sum_{i=1}^{n}S_i\mid N(t)=n\right\}&=E\left(\sum_{i=1}^{n}Y_{(i)}\right)\\&=E\left(\sum_{i=1}^{n}Y_i\right)=\frac{nt}{2}\end{aligned}$$

故

$$E\left\{\sum_{i=1}^{n}(t-S_i)\mid N(t)=n\right\}=\frac{nt}{2}$$

所以

$$\begin{aligned}E[S(t)]&=\sum_{n=0}^{\infty}\left(P\{N(t)=n\}E\left[\sum_{i=1}^{N(t)}(t-S_i)\mid N(t)=n\right]\right)\\&=\sum_{n=0}^{\infty}P(N(t)=n)\cdot\frac{nt}{2}=\frac{t}{2}\cdot E(N(t))=\frac{\lambda}{2}t^2\end{aligned}$$

第十一章　平稳过程

11.1　严平稳过程

定义 11.1 (严平稳过程)**.** 设X_T是随机过程, 若对任意$n \geq 1$, 对任意的n个时刻$t_1, t_2, \cdots, t_n \in T$, 和任意的$\tau, (t_1+\tau, t_2+\tau, \cdots, t_n+\tau \in T)$, 都有

$$(X(t_1), X(t_2), \cdots, X(t_n)) \stackrel{d}{=} (X(t_1+\tau), X(t_2+\tau), \cdots, X(t_n+\tau))$$

例 11.1. （*Bernoulli* 序列）独立重复地进行某项试验, 每次试验成功的概率为 $p(0 < p < 1)$, 失败的概率为 $1-p$. 以 X_n 表示第 n 次试验成功的次数, 试验证 $\{X_n, n = 1, 2, 3, \cdots\}$ 是严平稳过程.

解:

$$X_n = \begin{cases} 1 & \text{第 } n \text{ 次试验成功} \\ 0 & \text{第 } n \text{ 次试验失则,} \end{cases}$$

$P\{X_n = k\} = p^k(1-p)^{1-k}(k = 0, 1)$, 且 $\{X_n,\ n = 1, 2, \cdots\}$ 是独立随机序列. 任取 m 个正整数 $i_1, i_2, \cdots, i_m$, 则 m 维分布律为

$$\begin{aligned} P\{X_{i_1} = k_1, X_{i_2} = k_2, \cdots, X_{i_m} = k_m\} &= \prod_{r=1} P\{X_{i_r} = k_r\} \\ &= \prod_{r=1}^{m} p^{k_r}(1-p)^{1-k_r}\,(k_r = 0, 1)\,. \end{aligned}$$

对任意正整数 l,必有

$$\begin{aligned} &P\{X_{i_1+l} = k_1, X_{i_2+l} = k_2, \cdots, X_{i_m+l} = k_m\} \\ &\quad = \prod_{r=1}^{m} P\{X_{i_r+l} = k_r\} = \prod_{r=1}^{m} p^{k_1}(1-p)^{1-k_r} \\ &\quad = P\{X_{i_1} = k_1, X_{i_2} = k_2, \cdots, X_{i_m} = k_m\}\,, \end{aligned}$$

故 *Bernoulli* 序列 $\{X_n, n=1,2,3,\cdots\}$ 是严平稳过程.

11.2　宽平稳过程

定义 11.2 (宽平稳过程)**.** 设X_T是二阶矩过程, 若它满足

- 对任意$t\in T, E[X(t)]$为常数
- 对任意的$t\in T, t+\tau\in T, E[X(t)X(t+\tau)]$不依赖于$t$

则称X_T是宽平稳过程. 此时

$$R_X(\tau)=E[X(t)X(t+\tau)]$$

注意: 宽平稳过程未必是严平稳过程, 严平稳过程未必是宽平稳过程.

命题 11.3. 若X_T是正态过程, 则其为严平稳过程的充要条件为其为宽平稳过程.

定义 11.4. 设$\{X(n), n\in \boldsymbol{Z}\}$是互不相关的随机变量序列, 且$E[X(n)]=0, D[X(n)]=\sigma^2$, 容易验证它是宽平稳序列, 故其被称为白噪声序列.

例 11.2. 设 $Z(t)=X\sin\omega t+Y\cos\omega t$, 其中 ω 是常数, X 与 Y 是相互独立的随机变量, 且 $X\sim N(0,1), Y\sim U[-\sqrt{3},\sqrt{3}]$, 试证: $Z(t)$ 是广义平稳过程.

证明:

由题设条件得$EX=0, DX=1, EX^2=1; EY=0, EY^2=1, DY=1, E[XY]=EX\cdot EY=0$;

$$\begin{aligned}
E[Z(t)] &= \sin\omega t\cdot EX+\cos\omega t\cdot EY=0\\
E[Z(t)Z(t+\tau)] &= \sin\omega t\cdot\sin\omega(t+\tau)\cdot EX^2+[\sin\omega t\cdot\cos\omega(t+\tau)\\
&\quad +\cos\omega t\cdot\sin\omega(t+\tau)]E[XY]+\cos\omega t\cdot\cos\omega(t+\tau)\cdot EY^2\\
&= \sin\omega t\cdot\sin\omega(t+\tau)+\cos\omega t\cdot\cos\omega(t+\tau)=\cos\omega\tau\\
E\left[Z^2(t)\right] &= 1
\end{aligned}$$

于是 $Z(t)$ 是广义平稳过程.