No.

Date

概率与统计

随机变量的常见分布·记忆

一维常见分布　　符号　　分布　　$E(X)$　$D(X)$

1、离散型

① 0-1分布（两点分布）　$X\sim B(1,p)$　$P(X=k)=p^k q^{1-k}\ (k=0,1)$　p　pq

② 二项分布　$X\sim B(n,p)$　$P(X=k)=C_n^k p^k q^{n-k}\ (k=0,1,\cdots,n)$　np　npq

用于有放回抽样

P先上升后下降，当$[(n+1)p]$为整数时，在$k=(n+1)p$，$(n+1)p-1$时最大；

$[(n+1)p]$不为整数时，在k为$[(n+1)p]$的整数部分时最大

③ 泊松分布　$X\sim P(\lambda)$　$P(X=k)=\frac{\lambda^k e^{-\lambda}}{k!}$，$\lambda>0\ (k=0,1,\cdots)$　λ　λ

若$X\sim P(\lambda_1)$，$Y\sim P(\lambda_2)$相互独立，则$X+Y\sim P(\lambda_1+\lambda_2)$

④ 几何分布　$X\sim G(p)$　$P(X=k)=q^{k-1}p\ (k=1,2,\cdots)$　$\frac{1}{p}$　$\frac{q}{p^2}$

小性质：无记忆性，即：$X\sim G(p)\Rightarrow P\{X>n+m\mid X>n\}=P\{X>m\}$

⑤ 超几何分布　$X\sim H(n,N,M)$　$P(X=k)=\frac{C_M^k C_{N-M}^{n-k}}{C_N^n}\ (k=0,1,\cdots,l)$　$\frac{nM}{N}$　$\times$

用于不放回抽样　抽样数，产品总数，次品数　$l=\min\{n,M\}$

定理：若$\lim\limits_{N\to\infty}\frac{M}{N}=p$，$0<p<1$，则$\lim\limits_{N\to\infty}\frac{C_M^k C_{N-M}^{n-k}}{C_N^n}=C_n^k p^k(1-p)^{n-k}\ (k=0,1,\cdots,n)$

2 连续型：

① 均匀分布　$X\sim U[a,b]$，$f(x)=\begin{cases}\frac{1}{b-a}, & a\le x\le b\\ 0 & 其他\end{cases}$　$F(x)=\begin{cases}0 & x<a\\ \frac{x-a}{b-a} & a\le x<b\\ 1 & b\le x\end{cases}$　$\frac{a+b}{2}$　$\frac{(b-a)^2}{12}$

② 指数分布　$X\sim E(\lambda)$ $(\lambda>0)$　$f(x)=\begin{cases}\lambda e^{-\lambda x} & x>0\\ 0 & x\le 0\end{cases}$　$F(x)=\begin{cases}1-e^{-\lambda x} & x>0\\ 0 & x\le 0\end{cases}$　$\frac{1}{\lambda}$　$\frac{1}{\lambda^2}$

小性质：无记忆性，即$X\sim E(\lambda)\Rightarrow P\{X>n+m\mid X>n\}=P\{X>m\}$

③ 正态分布　$X\sim N(\mu,\sigma^2)$，$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$　$x\in R$　μ　σ^2

$\Rightarrow\int_{-\infty}^{+\infty}e^{-\frac{t^2}{2}}dt=\sqrt{2\pi}$

标准正态分布 $X\sim N(0,1)$　$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$　$x\in R$，$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt$，$x\in R$

$\Rightarrow\varphi(-x)=\varphi(x)$，$\Phi(-x)+\Phi(x)=1$

性质：对于$X\sim N(\mu,\sigma^2)$，有$F_X(x)=\Phi\left(\frac{x-\mu}{\sigma}\right)$

$\Rightarrow P(x_1<X\le x_2)=F_X(x_2)-F_X(x_1)=\Phi\left(\frac{x_2-\mu}{\sigma}\right)-\Phi\left(\frac{x_1-\mu}{\sigma}\right)$

常用：对$X\sim N(0,1)$有$E(X)=0$，$D(X)=1\Rightarrow E(X^2)=1$，$E(|X|)=\sqrt{\frac{2}{\pi}}$，$D(|X|)=1-\frac{2}{\pi}$

No.

Date

性质：正态随机变量的线性函数仍为正态随机变量

若 $X\sim N(\mu,\sigma^2)$，则 $Y=aX+b\sim N(a\mu+b,a^2\sigma^2)$

二维正态分布：$(X,Y)\sim N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$，$\Rightarrow X\sim N(\mu_1,\sigma_1^2)$ $Y\sim N(\mu_2,\sigma_2^2)$

特别地，当 $\rho=0$ 时，$(X,Y)\sim N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;0)$

此时 $f(x,y)=f_X(x)f_Y(y)$，X、Y 独立，即

$\rho=0 \Leftrightarrow$ X、Y独立 $\Leftrightarrow \rho_{XY}=0$，即X、Y不相关

性质：多个相互独立的正态变量的线性组合仍是正态变量

$X_i\sim N(\mu_i,\sigma_i^2)$ $(i=1,2,\cdots,n)$ 且 X_i 独立，则

$\sum_{i=1}^{n}X_i=X_1+X_2+\cdots+X_n\sim N(\sum_{i=1}^{n}\mu_i,\sum_{i=1}^{n}\sigma_i^2)$

随机变量的和的分布

① $X_1,X_2,\cdots X_n$ 相互独立且 $X_i\sim P(\lambda_i)$ $i=1,2,\cdots,n$，则

$X_1+X_2+\cdots+X_n\sim P(\lambda_1+\lambda_2+\cdots+\lambda_n)$

② $X_1,X_2\cdots X_n$ 相互独立且 $X_i\sim B(1,p)$，则

$X_1+X_2+\cdots+X_n\sim B(n,p)$

③ $X_1,X_2,\cdots,X_k$ 相互独立且 $X_i\sim B(n_i,p)$，$i=1,2,\cdots,k$，则

$X_1+X_2+\cdots+X_k\sim B(n_1+n_2+\cdots+n_k,p)$

No.

Date

常用表述·记忆

1.概率的公理化定义

称实数 $P(A)$ 为事件A的概率，当

① 规范性 $P(\Omega)=1$ ② $P(A)\geq 0$

③ 可列可加性：对互斥事件 $A_1, A_2, \cdots A_n, \cdots$，有

$$P(A_1+A_2+\cdots+A_n+\cdots)=P(A_1)+P(A_2)+\cdots+P(A_n)+\cdots$$

2.分布函数

某一函数 $F(x)$ 满足：① $0\leq F(x)\leq 1$ $x\in R$ ② $F(x)$ 单调非减

③ $F(-\infty)=0$, $F(+\infty)=1$ ④ $F(x)$ 右连续

$\Longleftrightarrow$ 存在随机变量X以 $F(x)$ 为其分布函数

3.分布列

某一数列 $\{p_k\}$ 满足：① $p_k\geq 0$ $k=1,2,\cdots$ ② $\sum_k p_k=1$

$\Longleftrightarrow$ 存在离散型一X以 $\{p_k\}$ 为其分布列

4.概率密度

某一函数 $f(x)$ 满足：① $f(x)\geq 0$ ② $\int_{-\infty}^{+\infty} f(x)dx=1$

$\Longleftrightarrow$ 存在X以 $f(x)$ 为概率密度

5.二维分布函数

某一二元函数 $F(X,Y)$ 满足 ① $0\leq F(x,y)\leq 1$, $\forall x,y\in R$ ② $F(x,y)$ 对 x,y 都是单调非减

③ $F(-\infty,y)=F(x,-\infty)=F(-\infty,-\infty)=0$, $F(+\infty,+\infty)=1$

④ $F(x,y)$ 关于 x、y 都是右连续的

⑤ $F(x_2,y_2)-F(x_2,y_1)-F(x_1,y_2)+F(x_1,y_1)\geq 0$, $\forall x_1\leq x_2, y_1\leq y_2$

$\Longleftrightarrow$ 存在 (X,Y) 以 $F(X,Y)$ 为其分布函数

6.切比雪夫不等式

任意X，若 $D(X)$ 存在，则 $\forall \varepsilon>0$，有 $P\{|X-E(X)|\geq\varepsilon\}\leq\frac{D(X)}{\varepsilon^2}$

$$P\{|X-E(X)|<\varepsilon\}\geq 1-\frac{D(X)}{\varepsilon^2}$$

7.伯努利大数定律

设在n重伯努利试验中成功次数记为 Y_n，每次试验的成功概率为 p，$(0<p<1)$

$\forall\varepsilon>0$，有 $\lim\limits_{n\to\infty}P\{|\frac{Y_n}{n}-p|\geq\varepsilon\}=0$ $\lim\limits_{n\to\infty}P\{|\frac{Y_n}{n}-p|<\varepsilon\}=1$

KOKUYO

8、独立同分布的中心极限定理

若随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 独立同分布，且 $E(X_i)=\mu$, $D(X_i)=\sigma^2>0$ 有限，则

$\forall x \in R$，有 $\lim\limits_{n\to\infty} P\left\{\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \le x\right\} = \Phi(x)$

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

9、棣莫弗一拉普拉斯定理

设在 n 重伯努利试验中，$\cdots Y_n \cdots$，p，$q=1-p$，则

$\forall x \in R$，有 $\lim\limits_{n\to\infty} P\left\{\frac{Y_n - np}{\sqrt{npq}} \le x\right\} = \Phi(x)$

第1章 随机事件与概率

一、随机事件

1. 称满足以下三个条件的试验为随机试验，简称试验，用E表示

   ① 试验可在相同条件下重复进行

   ② 试验的所有可能结果不止一个，且所有可能结果是事先已知的

   ③ 每次试验总是恰好出现这些可能结果之一，但究竟是哪一个结果无法提前预言

2. 基本事件（样本点）：随机试验的每一个可能结果，用 e 表示

3. 样本空间：样本点的全体，记为S

4. 事件：基本事件的集合。

   当事件A当中的某一基本事件出现时称事件A发生；否则称事件A不发生

5. S是必然事件，空集 ϕ 是不可能事件

二、事件的关系与运算

1. 对立事件一定是互斥的，互斥事件不一定是对立的

2. 若 $A \subset B$，则有 $AB=A$，$A \cup B=B$

3. 若A与B互斥，则记 $A \cup B$ 为 $A+B$

4. $A-B=A\overline{B}=A-AB=A \cup B-B$

5. 多画图

6. 事件的运算性质：

   (i) 交换律：$A \cup B=B \cup A$，$AB=BA$

   (ii) 结合律：$(A \cup B) \cup C=A \cup (B \cup C)$；$(AB)C=A(BC)$

   (iii) 分配律：$(A \cup B) \cap C=AC \cup BC$；$(A \cap B) \cup C=(A \cup C) \cap (B \cup C)$

   (iv) 德·摩根律（对偶原理）

   $\overline{\bigcup_{i=1}^{n} A_i}=\bigcap_{i=1}^{n}\overline{A_i}=\overline{A_1}\,\overline{A_2}\cdots\overline{A_n}$ 事件的和的对立等于对立事件的积

   $\overline{\bigcap_{i=1}^{n} A_i}=\bigcup_{i=1}^{n}\overline{A_i}=\overline{A_1}\cup\overline{A_2}\cdots\cup\overline{A_n}$ 事件的积的对立等于对立事件的和

No.

Date

三、概率

1.概率的概念与计算

①古典概型：　其样本空间S只含有限个基本事件，且每个基本事件是等可能的

$$P(A)=\frac{A\text{包含的基本事件的个数}}{\text{基本事件的总数}}$$

②几何概型：　无限性、等可能性

$$P(A)=\frac{\text{子区域}A\text{的度量}}{\text{样本空间}S\text{的度量}}$$　其中度量可能为长度、面积…

③统计概率：　概率是频率的极限

2.概率的性质

三个公理
- (i) 对任一事件A，有 $0\le P(A)$　　(ii) 规范性：$P(S)=1$
- (iii) 可列可加性：若 $A_1,A_2,\cdots,A_n,\cdots$ 互斥，则有 $P(A_1+A_2+\cdots+A_n+\cdots)=P(A_1)+P(A_2)+\cdots+P(A_n)+\cdots$

(iv) $P(\bar{A})+P(A)=1$　　(v) $P(\phi)=0$

(vi) 对任意两事件A、B，$P(A\bar{B})=P(A-B)=P(A)-P(AB)$　$\Rightarrow P(A)=P(AB)+P(A\bar{B})$

若 $A\subset B$，则 $P(A)\le P(B)$ 且 $P(B-A)=P(B)-P(A)$

(vii) 一般概率加法公式：$P(A\cup B)=P(A)+P(B)-P(AB)$

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

$+(-1)^{n-1}P(A_1A_2\cdots A_n)$

3.排列与组合

排列 $A_n^k=n(n-1)\cdots(n-k+1)=\frac{n!}{(n-k)!}$

当 $k=n$ 时称为全排列 $A_n^n=n(n-1)\cdots2\cdot1=n!$

组合 $C_n^k=\binom{n}{k}=\frac{A_n^k}{k!}=\frac{n!}{k!(n-k)!}=C_n^{n-k}$　　$A_n^k=C_n^k\cdot k!$

4.常用概率不等式

事件A、B，有 $P(AB)\le P(B)$、$P(A)\le P(A\cup B)$

第2章 条件概率与独立性

一、条件概率

1. 定义：设 A、B 为两个事件，满足 $P(B)>0$，称比值 $\frac{P(AB)}{P(B)}$ 为事件 A 在事件 B 已发生的条件下的条件概率，记作 $P(A|B)=\frac{P(AB)}{P(B)}$

条件概率也是概率，满足前述概率的性质

(i) $P(A|B)\geq 0$　　(ii) $P(S|B)=1$

(iii) 若 $A_1, A_2, \cdots$ 互不相容，则 $P((A_1+A_2+\cdots)|B)=P(A_1|B)+P(A_2|B)+\cdots$

(iv) $P(\bar{A}|B)+P(A|B)=1$　　(v) $P(\phi|B)=0$

(vi) $P((A_1-A_2)|B)=P(A_1|B)-P(A_1A_2|B)$　　当 $A_1\supset A_2$ 时，$P(A_1|B)\geq P(A_2|B)$

(vii) $P(A_1\cup A_2|B)=P(A_1|B)+P(A_2|B)-P(A_1A_2|B)$

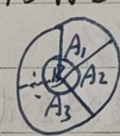
2. 乘法定理：$P(AB)=P(A|B)P(B)=P(B|A)P(A)$

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

二、全概率公式

设事件 $A_1, A_2, \cdots, A_n$ 互不相容，且 $P(A_i)>0$，$(i=1,2,\cdots,n)$ 若对事件 B 有 $B\subset A_1+A_2+\cdots+A_n$

则　$P(B)=\sum_{i=1}^{n}P(A_i)P(B|A_i)$



三、贝叶斯公式

引出：在试验中 B 发生了，问引起 B 发生的原因是 A_i 的概率有多大？哪一个原因发生的可能性最大？

设事件 $A_1, A_2, \cdots, A_n$ 互不相容，且 $P(A_i)>0$　$(i=1,2,\cdots,n)$，若对事件 B 有 $B\subset A_1+A_2+\cdots+A_n$ 且 $P(B)>0$

则　$P(A_i|B)=\frac{P(A_iB)}{P(B)}=\frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n}P(A_j)P(B|A_j)}$　$(i=1,2,\cdots,n)$

$P(A_i)$ 称先验概率，一般是在试验前就已知的，$P(A_i|B)$ 为后验概率

贝叶斯公式就是根据先验概率求后验概率的公式

No.
Date

四、事件的独立性

1. 定义：设A、B为两个事件，称A与B是相互独立的，当 $P(AB)=P(A)P(B)$　　若 $P(A)>0$，等价于 $P(B|A)=P(B)$

2. 定理：若A与B相互独立，则 A与$\bar{B}$、$\bar{A}$与B、$\bar{A}$与$\bar{B}$ 也分别相互独立

3. 性质：若A与B中有一个的概率为0（或为1），则A、B相互独立
　　若 $0<P(A)、P(B)<1$，则"A、B相互独立"与"A、B互斥"不能同时成立

注：在实际问题中，对事件A、B，常常根据其实际意义来看彼此是否有影响，从而判断它们是否独立，若独立，则可使用 $P(AB)=P(A)P(B)$

4. 多个事件的独立性
　定义：设A、B、C为三个事件，称A、B、C两两独立，当
$$P(AB)=P(A)P(B)\quad P(BC)=P(B)P(C)\quad P(AC)=P(A)P(C)$$
　；称A、B、C相互独立，当　上三式满足且 $P(ABC)=P(A)P(B)P(C)$　相互独立 $\Rightarrow$ 两两独立
　定理：若事件 $A_1, A_2, \cdots A_n$ 相互独立，则将其中任意个事件换成其对立事件后，这些事件仍相互独立

五、重复独立试验

1. n次重复独立试验：进行n次试验，在每次试验中，任一事件出现的概率与其他各次结果无关
2. 伯努利试验：试验的结果只有两个 A 和 $\bar{A}$
3. n重伯努利试验：伯努利试验的n次重复独立试验.
4. 二项概率公式：在n重伯努利试验中，设每次试验的成功概率为 $p\ (0<p<1)$，则成功恰好发生k次的概率为 $P_n(k)=C_n^k p^k q^{n-k}$
5. 二项概率的泊松逼近：
　　当试验次数很多，成功概率很小时，令 $\lambda=np$，有 $P_n(k)\approx\frac{\lambda^k}{k!}e^{-\lambda}$

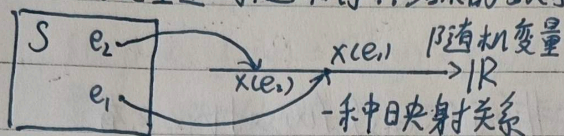
注：二项展开式：$1=(p+q)^n=\sum_{k=0}^{n} C_n^k p^k q^{n-k}=\sum_{k=0}^{n} P_n(k)$

第3章 随机变量及其分布

一、随机变量

1、概念：设E是随机试验，其样本空间是S，若对S中每一个基本事件e，都有唯一的实数值X(e)与之对应，则称 X=X(e)为随机变量

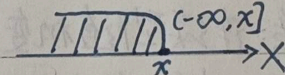
随机变量是对随机事件结果的数字化表示



2、随机变量的分布函数

定义：设X为一随机变量，称 $F(x)=P(X\le x)$ 为X的分布函数，其中x为任意实数

所有随机变量都有其分布函数



计算：$P(x_1<X\le x_2)=F(x_2)-F(x_1)$

$P(x_1<X<x_2)=F(x_2)-F(x_1)-P(X=x_1)$

$P(X=a)=P(X\le a)-P(X<a)=F(a)-F(a^-)$

$P(X>b)=1-P(X\le b)=1-F(b)$ ； $P(X\ge b)=1-F(b^-)$

性质：(i) $0\le F(x)\le 1$ (ii) $F(x)$是单调非减的，即 $\forall x_1<x_2$，有 $F(x_1)\le F(x_2)$

(iii) $F(-\infty)=\lim\limits_{x\to-\infty}F(x)=0$ $F(+\infty)=\lim\limits_{x\to+\infty}F(x)=1$

(iv) $F(x)$是右连续的，即 $F(x)=F(x^+)$

3、随机变量的分类
- 离散型：有限个/可列无穷多个 → 有分布列
- 非离散型
  - 连续型：所取的值连续地充满一个区间 → 有概率密度
  - ……

二、离散型随机变量

1、概念：只能取有限多个或可列无穷多个值的随机变量X 称离散型随机变量

设X所有可能取值 $x_1, x_2, \cdots, x_k, \cdots$，设事件"$X=x_k$"的概率为$p_k$，则

X的分布列 $P(X=x_k)=p_k \quad (k=1,2,\cdots)$

其表格形式

X	x_1	x_2	$\cdots$	x_k	$\cdots$
P	p_1	p_2	$\cdots$	p_k	$\cdots$

2、性质 (i) $p_k\ge 0$ (ii) $\sum p_k=1$ 作用：求分布列中缺少的数

3.分布列与分布函数　$F(x)=P(X\le x)=\sum_{x_i\le x}P(X=x_i)$

已知分布列求分布函数：用X的取值来划分实数轴的区间，并注意F(x)的右连续性

已知分布函数求分布列：利用F(x)在$x=x_i$处的跳跃值，$P(X=x_i)=F(x_i)-F(x_i^-)$

三、连续型随机变量

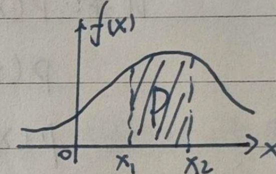
1.定义：设F(x)为X的分布函数，若存在一个非负可积函数f(x)，对$\forall x\in R$，有

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

则称X为连续型～，称f(x)为X的概率密度

2.性质：连续型随机变量的分布函数F(x)一定是连续的　但不一定是可导的

对f(x)的连续点，有　$F'(x)=f(x)$



3.性质：(i) $f(x)\ge 0$　　(ii) $\int_{-\infty}^{+\infty}f(x)dx=F(+\infty)=1$

(iii) $P(x_1<X\le x_2)=F(x_2)-F(x_1)=\int_{x_1}^{x_2}f(x)dx$

(iv) X为连续型，则$\forall x\in R$，$P(X=x)=0$

从而 $P(x_1\le X\le x_2)=P(x_1\le X<x_2)=P(x_1<X\le x_2)=P(x_1<X<x_2)$

从而 事件A $P(A)=1\not\Rightarrow$ A是必然事件；$P(A)=0\not\Rightarrow$ A是不可能事件

四、随机变量函数的分布

设g(x)是定义在随机变量X的一切可能值x的集合上的函数。若随机变量Y随着X取x的值而取$y=g(x)$的值，则称随机变量Y为随机变量X的函数，记作$Y=g(X)$

问题：对$Y=g(X)$，如何根据已知的随机变量X的分布寻求随机变量Y的分布

此处的"分布"对离散型来说是分布列，对连续型来说是概率密度

一、X为离散型　逐点代入合并法

$y=g(x)$

P	p_1 ··· p_n
X	x_1 ··· x_n
Y	y_1 ··· y_n

$\xrightarrow[\text{概率直接相加}]{\text{合并相同的}y_i}$

Y	y_1　y_2 ··· y_n
P	p_1'　p_2' ··· p_n'

二、X为连续型

1、分布函数求导法

Step 1. 已知 $F_X(x)$，根据分布函数定义，求 $Y=g(X)$ 的分布函数

$$F_Y(y)=P(Y\le y)\xlongequal{代入}P(g(X)\le y)\xlongequal{反解出X}P(X\le \varphi(y))=F_X(\varphi(y))$$

Step 2. $f_Y(y)=F_Y'(y)=\varphi'(y)f_X(\varphi(y))$

要点：把Y的分布函数 $F_Y(y)$ 用X的分布函数来表示　不同随机变量的分布函数相互表示
⇒用概率P作桥梁

注：注意复合函数求导

上述步骤只是表示，使用时注意分段，或直接使用公式法

2、公式法

设连续型X有 $f_X(x)$，又设 $y=g(x)$ 在 (a,b) 上严格单调可微，其反函数为 $x=h(y)$，有

$$f_Y(y)=\begin{cases} f_X(h(y))|h'(y)| & A<y<B \\ 0 & 其他 \end{cases}$$

其中 $A=\min\{g(a),g(b)\}$，$B=\max\{g(a),g(b)\}$

若 $g(x)$ 在不重叠的区间 I_1、I_2、……上逐段严格单调可微，其反函数分别为 $h_1(y)$、$h_2(y)$……有

$$f_Y(y)=f_X(h_1(y))|h_1'(y)|+f_X(h_2(y))|h_2'(y)|+\cdots$$

注：若 $f_X(x)$ 的表达式是分段的，则代入 $f_X(h(y))$ 时应代入对应的段

No.

Date

第4章 多维随机变量及其分布

一、多维随机变量

1.概念：若 $X_1(e), X_2(e), \cdots, X_n(e)$ 是定义在同一个样本空间 S 上的 n 个随机变量，$e \in S$，
（随机变量）（样本点）
则由它们构成的一个 n 维向量 $(X_1(e), X_2(e) \cdots X_n(e))$ 称为 n 维随机变量

例：研究某班学生的发育情况，对每人测其身高 H、体重 W

$S=\{e\}=\{$某班全部学生$\}$ $\quad e \to H(e),\ W(e)$

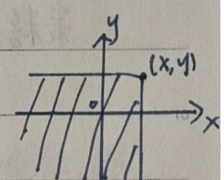
则 $(H(e), W(e))$ 为二维随机变量

2、二维随机变量的分布函数

定义：设 (X, Y) 为二维随机变量，称 $F(x,y)=P(X\le x, Y\le y)\quad (x,y)\in R^2$

为 (X,Y) 的分布函数或称为 X 和 Y 的联合分布函数

其中 $P(X\le x, Y\le y)$ 是事件 $\{X\le x\}$、$\{Y\le y\}$ 同时成立的概率



"联合"即把 (X,Y) 看作一个整体来研究

计算：$P(x_1<X\le x_2, y_1<Y\le y_2)=F(x_2,y_2)-F(x_2,y_1)-F(x_1,y_2)+F(x_1,y_1)$

性质：(i) 有界性：$\forall x,y\in R,\quad 0\le F(x,y)\le 1$

(ii) $\forall y,\ x_1<x_2$ 有 $F(x_1,y)\le F(x_2,y)$

$\forall x,\ y_1<y_2$ 有 $F(x,y_1)\le F(x,y_2)$

即 $F(x,y)$ 对每个自变量都是单调不减的

(iii) $\forall x,y$，有 $F(-\infty,y)=F(x,-\infty)=F(-\infty,-\infty)=0$

$F(+\infty,+\infty)=1$ ← 无穷矩形域扩展至整个平面

注：$F(+\infty,y)$ 和 $F(x,+\infty)$ 无法确定

(iv) $F(x,y)$ 对每个自变量都是右连续的，即

$F(x,y)=F(x^+,y)\qquad F(x,y)=F(x,y^+)$

3、二维随机变量的边缘分布函数

$F_X(x)=F(x,+\infty)\qquad F_Y(y)=F(+\infty,y)$

"边缘"即把 (X,Y) 中的 X 和 Y 分开来单独研究

No.

Date

二、二维离散型随机变量

1.概念：(X,Y) 的所有可能取值是有限对或可列无穷多对 $(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n),\cdots$

2、联合分布列：$p_{ij}=P(X=x_i,Y=y_j)\quad(i,j=1,2,\cdots)$

3.性质：(i) $p_{ij}\geq 0\quad(i,j=1,2,\cdots)$

(ii) $P(X=x_i)=\sum_{j=1}^{\infty}p_{ij}\triangleq p_{i\cdot}$

$P(Y=y_j)=\sum_{i=1}^{\infty}p_{ij}\triangleq p_{\cdot j}$

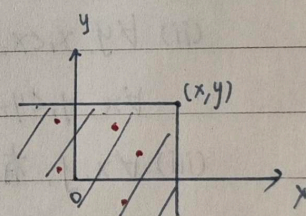
称 $p_{i\cdot}$，$p_{\cdot j}$ 为 (X,Y) 的边缘分布列

(iii) $\sum_i p_{i\cdot}=\sum_j p_{\cdot j}=\sum_j\sum_i p_{ij}=1$

表格形式

X \ Y	y_1	y_2	$\cdots$	y_j	$\cdots$	$p_{i\cdot}$ ← X的边缘分布
x_1	p_{11}	p_{12}	$\cdots$	p_{1j}	$\cdots$	$p_{1\cdot}$
x_2	p_{21}	p_{22}	$\cdots$	p_{2j}	$\cdots$	$p_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
x_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$	$p_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
Y的边缘分布→ $p_{\cdot j}$	$p_{\cdot 1}$	$p_{\cdot 2}$	$\cdots$	$p_{\cdot j}$	$\cdots$	1

4.分布列与分布函数



$$F(x,y)=P(X\leq x,Y\leq y)=\sum_{x_i\leq x}\sum_{y_j\leq y}p_{ij}$$

三、二维连续型随机变量

1.定义：设 (X,Y) 的分布函数为 $F(x,y)$，若存在非负可积函数 $f(x,y)$，对 $\forall\, x,y\in R$，有

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

则称 (X,Y) 为 ~，称 $f(x,y)$ 为 (X,Y) 的概率密度，或称 X 与 Y 的联合概率密度

2.意义：$f(x,y)$ 相当于质量的面密度，$F(x,y)$ 相当于无穷矩形域上的质量

3.性质：对 $f(x,y)$ 的连续点 (x,y)，有 $\dfrac{\partial^2F(x,y)}{\partial x\partial y}=f(x,y)$

4.性质：(i) $f(x,y)\geq 0$ (ii) $F(+\infty,+\infty)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)\,dx\,dy=1$

(iii) 设 G 是 xOy 面上的一个区域，则点 (X,Y) 落在 G 内的概率

$$P((X,Y)\in G)=\iint_G f(x,y)\,dx\,dy$$

几何意义：$P((X,Y)\in G)$ 在数值上等于以曲面 $z=f(x,y)$ 为顶，以平面区域 G 为底的曲顶柱体的体积

5、边缘概率密度

$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy$　　$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)dx$　　☆ 注意x与y的界限的分别处理

称$f_X(x)$、$f_Y(y)$为(X,Y)的边缘概率密度

总结：

$$f(x,y)\xrightarrow{\text{二重积分}}F(x,y)\xrightarrow{\text{代入}+\infty}F_X(x)、F_Y(y)\xrightarrow{\text{求导}}f_X(x)、f_Y(y)$$

$$f(x,y)\xrightarrow{\text{对另一变量作一重积分}}f_X(x)、f_Y(y)$$

联合分布 $\underset{\times}{\overset{\text{唯一确定}}{\rightleftarrows}}$ 边缘分布

$F(x,y)$、P_{ij}、$f(x,y)$　　$F_X(x)$　$P_{i\cdot}$　$f_X(x)$；$F_Y(y)$　$P_{\cdot j}$　$f_Y(y)$

四、随机变量的独立性

定义：两个随机变量X、Y独立，当且仅当X、Y的联合分布等于边缘分布的乘积，即

$F(x,y)=F_X(x)F_Y(y)$，$\forall x,y\in R$

or　$f(x,y)=f_X(x)f_Y(y)$，$\forall x,y\in R$

or　$P_{ij}=p_{i\cdot}\,p_{\cdot j}$　　注：若$i=1,2,\cdots,n$，$j=1,2,\cdots,m$，则该式要求$m\times n$个等式同时成立

定理：设随机变量X、Y独立，设$g(x)$是x的一元函数，$h(y)$是y的一元函数，则随机变量$g(X)$与$h(Y)$独立

五、推广至n维随机变量

设$(X_1,X_2,\cdots,X_n)$为n维随机变量，$x_1,x_2,\cdots,x_n$为n个任意实数

(1) 分布函数：$F(x_1,x_2,\cdots,x_n)=P(X_1\le x_1,X_2\le x_2,\cdots,X_n\le x_n)$

关于X_i的边缘分布函数 $F_{X_i}(x_i)=F(+\infty,+\infty,\cdots,x_i,\cdots,+\infty)$

(2) 概率密度：$F(x_1,x_2,\cdots,x_n)=\int_{-\infty}^{x_n}\int_{-\infty}^{x_{n-1}}\cdots\int_{-\infty}^{x_1}f(t_1,t_2,\cdots,t_n)dt_1dt_2\cdots dt_n$

关于X_i的边缘概率密度 $f_{X_i}(x_i)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}f(x_1,x_2,\cdots,x_n)dx_1dx_2\cdots dx_{i-1}dx_{i+1}\cdots dx_n$

(3) n个随机变量的独立性

$X_1,X_2,\cdots,X_n$是独立的 $\Longleftrightarrow F(x_1,x_2,\cdots,x_n)=F_{X_1}(x_1)F_{X_2}(x_2)\cdots F_{X_n}(x_n)$

对于连续型随机变量 $\Longleftrightarrow f(x_1,x_2,\cdots,x_n)=f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_n}(x_n)$

No.

Date

六、条件分布

两个事件的条件概率 $\xrightarrow{推广}$ 两个随机变量的条件分布

定义：离散型 (X,Y) $P(X=x_i|Y=y_j)=\frac{P(X=x_i,Y=y_j)}{P(Y=y_j)}=\frac{p_{ij}}{p_{\cdot j}}$ 称X在条件 $Y=y_j$ 下的条件分布列

j相对于固定的，$i=1,2,\cdots$

连续型 (X,Y) $f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}$，$f_Y(y)>0$ 为 $Y=y$ 条件下 X 的条件概率密度

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

注：$\underbrace{联合分布 \xrightarrow{唯一确定} 边缘分布}_{条件分布}$　　条件分布 $=\frac{联合分布}{边缘分布}$

七、二维随机变量的函数的分布

公式法：设 $(X,Y)\sim f(x,y)$，给定函数 $Z=g(X,Y)$，使 $z=g(x,y)$，若能反解出 $y=h(x,z)$，则

$$f_Z(z)=\int_{-\infty}^{+\infty}f(x,h(x,z))\left|\frac{\partial h(x,z)}{\partial z}\right|dx$$

或反解出 $x=\varphi(y,z)$，则 $f_Z(z)=\int_{-\infty}^{+\infty}f(\varphi(y,z),y)\left|\frac{\partial\varphi(y,z)}{\partial z}\right|dy$

Step 1. 用上述定理表示 $f_Z(z)$

Step 2. 注意可能有独立性，使 $f(x,h(x,z))=f_X(x)\cdot f_Y(h(x,z))$

Step 3. 注意 $f_Z(z)$ 只为 z 的函数，故只能对 z 分段讨论

注意对 z 讨论的目的是想找出使被积函数 $f(x,h(x,z))$ 非零的 z 的范围

2. 若无法反解，如瑞利分布 $Z=\sqrt{X^2+Y^2}$，则用基本方法：分布函数求导法

3. 对于离散型随机变量，基本方法：逐点代入合并法

4. 特殊分布——分布函数法

① $Z=\max\{X,Y\}$，且X、Y相互独立：$F_Z(z)=F_X(z)\cdot F_Y(z)$

且X、Y同分布：$F_Z(z)=[F(z)]^2$ $\Rightarrow$ $f_Z(z)=2F(z)f(z)$ 注意左右不同

$\Rightarrow$ 推广 $Z=\max\{X_1,\cdots X_n\}$，且 $X_1,\cdots X_n$ 相独立：$F_Z(z)=F_{X_1}(z)F_{X_2}(z)\cdots F_{X_n}(z)$

且 $X_1,\cdots X_n$ 同分布：$F_Z(z)=[F(z)]^n$

② $Z=\min\{X,Y\}$，且X，Y相互独立：$F_Z(z)=1-[1-F_X(z)][1-F_Y(z)]$

且X，Y同分布：$F_Z(z)=1-[1-F(z)]^2$

Campus 注：此处的 $F(z)$、$f(z)$ 指的是 X_1、$X_2\cdots X_n$ 的分布，并把自变量换成 z

第5章 随机变量的数字特征

一、数学期望 mathematical expectation

1.定义：离散型 X 其分布列 $P(X=x_i)=P_i\ (i=1,2,\cdots)$ 则X的数学期望

$$E(X)=\sum_{i=1}^{\infty}x_iP_i$$ 要求 $\sum_{i=1}^{\infty}|x_i|P_i<+\infty$

连续型 X 其概率密度 $f(x)$ 则X的数学期望

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

注：1.只需要在被积函数 $f(x)$ 非零的区间积分

2.$E(X)$的物理意义：质量密度为 $f(x)$ 的一维连续质点系的重心坐标

2.随机变量的函数的数学期望

→可以不求出随机变量的函数的分布，直接求其数学期望

定理：设 $Y=g(X)$，其中 $g(x)$ 连续，则

(i) 若 X 为离散型，其分布列 $P(X=x_i)=P_i\ (i=1,2,\cdots)$

$$E(Y)=E(g(X))=\sum_{i=1}^{\infty}g(x_i)P_i$$ ← 逐点代入 $g(x)$

(ii) 若 X 为连续型，其概率密度 $f_X(x)$

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

注：对于取有限个值的离散型，也可先用逐点代入合并法求 Y 的分布，再求 $E(Y)$

定理：设 $Z=g(X,Y)$，其中 $g(x,y)$ 连续则

(i) 若 (X,Y) 为二维离散型，其分布列 $P(X=x_i,Y=y_j)=P_{ij}\ \cdots$

$$E(Z)=E[g(X,Y)]=\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}g(x_i,y_j)P_{ij}$$ 同理

(ii) 若 (X,Y) 为二维连续型，其概率密度 $f(x,y)$

$$E(Z)=E[g(X,Y)]=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x,y)f(x,y)dxdy$$

⇒ 由 $(X,Y)\sim f(x,y)$ 求 X 与 Y 的数学期望：

$$E(X)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xf(x,y)dxdy \qquad E(Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}yf(x,y)dxdy$$

3.数学期望的性质

(i) 若C为常数，则 $E(C)=C$　　(ii) 若C为常数，则 $E(CX)=CE(X)$

(iii) $E(X_1+X_2+\cdots+X_n)=E(X_1)+E(X_2)+\cdots+E(X_n)$

(iv) 若 $X_1,X_2,\cdots,X_n$ 相互独立，则 $E(X_1X_2\cdots X_n)=E(X_1)E(X_2)\cdots E(X_n)$

注：1.线性性：$E(aX+bY)=aE(X)+bE(Y)$

2.X_1、X_2独立 $\xrightarrow[\times]{\checkmark}$ $E(X_1X_2)=E(X_1)E(X_2)$

No.

Date

二、方差

1.定义：设X为一随机变量，其方差 $D(X)=E\{[X-E(X)]^2\}$ 能反映X离开E(X)的平均偏离大小

标准差(均方差) $\sigma_X=\sqrt{D(X)}$

注：D(X)是一个数且 $D(X)\geq 0$

→ 离散型 $D(X)=\sum_{i=1}^{\infty}[x_i-E(X)]^2p_i$ 其中 $p_i=P(X=x_i)$ 为X的分布列

连续型 $D(X)=\int_{-\infty}^{+\infty}[x-E(X)]^2f(x)dx$ 其中 f(x)为X的概率密度

2.计算：$D(X)=E(X^2)-[E(X)]^2$

→ 离散型 $=\sum_{k=1}^{\infty}x_k^2p_k-[\sum_{k=1}^{\infty}x_kp_k]^2$

连续型 $=\int_{-\infty}^{+\infty}x^2f(x)dx-[\int_{-\infty}^{+\infty}xf(x)dx]^2$

3.方差的性质

(i) 若C为常数，则 $D(C)=0$ (ii) 若C为常数，则 $D(CX)=C^2D(X)$

(iii) $D(X\pm Y)=D(X)+D(Y)\pm 2E\{[X-E(X)][Y-E(Y)]\}=D(X)+D(Y)\pm 2[E(XY)-E(X)E(Y)]$

(iv) 若 $X_1,X_2,\cdots,X_n$ 相互独立，则 $D(X_1\pm X_2\pm\cdots\pm X_n)=D(X_1)+D(X_2)+\cdots+D(X_n)$

注：反之不一定成立

(v) $D(X)=0 \iff P(X=E(X))=1$

三、协方差和相关系数

研究二维随机变量(X,Y)时，协方差Cov(X,Y)和相关系数 ρ_{XY} 是刻画X与Y之间线性相关的程度的两个数字特征

1.协方差

定义式 $Cov(X,Y)=E\{[X-E(X)][Y-E(Y)]\}$

计算式 $Cov(X,Y)=E(XY)-E(X)E(Y)$

与方差的关系 $D(X\pm Y)=D(X)+D(Y)\pm 2Cov(X,Y)$ ⇒ $D(X+Y+Z)=D(X)+D(Y)+D(Z)+2Cov(X,Y)+2Cov(X,Z)+2Cov(Y,Z)$

性质：(i) $Cov(X,Y)=Cov(Y,X)$ (ii) $Cov(X,X)=D(X)$ (iii) 对常数C，$Cov(X,C)=0$

(iv) 对常数a,b $Cov(aX,bY)=ab\cdot Cov(X,Y)$

(v) $Cov(X_1+X_2,Y)=Cov(X_1,Y)+Cov(X_2,Y)$ ⇒ $Cov(X_1+X_2,Y_1+Y_2)$

注：协方差Cov(X,Y)一般是有量纲的

No.

Date

2、(线性)相关系数

定义：$\rho_{XY}=\frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$　　注：ρ_{XY} 无量纲

定理 i　有界性 $|\rho|\le 1$

ii　$|\rho|=1 \iff$ 存在常数 a,b. 使得 $P(Y=a+bX)=1$

即：当 $|\rho|=1$ 时，X与Y存在线性相关关系，这个事件的概率为1

特别地：当 $\rho=1$ 时，$b>0$，称X、Y为正相关

当 $\rho=-1$ 时，$b<0$，称X、Y为负相关

定义：若 $\rho_{XY}=0$，则称X与Y不相关

X、Y相互独立 $\underset{\times}{\overset{\checkmark}{\rightleftharpoons}}$ X、Y不相关
(X与Y无任何关系)　　(X与Y无线性相关关系)

X与Y之间重要关系的判断：

X与Y相互独立 $\iff F(x,y)=F_X(x)F_Y(y)\quad \forall x,y$
$p_{ij}=p_{i\cdot}\cdot p_{\cdot j}\quad \forall i,j$
$f(x,y)=f_X(x)f_Y(y)\quad \forall x,y$
} 用随机变量的分布来描述

X与Y不相关 $\iff \rho_{XY}=0$ 、$Cov(X,Y)=0$
$\iff D(X\pm Y)=D(X)+D(Y)$
$\iff E(XY)=E(X)E(Y)$：最常用
} 用随机变量的数字特征来描述

四、矩

1. 随机变量X的k阶原点矩 $\alpha_k=E(X^k)$　　若 $X\sim f(x)$，则 $E(X^k)=\int_{-\infty}^{+\infty}x^kf(x)dx$

1阶原点矩 $E(X)$；2阶原点矩 $E(X^2)$

2. X的k阶中心矩：$\beta_k=E\{[X-E(X)]^k\}$　　若 $X\sim f(x)$，则 $E\{[X-E(X)]^k\}=\int_{-\infty}^{+\infty}[x-E(x)]^kf(x)dx$

2阶中心矩：$D(X)$　　注：$\beta_1=E\{X-E(X)\}=0$

3. X、Y的k+l阶混合原点矩 $\alpha_{k,l}=E(X^kY^l)$

4. X、Y的k+l阶混合中心矩 $\beta_{k,l}=E\{[X-E(X)]^k[Y-E(Y)]^l\}$

$k=l=1$ 时，2阶混合中心矩 $Cov(X,Y)$

No.

Date

随机变量——概率分布(分布函数或分布列和概率密度)

完整地描述随机变量的统计规律

——数字特征(数学期望、方差、协方差、相关系数、矩)

集中地反映随机变量的某些统计特性

极限定理

1.切比雪夫不等式

对任意随机变量X,方差为D(X),对 $\forall\varepsilon>0$,有 $P\{|X-E(X)|\geq\varepsilon\}\leq\frac{D(X)}{\varepsilon^2}$

作用:在X的概率分布未知的情况下,用E(X)和D(X)估计X的概率分布

2.定义:设 $\{Z_n\}$ $n=1,2,\cdots$ 是一个随机变量序列,a是一个常数,若 $\forall\varepsilon>0$,有

$$\lim_{n\to\infty}P\{|Z_n-a|<\varepsilon\}=1$$

则称序列 $\{Z_n\}$ 依概率收敛于a,记为 $Z_n\xrightarrow{P}a$,$(n\to\infty)$

(只需理解)

3.辛钦大数定理

设 $X_1,X_2,\cdots,X_n,\cdots$ 独立同分布,其期望 $E(X_k)=\mu$,则对 $\forall\varepsilon>0$,有:$\lim_{n\to\infty}P\{|\frac{1}{n}\sum_{k=1}^{n}X_k-\mu|<\varepsilon\}=1$

理解:$\frac{1}{n}\sum_{i=1}^{n}X_i$ 即随机变量X在n次重复独立试验中n个观测值的算术平均值

当试验次数足够多大时,$\frac{1}{n}\sum_{k=1}^{n}X_k\xrightarrow{P}\mu$,$(n\to\infty)$ 实践:用测量值的算术平均值作为近似值

4.切比雪夫大数定律

设 $X_1,X_2,\cdots,X_n,\cdots$ 独立,其期望 $E(X_k)$,$k=1,2,\cdots$,则对 $\forall\varepsilon>0$,有 $\lim_{n\to\infty}P\{|\frac{1}{n}\sum_{k=1}^{n}X_k-\frac{1}{n}\sum_{k=1}^{n}E(X_k)|<\varepsilon\}=1$

理解:当试验次数足够多大时,$\frac{1}{n}\sum_{k=1}^{n}X_k\xrightarrow{P}\frac{1}{n}\sum_{k=1}^{n}E(X_k)$,$(n\to\infty)$

称满足上式的随机变量序列 $\{X_k\}$ 服从大数定律

5.伯努利大数定律

设 f_A 是n次独立重复试验中事件A发生的次数,p是每次试验中事件A发生的概率,则对 $\forall\varepsilon>0$

$$\lim_{n\to\infty}P\{|\frac{f_A}{n}-p|<\varepsilon\}=1 \quad 即 \quad \frac{f_A}{n}\xrightarrow{P}P(A)\quad(n\to\infty)$$

理论上:当试验次数足够多大时,某事件发生的频率与概率之差可以无限小

实践上:当试验次数很大时,可以用事件的频率来近似代替事件的概率

No.

Date

中心极限定理

1.标准化变量☆

对任意随机变量X,若 $E(X)=\mu$, $D(X)=\sigma^2\neq 0$, 称 $X^*=\frac{X-\mu}{\sigma}=\frac{X-E(X)}{\sqrt{D(X)}}$ 为X的标准化变量.

则 X^* 满足 $E(X^*)=0$, $D(X^*)=1$

2.独立同分布的中心极限定理

若随机变量序列 $\{X_n\}$ $n=1,2,\cdots$ 独立同分布,且 $E(X_i)=\mu$, $D(X_i)=\sigma^2>0$,

则随机变量之和 $\sum_{i=1}^{n}X_i$ 的 $E(\sum_{i=1}^{n}X_i)=n\mu$, $D(\sum_{i=1}^{n}X_i)=n\sigma^2$

其标准化变量 $Y_n=\frac{\sum_{i=1}^{n}X_i-E(\sum_{i=1}^{n}X_i)}{\sqrt{D(\sum_{i=1}^{n}X_i)}}=\frac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{n}\sigma}$ 的分布函数记为 $F_n(x)$

对 $\forall x\in R$, 满足 $\lim_{n\to\infty}F_n(x)=\Phi(x)$, 即

$\frac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{n}\sigma}\overset{近似}{\sim}N(0,1)$ 标准正态分布　　公式: $\lim_{n\to\infty}P\left\{\frac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{n}\sigma}\leq x\right\}=\Phi(x)$

理解:若某被研究的随机变量是大量($n\to\infty$)独立随机变量的和,其中每一个随机变量对总和只起微小作用,则可以认为这个随机变量近似服从于正态分布

变形:随机变量 $\{X_n\}$ 的算术平均记为 $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$, 记 $Y=\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\overset{近似}{\sim}N(0,1)$

$\Rightarrow \overline{X}=\frac{\sigma}{\sqrt{n}}Y+\mu \overset{近似}{\sim} N(\mu,\frac{\sigma^2}{n})$

题型: $\{X_n\}$ $n=1,2,\cdots$ 独立同分布,随机变量 $Z=\sum_{i}X_i$, 求 $P(|Z|\leq m)$

解法: 由 $\frac{Z-n\mu}{\sqrt{n}\sigma}\sim N(0,1)$ 知 $P(|Z|\leq m)=P(-m\leq Z\leq m)=P\left(\frac{-m-n\mu}{\sqrt{n}\sigma}\leq\frac{Z-n\mu}{\sqrt{n}\sigma}\leq\frac{m-n\mu}{\sqrt{n}\sigma}\right)$

$=\Phi\left(\frac{m-n\mu}{\sqrt{n}\sigma}\right)-\Phi\left(\frac{-m-n\mu}{\sqrt{n}\sigma}\right)$　　Φ 可查表

3.棣莫弗-拉普拉斯定理

设 Y_n $(n=1,2,\cdots)$ 服从参数为 n,p (变化 / 固定) $(0<p<1)$ 的二项分布,即 $Y_n\sim B(n,p)$, $E(Y_n)=np$, $D(Y_n)=npq$

则 Y_n 的标准化变量 $\frac{Y_n-np}{\sqrt{npq}}\overset{近似}{\sim}N(0,1)$　　公式: $\lim_{n\to\infty}P\left\{\frac{Y_n-np}{\sqrt{npq}}\leq x\right\}=\Phi(x)$

用法:随机变量 $X\sim B(n,p)$, 当n充分大时, $\frac{X-np}{\sqrt{npq}}\sim N(0,1)$

总结:二项分布 $X\sim B(n,p)$, 成功次数为k的概率 $P_n(k)$ 的计算

① n很小时,直接计算　$P_n(k)=P\{X=k\}=C_n^k p^k q^{n-k}$

② n很大p很小时,(且 $\lambda=np<10$),用泊松逼近　$P_n(k)=P\{X=k\}=\frac{\lambda^k}{k!}e^{-\lambda}$

③ n很大p不是很小时, $\frac{X-np}{\sqrt{npq}}\overset{近似}{\sim}N(0,1)$

No.

Date

切比雪夫不等式的证明 $P\{|X-E(X)|\geq\varepsilon\}\leq\frac{D(X)}{\varepsilon^2}$

$$P\{|X-E(X)|\geq\varepsilon\}=\int_{|x-E(X)|\geq\varepsilon}f(x)dx\leq\int_{|x-E(X)|\geq\varepsilon}\frac{(x-E(X))^2}{\varepsilon^2}f(x)dx\leq\frac{1}{\varepsilon^2}\int_{-\infty}^{+\infty}(x-E(X))^2f(x)dx=\frac{D(X)}{\varepsilon^2}$$

例：证明 $P\{|X-C|\geq\varepsilon\}\leq\frac{E|X-C|}{\varepsilon}$，其中 C为常数 $\varepsilon>0$

$$P\{|X-C|\geq\varepsilon\}=\int_{|x-c|\geq\varepsilon}f(x)dx\leq\int_{|x-c|\geq\varepsilon}\frac{|x-c|}{\varepsilon}f(x)dx\leq\frac{1}{\varepsilon}\int_{-\infty}^{+\infty}|x-c|f(x)dx=\frac{E|X-C|}{\varepsilon}$$

No.

Date

第6章 数理统计的基本概念

数理统计：以概率论为理论基础，根据试验或观测到的数据，研究如何利用有效的方法对这些数据进行整理、分析和推断，从而对研究对象的性质和统计规律作出合理、科学的估计和推断

一、总体与样本

1、总体：所研究的全体元素构成的集合　　个体：组成总体的每个元素

总体分为有限总体和无限总体，容量很大的有限总体可以看作无限总体

总体和服从某个概率分布的随机变量 X 是一一对应的 $\Rightarrow$ 总体 X 与随机变量 X 等同

2、样本

把从总体 X 中随机抽检 n 个个体的试验称为抽样，n 称为容量

抽样结果是 n 个随机变量 $X_1, X_2, \cdots, X_n$，称为来自于总体 X 的一个容量为 n 的样本

对于某次具体的抽样，结果是 n 个确定数值 $x_1, x_2, \cdots, x_n$ 称为样本的一个观测值（或一个实现）

3、简单随机样本

若对总体 X 的 n 次观测是在相同条件下独立重复进行，则得到的样本 $X_1, X_2, \cdots X_n$ 满足

$$\left.\begin{array}{l}(1)\ X_1, X_2, \cdots, X_n \text{ 相互独立} \\ (2)\ X_1, X_2, \cdots, X_n \text{ 与总体 } X \text{ 同分布}\end{array}\right\} \text{样本} \Rightarrow \text{独立同分布}$$

称为简单随机样本 $\Rightarrow$ 只研究简单随机样本

4、样本的分布

设总体 X 的分布函数为 $F(x)$（有 $f(x)$ 或 $P\{X=a_i\}=p_i$）样本 $X_1, X_2, \cdots, X_n$ 取自总体 X

① $(X_1, X_2, \cdots, X_n)$ 的联合分布函数

$$F(x_1, x_2, \cdots, x_n) \overset{\text{独立}}{=} F_{X_1}(x_1) F_{X_2}(x_2) \cdots F_{X_n}(x_n) \overset{\text{同分布}}{=} \prod_{i=1}^{n} F(x_i)$$ 注意把 x 换为 x_i

② 对离散型样本 $X_1, X_2 \cdots X_n$，联合分布列

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

③ 对连续型样本 $X_1, X_2 \cdots X_n$，联合概率密度

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

No.

Date

二、直方图与经验分布函数

1、直方图

可以估计概率密度曲线

作法：给定总体 X 的一个容量为 n 的样本观测值 $x_1, x_2, \cdots, x_n$

Step 1. 找出样本观测值的最小值 $x_{(1)}$，最大值 $x_{(n)}$

Step 2. 确定组数、组距、组限

选取 a（略小于 $x_{(1)}$）和 b（略大于 $x_{(n)}$），区间 $(a,b]$ 为作图区间

将 $(a,b]$ 等分为 $\underbrace{m}_{组数}$ 个小区间，分点为 $a=t_0<t_1<t_2\cdots<t_m=b$ 组限

注意：某小区间内不能没有观测值；分点要比观测值多取一位小数

Step 3、数出观测值落在区间 $(t_i, t_{i+1}]$ 中的个数 $\underbrace{n_i}_{频数}$，计算 $\underbrace{\frac{n_i}{n}}_{频率}$

Step 4、画图，在横坐标上标出各分点 t_i，以区间 $\underbrace{(t_i, t_{i+1}]}_{分组区间}$ 为底，画出高度为 $\underbrace{\frac{n_i}{\Delta t_i}}_{\frac{频率}{组距}}$ 的矩形

2、经验分布函数

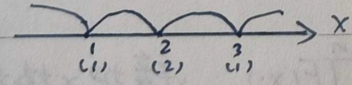
设总体 X 的一个容量为 n 的样本观测值 $x_1, x_2, \cdots, x_n$.

经验分布函数 $F_n(x)=\frac{N_n(x)}{n}$，$(-\infty<x<+\infty)$

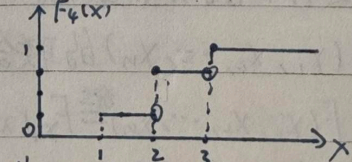
其中 $N_n(x)$ 为观测值 $x_1, x_2, \cdots, x_n$ 中不大于 x 的个数

定理：$\forall x$，$\lim\limits_{n\to\infty} F_n(x) = F(x)$

eg. 设总体 X 有一个容量为 4 的样本观测值 1, 2, 2, 3.



$$F_4(x)=\begin{cases} 0 & x<1 \\ \frac{1}{4} & 1\le x<2 \\ \frac{3}{4} & 2\le x<3 \\ 1 & 3\le x \end{cases}$$



$\Rightarrow F_n(x)$ 分为多个左闭右开的小区间；图像是阶梯型曲线（很像离散型的分布函数）

三、统计量

定义：设 $X_1, X_2, \cdots, X_n$ 为总体 X 的容量为 n 的样本，$T(x_1, x_2, \cdots, x_n)$ 是一个不依赖于未知参数的一个连续函数，则称随机变量 $T(X_1, X_2, \cdots, X_n)$ 为一个统计量

1、样本均值 $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i=\frac{1}{n}(X_1+X_2+\cdots+X_n)$

样本方差 $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2=\frac{1}{n-1}(\sum_{i=1}^{n}X_i^2-n\overline{X}^2)$ ←是总体方差的无偏估计

注意不是n

样本标准差 $S=\sqrt{S^2}$

注：若 $x_1,x_2,\cdots,x_n$ 是 $X_1,X_2,\cdots,X_n$ 的观测值，代入上式即能得到对应的观测值 $\overline{x}$，s^2，s

2、样本k阶原点矩 $A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$

样本k阶中心矩 $B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k$

样本均值即为样本1阶原点矩 $A_1=\frac{1}{n}\sum_{i=1}^{n}X_i=\overline{X}$

样本2阶中心矩 $B_2\triangleq S^{*2}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}^2=A_2-A_1^2$ ←是总体方差的矩估计（有偏）

四、抽样分布

即统计量的分布（即样本函数的分布）

三大抽样分布：

1、卡方分布：

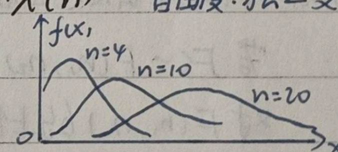
① 设 $X_1,X_2,\cdots,X_n$ 是来自于标准正态总体 $X\sim N(0,1)$ 的容量为n的样本

$Y=\sum_{i=1}^{n}X_i^2=X_1^2+X_2^2+\cdots+X_n^2$

则统计量Y的分布称为自由度为n的 χ^2 分布，记为 $Y\sim\chi^2(n)$ 自由度：独立变量的个数

特别地，若 $X\sim N(0,1)$ 则 $X^2\sim\chi^2(1)$

$\chi^2(n)$ 的概率密度是确定的，无须记



② 卡方分布的数字特征：

对于 $Y\sim\chi^2(n)$，有 $E(Y)=n$，$D(Y)=2n$

③ 卡方分布关于自由度的可加性

设随机变量 $X\sim\chi^2(m)$，$Y\sim\chi^2(n)$ 相互独立，则有：$X+Y\sim\chi^2(m+n)$

eg. 已知总体 $X\sim N(0,1)$，X_1,X_2,X_3,X_4 为其样本

① $X_1^2+X_2^2+X_3^2\sim$ $\underline{\chi^2(3)}$ ② $X_1^2+\frac{1}{2}(X_2+X_3)^2\sim$ $\underline{\chi^2(2)}$

解②：由于 $X_2\sim N(0,1)$、$X_3\sim N(0,1)$ 相互独立，$X_2+X_3\sim N(0,2)$

则标准化变量 $\frac{(X_2+X_3)-0}{\sqrt{2}}\sim N(0,1)$

$X_1^2+\frac{1}{2}(X_2+X_3)^2=X_1^2+[\frac{(X_2+X_3)-0}{\sqrt{2}}]^2\sim\chi^2(2)$

No.

Date

2、t分布

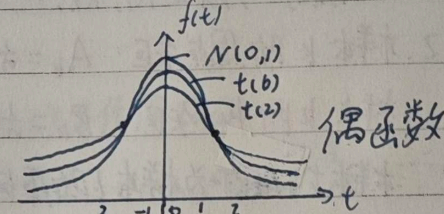
① 设随机变量X, Y相互独立, 且$X\sim N(0,1)$, $Y\sim \chi^2(n)$, 则

$$T=\frac{X}{\sqrt{\frac{Y}{n}}} \qquad 即 \frac{N(0,1)}{\sqrt{\frac{\chi^2(n)}{n}}}$$

称T为服从自由度为n的t分布, 记为$T\sim t(n)$

$t(n)$的概率密度也是确定的, 有以下性质:

(1) $f(t)$关于$t=0$对称



(2) $\lim\limits_{n\to\infty} f(t)=\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$ $t\in(-\infty,+\infty)$

即$n\to\infty$, $t(n)\overset{近似}{\sim}N(0,1)$

(3) 称$P(T>t_\alpha(n))=\int_{t_\alpha(n)}^{+\infty}f(t)dt=\alpha$的$t_\alpha(n)$为$t(n)$的上侧$\alpha$分位数,

由于对称性, $t_{1-\alpha}(n)=-t_\alpha(n)$

3. F分布

① 设随机变量X, Y相互独立, 且$X\sim \chi^2(n_1)$, $Y\sim\chi^2(n_2)$, 则

$$F=\frac{\frac{X}{n_1}}{\frac{Y}{n_2}} \qquad 即\frac{\frac{\chi^2(n_1)}{n_1}}{\frac{\chi^2(n_2)}{n_2}}$$

称F为服从自由度(n_1,n_2)的F分布, 记为$F\sim F(n_1,n_2)$

$F(n_1,n_2)$的概率密度也是确定的

② 性质: 若$X\sim F(n_1,n_2)$ 则$\frac{1}{X}\sim F(n_2,n_1)$

对$F(n_1,n_2)$的上侧α分位数$F_\alpha(n_1,n_2)$, 有$F_{1-\alpha}(n_1,n_2)=\frac{1}{F_\alpha(n_2,n_1)}$

例: 利用抽样分布的定义证明分布

1. 设$X\sim t(n)$, $Y=\frac{1}{X^2}$, 求证$Y\sim F(n,1)$

因为$X\sim t(n)$, $X=\frac{U}{\sqrt{\frac{V}{n}}}$, 其中$U\sim N(0,1)$, $V\sim\chi^2(n)$ 且U与V独立

故$Y=\frac{1}{X^2}=\frac{\frac{V}{n}}{U^2}=\frac{V/n}{U^2/1}\sim F(n,1)$, 其中$U^2\sim\chi^2(1)$

No.

Date

五、一般总体的样本均值和样本方差

1.定理：设总体X的期望 $E(X)=\mu$，方差 $D(X)=\sigma^2$，$X_1, X_2, \cdots, X_n$ 是来自总体X的样本，则

样本均值的期望 $E(\overline{X})=\mu$. 方差 $D(\overline{X})=\frac{\sigma^2}{n}$

样本方差的期望 $E(S^2)=\sigma^2$

六、正态总体的样本均值和样本方差

1、一个样本

定理：设 $X_1, X_2, \cdots, X_n$ 为正态总体 $N(\mu, \sigma^2)$ 的一个样本，则

① $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \sim N(\mu, \frac{\sigma^2}{n})$ 推论：$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

② 样本均值 $\overline{X}$ 与样本方差 S^2 相互独立

③ $\frac{n-1}{\sigma^2}S^2 \sim \chi^2(n-1)$ 注：同理，$\frac{n}{\sigma^2}S^{*2} \sim \chi^2(n-1)$ 这条用于求 $D(S^2)$

④ $\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

2、两个样本

定理：设 $X_1, X_2 \cdots, X_{n_1}$ 和 $Y_1, Y_2, \cdots, Y_{n_2}$ 分别是来自正态总体 $N(\mu_1, \sigma_1^2)$ 和 $N(\mu_2, \sigma_2^2)$ 的两个样本，它们相互独立，则

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

② 当 $\sigma_1=\sigma_2=\sigma$ 时 $\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$

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

KOKUYO

No.

Date

第7章 参数估计

在实际问题中,所研究的总体分布类型往往是已知的,但依赖于一个或几个未知参数,这时,从样本估计总体分布中的未知参数就是参数估计问题

一、点估计

1.点估计问题:

(1) 总体X的分布形式已知,θ是总体X的未知参数,可以用X的样本$X_1, X_2, \cdots, X_n$的一个统计量$\hat{\theta}=\hat{\theta}(X_1, X_2, \cdots, X_n)$来估计$\theta$,称$\hat{\theta}$为$\theta$的估计量

对具体的样本观测值$x_1, x_2, \cdots, x_n$,估计量$\hat{\theta}$的值$\hat{\theta}(x_1, x_2, \cdots, x_n)$称为$\theta$的估计值

估计量(一个函数)和估计值(函数值)统称为估计

注:若未知参数有多个,θ可理解为向量 $\theta=(\theta_1, \theta_2, \cdots, \theta_n)$

(2) 点估计的两种常用方法:

2、矩估计法

(1) 理论基础: $n\to\infty$时, $A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k \xrightarrow{P} \alpha_k=E(X^k)$ $k=1,2,\cdots$

(2) 矩估计法:用样本k阶原点矩$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$估计总体k阶原点矩$\alpha_k=E(X^k)$

(3) 矩估计求解步骤: 设总体X的分布中有m个未知参数$\theta_1, \theta_2, \cdots, \theta_m$ (一般最多为2个)

Step 1. 求总体各阶原点矩 $E(X^k)$, $k=1,2,\cdots,m$ 这个一定是第一步

Step 2. 令样本的各阶原点矩等于总体各阶原点矩,得到含m个未知参数的m个方程

$$\begin{cases} A_1=E(X) \\ A_2=E(X^2) \\ \vdots \\ A_m=E(X^m) \end{cases}$$

Step 3. 解上述方程,得到$\hat{\theta}_1, \hat{\theta}_2 \cdots \hat{\theta}_m$

注:总体有几个未知参数就建立几个方程;也可使用到$E(X^k)\to$反解$\theta_k\to$估计的步骤

(4) 矩估计常用公式: $E(X^2)=D(X)+E^2(X)$ $S^{*2}=A_2-\overline{X}^2$

(5) 常用结论:对服从任何分布的X,均值μ,方差$\sigma^2>0$都存在且均未知,样本为$X_1, X_2, \cdots, X_n$,则

μ, σ^2的矩估计: $\hat{\mu}=\overline{X}$

$$\hat{\sigma}^2=S^{*2}=A_2-A_1^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$$

样本2阶中心矩

$\Rightarrow$总体方差的矩估计是有偏的

No.

Date

3、最大似然估计法

(1) 思想：设总体 X 的分布类型已知，但分布中含有 m 个未知参数 $\theta_1, \theta_2, \cdots, \theta_m$，$X_1, X_2 \cdots, X_n$ 为来自 X 的一个样本，$x_1, x_2, \cdots, x_n$ 为一个样本值，

称 $L(x_1, x_2, \cdots, x_n; \theta_1, \theta_2, \cdots, \theta_m) = \prod_{i=1}^{n} P\{X = x_i\}$ 或 $\prod_{i=1}^{n} f(x_i; \theta_1, \theta_2, \cdots, \theta_m)$

为样本的似然函数，使得似然函数取到最大值的参数值 $\hat{\theta}_i(x_1, x_2, \cdots, x_n)$ $i=1,2,\cdots,m$

称为 θ_i 的最大似然估计值，相应的统计量 $\hat{\theta}_i(X_1, X_2, \cdots, X_n)$ 称为最大似然估计量

(2) 解题步骤

Step 1：若题目没给出样本值，则设为 $x_1, x_2, \cdots, x_n$

Step 2：写出似然函数 L

若 X 为离散型 $L(x_1, x_2, \cdots, x_n; \theta_1, \theta_2, \cdots, \theta_m) = \prod_{i=1}^{n} P\{X = x_i\} \longrightarrow$ 就是联合分布列

若 X 为连续型 $L(x_1, x_2, \cdots, x_n; \theta_1, \theta_2, \cdots, \theta_m) = \prod_{i=1}^{n} f(x_i; \theta_1, \theta_2, \cdots, \theta_m) \longrightarrow$ 联合概率密度

Step 3：求出对数似然函数 $\ln L$

利用 $\ln(uv) = \ln u + \ln v$. $\ln a^b = b\ln a$

取对数方便求偏导数，求最值点；取对数不会改变最值点

$\ln L(x_1, x_2, \cdots, x_n; \theta_1, \theta_2, \cdots, \theta_m) = \sum_{i=1}^{n} \ln P\{X = x_i\}$ 或 $\sum_{i=1}^{n} \ln f(x_i; \theta_1, \theta_2, \cdots, \theta_m)$

Step 4：对数似然函数关于各未知参数求偏导数，令其为0，得到对数似然方程

$$\frac{\partial \ln L}{\partial \theta_i} = 0 \quad (i = 1, 2, \cdots, m)$$

Step 5：求解上述方程组，若有解，则解 $\hat{\theta}_i$ 就是 $\theta_1, \theta_2, \cdots, \theta_m$ 的最大似然估计量

若方程组无解：则观察似然函数 L 达到最大值时的 θ_i 即可

(3) 最大似然估计的不变性

设 $\hat{\theta}$ 是 X 的概率分布中参数 θ 的最大似然估计，又假设 $u=u(\theta)$ 具有单值的反函数 $\theta=\theta(u)$，则 $\hat{u}=u(\hat{\theta})$ 就是 u 的最大似然估计

例：若已得到方差 σ^2 的最大似然估计量 $\hat{\sigma}^2$，

由 $\sigma=\sqrt{\sigma^2}$，即 $u=u(\sigma^2)=\sqrt{\sigma^2} \xrightarrow{\text{反函数}} \sigma^2=u^2 \ (u \geq 0)$ 是单值的，

故标准差 σ 的最大似然估计为 $\hat{\sigma}=\sqrt{\hat{\sigma}^2}$

Campus

No.

Date

例：设总体 $X\sim U[a,b]$，a、b为未知参数，样本 $X_1,X_2,\cdots,X_n$ 来自 X，求a、b的点估计

1、矩估计法：

由 $X\sim U[a,b]$，$f_X(x)=\begin{cases}\frac{1}{b-a} & a\le x\le b\\ 0 & 其他\end{cases}$ $E(X)=\frac{a+b}{2}$ $D(X)=\frac{(b-a)^2}{12}$

因为X有两个未知参数a、b，故计算X的一阶、二阶原点矩

$E(X)=\frac{a+b}{2}$，$E(X^2)=D(X)+E^2(X)=\frac{(b-a)^2}{12}+(\frac{a+b}{2})^2$

令 $\begin{cases}A_1=\overline{X}=E(X)\\ A_2=\frac{1}{n}\sum_{i=1}^{n}X_i^2=E(X^2)\end{cases}$ 即 $\begin{cases}A_1=\frac{a+b}{2} & ①\\ A_2=\frac{(b-a)^2}{12}+(\frac{a+b}{2})^2 & ②\end{cases}$ 是关于a、b的方程组

解法：先将①代入②，得 $(b-a)^2=12(A_2-A_1^2)=12S^{*2}$

故 $b-a=2\sqrt{3}S^*$，又由①有 $b+a=2\overline{x}$，从而

a、b的矩估计量为 $\hat{a}=\overline{x}-\sqrt{3}S^*$ $\hat{b}=\overline{x}+\sqrt{3}S^*$

2、最大似然估计法

设样本值为 $x_1,x_2,\cdots,x_n$，$f(x_i)=\begin{cases}\frac{1}{b-a} & a\le x_i\le b\\ 0 & 其他\end{cases}$

① 似然函数 $L(a,b)=\prod_{i=1}^{n}f(x_i)=\begin{cases}\frac{1}{(b-a)^n} & a\le x_1\le b, a\le x_2\le b,\cdots,a\le x_n\le b\\ 0 & 其他\end{cases}$

② 对数似然函数，当 $a\le x_1\le b, a\le x_2\le b,\cdots,a\le x_n\le b$ 时

$\ln L(a,b)=-n\ln(b-a)$

③ 对数似然方程 $\begin{cases}\frac{\partial\ln L(a,b)}{\partial a}=\frac{n}{b-a}=0\\ \frac{\partial\ln L(a,b)}{\partial b}=-\frac{n}{b-a}=0\end{cases}$ 是无解的（说明不能用求偏导的方式得到解，不代表最大似然估计法失效）

观察 $L(a,b)=\frac{1}{(b-a)^n}$ $a\le x_1\le b, a\le x_2\le b,\cdots a\le x_n\le b$，要 $L(a,b)$ 最大，则b应最小，a应最大

a满足 $a\le x_1,x_2,\cdots,x_n\Rightarrow a\le\min\{x_1,x_2,\cdots,x_n\}\Rightarrow\hat{a}=\min\{x_1,x_2,\cdots,x_n\}$

b满足 $x_1,x_2,\cdots,x_n\le b\Rightarrow b\ge\max\{x_1,x_2,\cdots,x_n\}\Rightarrow\hat{b}=\max\{x_1,x_2,\cdots,x_n\}$

$\Rightarrow$ a、b的最大似然估计量 $\hat{a}=\min\{X_1,X_2,\cdots,X_n\}$，$\hat{b}=\max\{X_1,X_2,\cdots,X_n\}$

KOKUYO

例. 已知离散型X的分布列如下，其中$\theta\in(0,\frac{1}{2})$未知，X的一个样本值为 3,1,3,0,3,1,2,3

求θ的点估计

X	0	1	2	3
P	θ^2	$2\theta(1-\theta)$	θ^2	$1-2\theta$

1. 矩估计法：

只有1个未知参数θ，只需求一阶原点矩

① $E(X)=0\cdot\theta^2+1\cdot2\theta(1-\theta)+2\cdot\theta^2+3\cdot(1-2\theta)=3-4\theta$

② 令$A_1=\overline{X}=E(X)$，有$\overline{X}=3-4\theta$

即θ的矩估计量 $\hat{\theta}=\frac{3-\overline{X}}{4}$

又$\overline{x}=\frac{1}{8}\sum_{i=1}^{8}x_i=2$，可得$\theta$的矩估计值$\hat{\theta}=\frac{1}{4}$

2. 最大似然估计法：

① $L(\theta)=P\{X_1=3,X_2=1,X_3=3,X_4=0,X_5=3,X_6=1,X_7=2,X_8=3\}=4\theta^6(1-\theta)^2(1-2\theta)^4$

② $\ln L(\theta)=\ln4+6\ln\theta+2\ln(1-\theta)+4\ln(1-2\theta)$

③ $\frac{d\ln L(\theta)}{d\theta}=\frac{6}{\theta}-\frac{2}{1-\theta}-\frac{8}{1-2\theta}=\frac{24\theta^2-28\theta+6}{\theta(1-\theta)(1-2\theta)}=0$ 解这个方程，可得 $\theta_1=\frac{7-\sqrt{13}}{12}$，$\theta_2=\frac{7+\sqrt{13}}{12}>\frac{1}{2}$

故θ的最大似然估计值 $\hat{\theta}=\frac{7-\sqrt{13}}{12}$

No.

Date

4、估计量的评选标准

(1) 无偏性

引入：估计量是一种统计量，所以是随机变量，对不同的样本观测值，有不同的估计值，希望这些估计值最好在待估参数的真值附近，即：

定义：设 $\hat{\theta}=\theta(X_1,X_2,\cdots,X_n)$ 是 θ 的估计量，若 $\hat{\theta}$ 满足 $E(\hat{\theta})=\theta$，则称 $\hat{\theta}$ 是 θ 的无偏估计

定理：无论总体服从什么分布，都有：

1、样本k阶原点矩 $A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$ 是总体k阶原点矩 $\alpha_k=E(X^k)$ 的无偏估计

特别地，样本均值是总体的数学期望的无偏估计（样本的组合系数和为1的线性组合也是）

2、样本方差 $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ 是总体方差 $\sigma^2=D(X)$ 的无偏估计

而样本二阶中心矩 $S^{*2}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$ 估计 σ^2 是有偏的

(2) 有效性

引入：若 $E(\hat{\theta}_1)=\theta$，$E(\hat{\theta}_2)=\theta$，两个无偏估计量如何评价？

若 $\hat{\theta}_1$ 的取值比 $\hat{\theta}_2$ 的取值更集中在 θ 附近，则认为 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 更理想：

定义：设 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 均为 θ 的无偏估计，即 $E(\hat{\theta}_1)=E(\hat{\theta}_2)=\theta$，若 $D(\hat{\theta}_1)\le D(\hat{\theta}_2)$

则称 $\hat{\theta}_1$ 较 $\hat{\theta}_2$ 有效

(3) 相合性

引入：无偏性、有效性都是在样本容量n固定的前提下讨论的，则：

定义：设 $\hat{\theta}=\theta(X_1,X_2,\cdots,X_n)$ 是 θ 的估计量，若 $\hat{\theta}$ 依概率收敛于 θ，即 $\forall\varepsilon>0,\lim_{n\to\infty}P\{|\hat{\theta}-\theta|\ge\varepsilon\}=0$

则称 $\hat{\theta}$ 是 θ 的相合估计量

注：相合性是对估计量的一个基本要求

性质：1、样本k阶原点矩 $A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$ 是总体k阶原点矩 $\alpha_k=E(X^k)$ 的相合估计

2、样本方差 S^2 与样本二阶中心矩 S^{*2} 都是总体方差 σ^2 的相合估计

定理：若待估参数 $\theta=g(\alpha_1,\alpha_2,\cdots,\alpha_k)$，其中g为连续函数，则 θ 的矩估计量 $\hat{\theta}=g(A_1,A_2,\cdots,A_k)$ 是 θ 的相合估计

No.

Date

二、区间估计

置信区间　　　　　　　　　　　　　　　　置信水平

引入：区间估计：给出未知参数的一个估计范围，并给出此范围包含参数 θ 真值的可信程度

1.置信区间

定义：设总体 X 的分布函数为 $F(x;\theta)$，含有一个未知参数 θ，$\theta\in\Theta$

对于给定值 α $(0<\alpha<1)$ α 一般很小 若由来自 X 的样本 $X_1、X_2、\cdots X_n$ 确定两个

统计量 $\underline{\theta}=\underline{\theta}(X_1,X_2,\cdots,X_n)$，$\overline{\theta}=\overline{\theta}(X_1,X_2,\cdots,X_n)$，$\underline{\theta}<\overline{\theta}$，

对于任意的 $\theta\in\Theta$，满足 $P\{\underline{\theta}<\theta<\overline{\theta}\}=1-\alpha$

则称随机区间 $(\underline{\theta},\overline{\theta})$ 是 θ 的置信水平为 $1-\alpha$ 的置信区间.

$\underline{\theta}$，$\overline{\theta}$ 分别称为置信下限和置信上限

注：1.由于统计量 $\underline{\theta}$、$\overline{\theta}$ 是随机变量，故 $(\underline{\theta},\overline{\theta})$ 是一个随机区间

若反复抽样多次（样本容量均为 n），每一个样本观测值确定一个区间 $(\underline{\theta},\overline{\theta})$，这样的区间要么包含真值（约占 $100(1-\alpha)\%$），要么不包含真值（约占 $100\alpha\%$）

eg. 若 $\alpha=0.01$，反复抽样1000次，则相应得到的1000个区间中包含 θ 真值的区间约为990个，不包含 θ 真值的约为10个

2.若 X 为连续型，给定 α，利用 $P\{\underline{\theta}<\theta<\overline{\theta}\}=1-\alpha$ 求置信区间，若 X 为离散型，则可为 $\geq$

2.求未知参数 θ 的置信区间的方法

Step1. 寻找一个样本 $X_1、X_2、\cdots X_n$ 和 θ 的函数 $W=W(X_1,X_2,\cdots,X_n;\theta)$

使 W 的分布不依赖于任何未知参数，称函数 W 为枢轴量

Step2、对于给定的置信水平 $1-\alpha$，定出两个常数 a,b，使得 $P\{a<W(X_1,X_2,\cdots X_n;\theta)<b\}=1-\alpha$

Step3、从 $a<W(X_1,X_2,\cdots,X_n;\theta)<b$ 中反解得到等价的不等式 $\underline{\theta}<\theta<\overline{\theta}$，

则 $(\underline{\theta},\overline{\theta})$ 即为所求

注：1.枢轴量 W 的构造方法：从 θ 的点估计着手构造，利用正态总体的样本均值和样本方差的性质

2. a、b 一般取 W 的上分位点，方便查表

3、单个正态总体参数的区间估计

设总体 $X \sim N(\mu,\sigma^2)$，设 $X_1, X_2, \cdots, X_n$ 是来自X的一个样本，样本均值 $\overline{X}$，样本方差 S^2，下求的均为置信水平为 $1-\alpha$ 的置信区间

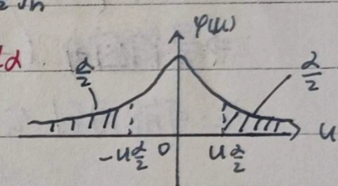
(1) 方差 σ^2 已知，期望 μ 未知，求 μ 的置信区间

由单个正态总体的样本均值 $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$ 知 $\frac{\overline{X}-\mu}{\sigma}\sqrt{n} \sim N(0,1)$

$P\{-u_{\frac{\alpha}{2}} < \frac{\overline{X}-\mu}{\sigma}\sqrt{n} < u_{\frac{\alpha}{2}}\} = 1-\alpha$，其中 $u_{\frac{\alpha}{2}}$ 为标准正态分布的上 $\frac{\alpha}{2}$ 分位点

由 $-u_{\frac{\alpha}{2}} < \frac{\overline{X}-\mu}{\sigma}\sqrt{n} < u_{\frac{\alpha}{2}}$ 反解 μ，得 $\overline{X} - u_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + u_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$

$\Rightarrow$ 置信区间 $(\overline{X} - u_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{X} + u_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}})$ 用 $\Phi(u_\alpha) = 1-\alpha \Rightarrow u_\alpha$



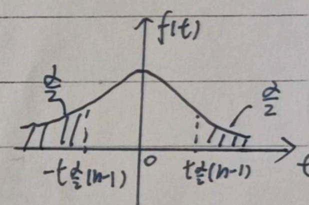
注：置信水平为 $1-\alpha$ 的置信区间是无穷多的，但由上述方法得到的区间长度是最短的

(2) 方差 σ^2 未知，期望 μ 未知，求 μ 的置信区间

由 $\frac{\overline{X}-\mu}{S}\sqrt{n} \sim t(n-1)$ 知 $P\{-t_{\frac{\alpha}{2}}(n-1) < \frac{\overline{X}-\mu}{S}\sqrt{n} < t_{\frac{\alpha}{2}}(n-1)\} = 1-\alpha$

反解 μ，得 $\overline{X} - t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}$

$\Rightarrow$ 置信区间 $(\overline{X} - t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}, \overline{X} + t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}})$

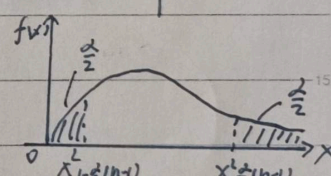


(3) 方差 σ^2、期望 μ 未知，求 σ^2 的置信区间

由 $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ 知 $P\{\chi^2_{1-\frac{\alpha}{2}}(n-1) < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\frac{\alpha}{2}}(n-1)\} = 1-\alpha$

反解 σ^2，得 $\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}$

$\Rightarrow$ 置信区间 $\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}, \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\right)$



4、两个正态总体参数的区间估计

设总体 $X \sim N(\mu_1, \sigma_1^2)$，$Y \sim N(\mu_2, \sigma_2^2)$ 相独立，分别有样本 $X_1, X_2, \cdots, X_{n_1}$ 和 $Y_1, Y_2, \cdots, Y_{n_2}$

设样本均值分别为 $\overline{X}$、$\overline{Y}$，样本方差分别为 S_1^2、S_2^2，求置信水平为 $1-\alpha$ 的置信区间

(1) σ_1^2、σ_2^2 均已知，求 $\mu_1-\mu_2$ 的置信区间

由 $\overline{X} \sim N(\mu_1, \frac{\sigma_1^2}{n_1})$，$\overline{Y} \sim N(\mu_2, \frac{\sigma_2^2}{n_2})$，$\overline{X}$、$\overline{Y}$ 相独立知 $\overline{X}-\overline{Y} \sim N(\mu_1-\mu_2, \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2})$

得枢轴量 $u = \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1)$ 由 $P\{-u_{\frac{\alpha}{2}} < u < u_{\frac{\alpha}{2}}\} = 1-\alpha$，反解得

$\Rightarrow$ 置信区间 $\left(\overline{X}-\overline{Y} - u_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}, \overline{X}-\overline{Y} + u_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$

No.

Date

(2) $\sigma_1^2=\sigma_2^2=\sigma^2$，但 σ^2 未知，求 $(\mu_1-\mu_2)$ 的置信区间

枢轴量 $t=\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$　　其中 $S_w^2=\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$

利用 $P\{-t_{\frac{\alpha}{2}}(n_1+n_2-2)<t<t_{\frac{\alpha}{2}}(n_1+n_2-2)\}=1-\alpha$ 得

$\Rightarrow$ 置信区间 $\left(\bar{X}-\bar{Y}-t_{\frac{\alpha}{2}}(n_1+n_2-2)S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}},\ \bar{X}-\bar{Y}+t_{\frac{\alpha}{2}}(n_1+n_2-2)S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$

(3) μ_1、μ_2 未知，求 $\dfrac{\sigma_1^2}{\sigma_2^2}$ 的置信区间

枢轴量 $F=\dfrac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$

利用 $P\{F_{1-\frac{\alpha}{2}}(n_1-1,n_2-1)<F<F_{\frac{\alpha}{2}}(n_1-1,n_2-1)\}=1-\alpha$

$\Rightarrow$ 置信区间 $\left(\dfrac{S_1^2}{S_2^2}\dfrac{1}{F_{\frac{\alpha}{2}}(n_1-1,n_2-1)},\ \dfrac{S_1^2}{S_2^2}\dfrac{1}{F_{1-\frac{\alpha}{2}}(n_1-1,n_2-1)}\right)$

注：可能用到公式：$F_{1-\alpha}(n_1,n_2)=\dfrac{1}{F_{\alpha}(n_2,n_1)}$