ELEYANG DESIGN

La Vita E Bella

MATH

ELEYANG DESIGN

12

Content
内容

【概率论】

Review
复习记录

/ / / / /

Title 概率论基本概念

Date /

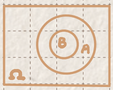
Keywords 关键词	Notes 笔记

Review 复习记录 / / / / /

- 集合论
- 概率

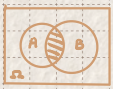
§. 集合论

▶ [Theorem] 集合论



$\subset$

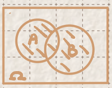
($\subset$ 则 $A \equiv B$)



A和B同时发生

$A \cap B$ / AB

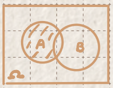
积事件



A和B至少发生一个

$A \cup B$

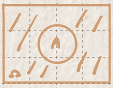
和事件



A发生B不发生

$A - B$

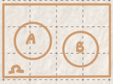
差事件



A不发生

$\bar{A}$

逆事件



$A \cap B = \varnothing$

互斥事件

[Prop] 交换律： $A \cup B = B \cup A$ $\quad A \cap B = B \cap A$

结合律： $A \cup (B \cup C) = (A \cup B) \cup C$ $\quad A \cap (B \cap C) = (A \cap B) \cap C$

分配律： $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ $\quad A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

德摩根律： $\overline{A \cup B} = \bar{A} \cap \bar{B}$ $\quad \overline{A \cap B} = \bar{A} \cup \bar{B}$

§. 概率

▶ [Theorem] 概率公理： 1. 非负性 $\quad 0 \le P(A) \le 1$

2. 正则性 $\quad$ 对必然事件 S $\quad P(S) = 1$

3. 可列可加性 $\quad$ 若有互不相容的事件 $A_1, A_2 \cdots$ $\quad P(\bigcup A_i) = \sum P(A_i)$

[Prop] (i) $P(\varnothing) = 0$

(ii) 有限可加性 (略)

(iii) $A \subset B$ $\quad P(B-A) = P(B) - P(A)$ $\quad P(B) \ge P(A)$

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

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

▶ [Def] 古典概型 $\Omega = \{e_1\ e_2 \cdots e_n\}$

$P(\{e_1\}) = P(\{e_2\}) = \cdots = P(\{e_n\})$ $\quad P(\Omega) = 1 = \sum P(\{e_i\})$

$\Rightarrow P(\{e_i\}) = \frac{1}{n}$

对事件 $A = \{e_{i_1}, e_{i_2} \cdots e_{i_k}\}$ $\quad P(A) = \frac{k}{n}$

则 $P(A) = \frac{A中基本点数}{\Omega中基本点数}$

▶ [Def] 几何概型：基于几何图形长度、面积、体积等算出的概率

Summary 总结

Content 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

Chapter I

Title 排列组合

Date /

Keywords 关键词	Notes 笔记

Review 复习记录 / / / / /

- 排列组合
- 条件概率
- 独立性

§. 排列组合

[Prop] (i) 加法原理：完成某件事 有 s 类 各有 $r_1 \sim r_s$ 种方式

分类 be like 方法数 $= r_1 + \cdots + r_s$

(ii) 乘法原理：完成某件事 有 s 步 各有 $r_1 \sim r_s$ 种方式

分步 be like 方法数 $= r_1 \cdot r_2 \cdots r_s$

[Theo] 排列 方法数 A_n^m （全排列 m!）

组合 方法数 C_n^m

[Remark] $C_n^m = C_n^{n-m}$　　$C_n^0 + C_n^1 + \cdots + C_n^n = 2^n$

§. 条件概率

[Def] $P(B|A)$ 描述 A 发生条件下 B 发生的概率　$P(B|A) = \frac{P(AB)}{P(A)}$

[Theorem] 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 $P(B_i) > 0$ 则

全概率公式　$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_n)P(B_n)$

[Def] E, S, $B_1 \sim B_n$ 为 E 的一组事件

if (i) $B_i B_j \ (i \neq j) = \varnothing$　(ii) $B_1 \cup B_2 \cup \cdots \cup B_n = S$

then 称 $B_1 \sim B_n$ 为 S 的 一个划分

[Theorem] E, S, A, $B_1 \sim B_n$　且 $P(A) > 0$　$P(B_i) > 0$

贝叶斯公式　then　$P(B_i|A) = \dfrac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)}$

§. 独立性

[Def] 设 A B 为两个事件　if $P(AB) = P(A)P(B)$　then A, B 相互独立

[Prop] if A B 相互独立　$P(A) > 0$, then $P(B|A) = P(B)$

[Remark] A B C 两两独立 与 相互独立：相互独立需 $P(ABC) = P(A)P(B)P(C)$

Summary 总结

Chapter Ⅱ

Title 随机变量及分布(离散 Part)　Date /

Keywords 关键词	Notes 笔记　Review 复习记录 / / / / /

- 随机变量
- 三种重要分布
- 分布函数

§.随机变量

▶[Def] 设随机试验的样本空间为 $S=\{e\}$　$X=X(e)$ 是定义在样本空间 S 上的实值单值函数

称 $X=X(e)$ 为随机变量

▶[Theorem] X所有可能取值为 $x_k\ (k=1,2,\cdots)$　事件 $\{X=x_k\}$ 概率为 $P\{X=x_k\}=p_k$　$k=1,2,\cdots$

p_k 满足 (i) $p_k\geq 0$　(ii) $\sum\limits_{k=1}^{\infty}p_k=1$　称为离散性随机变量 X 的分布律

三种重要分布

[0-1分布]　$P\{X=1\}=p$　$P\{X=0\}=1-p$

[伯努利试验]　对0-1分布试验重复 n 次(每次发生概率为 p)

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

此种情况下 X 服从参数为 n,p 的二项分布　记 $X\sim b(n,p)$

[泊松分布]　X 可能取值 $0,1,2\cdots$　$P(X=k)=\frac{\lambda^k e^{-\lambda}}{k!}$　$k=0,1,2\cdots$

称 X 服从参数为λ的泊松分布　记 $X\sim\pi(\lambda)$

[$\lim\limits_{n\to\infty}C_n^k p_n^k(1-p_n)^{n-k}=\frac{\lambda^k e^{-\lambda}}{k!}$　$(np_n=\lambda)$　即泊松分布为 n 很大 p 很小的二项分布]

§.分布函数

▶[Def] 设 X 为一个随机变量，x 为任意实数

$F(x)=P\{X\leq x\}$　$x\in R$　称为 X 的分布函数

此时 $P\{x_1\leq X\leq x_2\}=F(x_2)-F(x_1)$

[Prop] 1° $F(x)$ 为非减函数

2° $0\leq F(x)\leq 1$　且 $F(-\infty)=0$　$F(\infty)=1$

3° $F(x)$ 为右连续函数

Summary 总结

Title 连续性随机变量

Date /

Keywords 关键词 | Notes 笔记 | Review 复习记录 / / / / /

- 概率密度
- 三种重要分布
- 正态分布
- 随机变量函数

§. 概率密度

▶【Def】对随机变量 X 分布函数 $F(x)$ ∃非负函数 $f(x)$

对 $\forall x \in R$　$F(x) = \int_{-\infty}^{x} f(t)\,dt$　$f(x)$ 即为 X 概率密度

【Prop】1° $f(x) \geq 0$　$\int_{-\infty}^{+\infty} f(x)\,dx = 1$

2° $P(x_1 \leq X \leq x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(t)\,dt$

3° if $f(x)$ 在 x 处连续　$F'(x) = f(x)$

三种重要分布

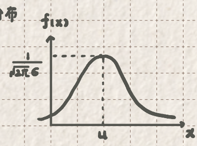
【均匀分布】 $f(x) = \begin{cases} \frac{1}{b-a} & a<x<b \\ 0 & \text{其它} \end{cases}$　称 X 均匀分布　记为 $X \sim U(a, b)$　$F(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a<x<b \\ 1 & x \geq b \end{cases}$

【指数分布】 $f(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & x>0 \\ 0 & x \leq 0 \end{cases}$　$\theta > 0$　称 X 指数分布　$F(x) = \begin{cases} 1 - e^{-\frac{x}{\theta}} & x>0 \\ 0 & x \leq 0 \end{cases}$

满足以下性质：$\forall s, t$　$P\{X > s+t \mid X > s\} = P\{X > t\}$　无记忆性

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

记为 $X \sim N(\mu, \sigma^2)$



$F(x) = \int_{-\infty}^{x} f(x)\,dx$

满足以下性质：1° 正态分布曲线关于 $x = \mu$ 对称

2° 当 $x = u$　$f(x)$ 取 max $\frac{1}{\sqrt{2\pi}\sigma}$

标准正态分布：$u = 0$　$\sigma^2 = 1$　$X \sim N(0, 1)$

$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$　$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}}\,dx$

3° if $X \sim N(u, \sigma^2)$　$Z = \frac{X-u}{\sigma} \sim N(0, 1)$

§. 正态分布

▶【Prop】$X \sim N(u, \sigma^2)$　$F(x) = P\{X \leq x\} = \Phi\left(\frac{x-u}{\sigma}\right)$

对 $(x_1, x_2]$　$P(x_1 < X \leq x_2) = F(x_2) - F(x_1) = \Phi\left(\frac{x_2-u}{\sigma}\right) - \Phi\left(\frac{x_1-u}{\sigma}\right)$

1σ 区间：$P\{\mu-\sigma < X \leq \mu+\sigma\} = \Phi(1) - \Phi(-1) = 68.28\%$

2σ 区间：$P\{\mu-2\sigma < x \leq \mu+2\sigma\} = 95.44\%$

3σ 区间：$P\{\mu-3\sigma < x \leq \mu+3\sigma\} = 99.74\%$

▶【Def】$X \sim N(0, 1)$　$P\{X \geq Z_\alpha\} = \alpha$　称 Z_α 为上 α 分位点

§. 随机变量的函数的分布

▶【Theorem】$X : f_X(x)$　设 $g(x)$ 处处可导且恒有 $g'(x) > 0$ / $g'(x) < 0$

then $Y = g(X)$ 为连续性随机变量

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

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

$h(y) = g^{-1}(x)$

Summary 总结

Content 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

Chapter Ⅲ

Title 二维随机变量 Date /

Keywords 关键词	Notes 笔记
• 二维随机变量 • pt.1 • pt.2	[Def] E随机试验 $S=\{e\}$ 样本空间 设 $X=X(e)$ $Y=Y(e)$ 定义在S上的随机变量 由它们构成向量 (X,Y) 称二维随机向量 [Def] (X,Y) 对 $\forall x,y\in R$ $F(x,y)=P\{(X\le x)\cap(Y\le y)\}\equiv P\{X\le x, Y\le y\}$ 称为二维随机变量 (X,Y) 的分布函数 / X、Y 的联合分布函数 它具有以下性质: [i] $F(x,y)$ 是 x、y 的不减函数 [ii] $0\le F(x,y)\le 1$ 且 $\forall$ fix y $F(-\infty,y)=0$ $\forall$ fix x $F(x,-\infty)=0$ $F(-\infty,-\infty)=0$ $F(\infty,\infty)=1$ [iii] $F(x,y)$ 关于 x、y 都右连续 [iv] $\forall (x_1,y_1)\ (x_2,y_2)$ $x_1<x_2$ $y_1<y_2$ $F(x_2,y_2)+F(x_1,y_1)-F(x_2,y_1)-F(x_1,y_2)\ge 0$ $P\{x_1<X\le x_2, y_1<Y\le y_2\} = F(x_2,y_2)-F(x_2,y_1)+F(x_1,y_1)-F(x_1,y_2)$
§. 二维随机变量	
§. pt.1 离散性	[Theorem] 设 (X,Y) 所有可能取值 (x_i,y_j) $i,j=1,2\cdots$ 记 $P(X=x_i,Y=y_j)=p_{ij}$ $(p_{ij}\ge 0$ $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}p_{ij}=1)$ 称 (X,Y) 分布律 $F(x,y)=\sum_{x_i\le x}\sum_{y_j\le y}p_{ij}$
§. pt.2 连续性	[Def] (X,Y) $F(x,y)$ if $\exists$ 非负函数 $f(x,y)$ s.t. $\forall x,y$ $F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(u,v)\,du\,dv$ then 称 (X,Y) 为连续型二维随机变量 $f(x,y)$ 称概率密度 [Prop] 1° $f(x,y)\ge 0$ $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(u,v)\,du\,dv=1$ 2° G为 xOy 上区域 $P\{(X,Y)\in G\}=\iint_G f(u,v)\,du\,dv$ 3° if $f(x,y)$ 在 (x,y) 上连续 $\frac{\partial^2 F}{\partial x\,\partial y}=f(x,y)$

Review 复习记录 / / / / /

Summary 总结

Title 边缘分布与条件分布

Date /

Keywords 关键词	Notes 笔记

Review 复习记录 / / / / /

- 边缘分布
- 条件分布

§. 边缘分布

[Def] (X,Y) X,Y 各自的分布函数 $F_X(x)$ $F_Y(y)$ 依次称为 (X,Y) 关于 X,Y 的边缘分布函数

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

[Remark] 对离散型 $F_X(x) = \sum\limits_{x_i \le x} \sum\limits_{j=1}^{\infty} p_{ij}$

记 $p_{i\cdot} = \sum\limits_{j=1}^{\infty} p_{ij} = P\{X = x_i\}$ $\quad$ $p_{\cdot j} = \sum\limits_{i=1}^{\infty} p_{ij} = P\{Y = y_j\}$

分别称 $p_{i\cdot}$ $p_{\cdot j}$ 为 (X,Y) 关于 X,Y 的边缘分布律

[Remark] 对连续性 $F_X(x) = \int_{-\infty}^{x}\int_{-\infty}^{+\infty} f(x,y)\,dx\,dy$

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

$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$ 分别称 (X,Y) 关于 X,Y 的边缘概率密度

[Prop] 设 (X,Y) $f(x,y) = \dfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{\dfrac{-1}{2(1-\rho^2)}\left[\dfrac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\dfrac{(x-\mu_1)(x-\mu_2)}{\sigma_1\sigma_2} + \dfrac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}$

then 称 (X,Y) 服从 $\mu_1\ \mu_2\ \sigma_1\ \sigma_2\ \rho$ 的二维正态分布 $\quad (X,Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$

$f_X(x) = \dfrac{1}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}$ $\quad$ $f_Y(y) = \dfrac{1}{\sqrt{2\pi}\,\sigma_2} e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}$

§. 条件分布

[Def] (X,Y) 离散 对 fix j if $P\{Y=y_j\} > 0$

then 称 $P\{X = x_i \mid Y = y_j\} = \dfrac{p_{ij}}{p_{\cdot j}}$ 为 $Y = y_j$ 下 X 的条件分布律 对 Y 同理

[Def] (X,Y) 连续 $f(x,y)$ $f_Y(y)$ 对 fix y if $f_Y(y) > 0$

then 称 $\dfrac{f(x,y)}{f_Y(y)}$ 为 $Y=y$ 条件下 X 的条件概率密度

记为 $f_{X|Y}(x|y) = \dfrac{f(x,y)}{f_Y(y)}$ 对 Y 同理

同时称 $\int_{-\infty}^{x} f_{X|Y}(x|y)\,dx = \int_{-\infty}^{x} \dfrac{f(x,y)}{f_Y(y)}\,dx$ 为 $Y=y$ 下的条件分布函数

记为 $P\{X \le x \mid Y = y\}$ 或 $F_{X|Y}(x|y)$

Summary 总结

Content
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24

Title 相互独立 Date /

Keywords 关键词 | Notes 笔记 | Review 复习记录 / / / / /

- 相互独立性

[Def] $F(x,y)$ $F_X(x)$ $F_Y(y)$

if $P(X\le x, Y\le y) = P(X\le x)P(Y\le y)$ or $F(x,y) = F_X(x)F_Y(y)$

then 称 X, Y 相互独立

[Remark] 对n维 $(X_1\ X_2 \cdots X_n)$

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

$= \int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_n} f(x_1\cdots x_n)\,dx_1\cdots dx_n$

§. 独立性

$F_{X_1}(x_1) = F(x_1, \infty, \infty \cdots \infty)$

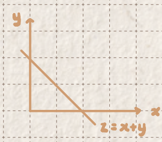
$F_{X_1X_2}(x_1,x_2) = F(x_1, x_2, \infty, \infty \cdots \infty)$

if $X_1 \sim X_n$ 相互独立 $F(x_1\cdots x_n) = F_{X_1}(x_1)\cdot F_{X_2}(x_2)\cdots F_{X_n}(x_n)$

$X_1\sim X_m$ $Y_1\sim Y_n$ $F(x_1\sim x_m, y_1\sim y_n) = F_1(x_1\sim x_m)F_2(y_1\sim y_n)$

则 $X_1\sim X_m$ $Y_1\sim Y_n$ 相互独立

if X, Y 独立 h, g 连续函数 h(X) g(Y) 独立



§. 两个随机变量函数分布 # 三种重要模型

[Z=X+Y] (X, Y) 二维连续 $f(x,y)$ $Z=X+Y$

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

if X, Y 相互独立 $f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y)f_Y(y)\,dy$ or $\int_{-\infty}^{\infty} f_X(x)f_Y(z-x)\,dx$ (卷积公式 $f_X * f_Y$)

[Remark] 独立 $X_i\sim N(\mu_i, \sigma_i^2)$ $i=1,2,3\cdots n$ $Z=\sum_{i=1}^{n} X_i \sim N(\sum_{i=1}^{n}\mu_i, \sum_{i=1}^{n}\sigma_i^2)$

[M=max{X,Y}, N=min{X,Y}] (X, Y) 独立 $F_X(x)$ $F_Y(y)$

$F_{max}(z) = P(M\le z) = P\{X\le z, Y\le z\} = F_X(z)F_Y(z)$

$F_{min}(z) = P(N\le z) = 1-P\{N>z\} = 1-[1-F_X(z)][1-F_Y(z)]$

// $F_{max}(z) = F_{X_1}(z)\cdot F_{X_2}(z)\cdots F_{X_n}(z)$

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

[Z = Y/X, Z = XY] (X, Y) $f(x,y)$ $Z=\frac{Y}{X}$ $Z=XY$ 连续

$f_{Y/X}(z) = \int_{-\infty}^{\infty} |x| f(x, zx)\,dx$

$f_{XY}(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f(x, \frac{z}{x})\,dx$

Summary 总结

Title 随机变量数字特征

Date /

Keywords 关键词	Notes 笔记	Review 复习记录 / / / / /

- 数学期望
- 方差
- 协方差.相关系数

§. 数学期望

▸ [Def] 对离散性随机变量 X　$P\{X=x_k\}=p_k$　$k=1,2\cdots$

if $\sum_{k=1}^{\infty} x_k p_k$ 绝对收敛　then 称 $\sum_{k=1}^{\infty} x_k p_k$ 为 X 的期望(均值)，记为 $E(X)$

[Remark] 绝对收敛　$\sum_{k=1}^{\infty}|u_k|$ 收敛　(本书P116)

对连续性 X　$f(x)$

if $\int_{-\infty}^{\infty} x f(x)$ 绝对收敛　then 称之为 X 的期望

▸ [Theorem] $Y=g(X)$　g 连续

(i) X 离散　$E(Y)=E[g(X)]=\sum_{k=1}^{\infty} g(x_k)p_k$

(ii) X 连续　$E(Y)=E[g(X)]=\int_{-\infty}^{\infty} g(x)f(x)\,dx$

▸ [Theorem] $Z=g(X,Y)$　g 连续

(i) X,Y 离散　$P\{X=x_i,Y=y_j\}=p_{ij}$　$E(Z)=\sum_i\sum_j g(x_i,y_j)p_{ij}$

(ii) X,Y 连续　$f(x,y)$　$E(Z)=E[g(X,Y)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)\,dx\,dy$

[Prop] (i) $E(C)=C$　(ii) $E(cX)=cE(X)$　(iii) $E(\alpha X+\beta Y)=\alpha E(X)+\beta E(Y)$

(iv) $E(XY) \overset{X,Y独立}{=\!=\!=} E(X)\cdot E(Y)$

§. 方差

▸ [Def] X 随机变量 if $E\{[X-E(X)]^2\}$ 存在　then 称为 X 的方差　记为 $D(X)$

$\sqrt{D(X)}$ 为标准差(均方差)

$D(X)=\sum_{k=1}^{\infty}[x_k-E(X)]^2 p_k$　//　$\int_{-\infty}^{\infty}[x-E(X)]^2 f(x)\,dx$

$D(X)=E(X^2)-[E(X)]^2$

[Prop] (i) $D(C)=0$　(ii) $D(cX)=c^2D(X)$　$D(c+X)=D(X)$

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

$D(X+Y) \overset{X,Y独立}{=\!=\!=} D(X)+D(Y)$

(iv) $D(X)=0 \iff P\{X=E(X)\}=1$

§. 重要分布的 E与D

[0-1分布]　$E(X)=p$

$D(X)=p-p^2$

[$X\sim b(n,p)$]　$E(X)=np$

$D(X)=np(1-p)$

[$X\sim \pi(\lambda)$ / $P\{X=k\}=\frac{\lambda^k e^{-\lambda}}{k!}$]　$E(X)=\lambda$

$D(X)=\lambda$

[$X\sim U(a,b)$]　$E(X)=\frac{a+b}{2}$

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

[指数分布 / $f(x)=\frac{1}{\theta}e^{-\frac{x}{\theta}}$]　$E(X)=\theta$

$D(X)=\theta^2$

Content 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

Keywords 关键词	Notes 笔记

[$X \sim N(u, \sigma^2)$ / $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-u)^2}{2\sigma^2}}$]

对 $Z \sim N(0,1)$ $E(Z) = 0$ $D(Z) = 1$ $X \sim N(u, \sigma^2)$ $E(X) = u$ $D(X) = \sigma^2$

对 $X_i \sim N(u_i, \sigma_i^2)$ $i = 1, 2, 3 \cdots n$ 有限.独立 $Y = C_1X_1 + C_2X_2 + \cdots + C_nX_n$

$E(Y) = \sum_i C_i u_i$ $D(Y) = \sum_i C_i^2 \sigma_i^2$

▶[Theorem] 切比雪夫不等式: X $E(X) = u$ $D(X) = \sigma^2$

$\forall \varepsilon > 0$ $P\{|X-u| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}$

§. 协方差与相关系数

▶[Def] 协方差 $Cov(X,Y) = E\{[X-E(X)][Y-E(Y)]\}$

[Prop] $Cov(X,Y) = E(XY) - E(X)E(Y)$

$Cov(X,X) = D(X)$

$D(X+Y) = D(X) + D(Y) + 2Cov(X,Y)$

[Prop] (i) $Cov(aX, bY) = ab\,Cov(X,Y)$

(ii) $Cov(X_1+X_2, Y) = Cov(X_1,Y) + Cov(X_2,Y)$

▶[Def] 相关系数 $\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{D(X)D(Y)}}$

$X^* = \frac{X-E(X)}{\sqrt{D(X)}}$ $Y^* = \frac{Y-E(Y)}{\sqrt{D(Y)}}$ $\rho_{XY} = Cov(X^*, Y^*)$

[Prop] (i) X, Y 独立 $\rho_{XY} = 0$ (ii) $|\rho_{XY}| \leq 1$ 且 $\rho_{XY} = 1 \Leftrightarrow \exists a > 0$ s.t. $Y = aX + b$

(iii) $|\rho_{XY}|$ 越大 X Y 相关性越强

$\rho_{XY} = -1 \Leftrightarrow \exists a < 0$ s.t. $Y = aX + b$

$\rho_{XY} = 0 \Leftrightarrow X$ Y 不相关 ($\neq$ 独立)

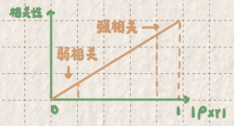
ρ_{XY} 的本质 $\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$

考虑 $a + bX$ 近似拟合 Y $e = E\{[Y-(a+bX)]^2\} = E(Y^2) + b^2E(X^2) + a^2 - 2bE(XY) + 2abE(X)$
$- 2aE(Y)$

$\frac{\partial e}{\partial a} = 2a + 2bE(X) - 2E(Y) = 0$

$\frac{\partial e}{\partial b} = 2bE(X^2) - 2E(XY) + 2aE(X) = 0$

$\Rightarrow \min_{a,b} E\{[Y-(aX+b)]^2\} = (1-\rho_{XY}^2)D(Y)$



[Remark] 相关性只描述随机变量间的线性关系

[Remark] 二维正态分布 $(X,Y) \sim N(u_1, \sigma_1^2, u_2, \sigma_2^2, \rho)$

$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\{\frac{1}{2\sqrt{1-\rho^2}}[\frac{(x-u_1)^2}{\sigma_1^2} - 2\rho\frac{(x-u_1)(x-u_2)}{\sigma_1\sigma_2} + \frac{(x-u_2)^2}{\sigma_2^2}]\}$

$\begin{cases} Cov(X,Y) = \rho\sigma_1\sigma_2 \\ \rho_{XY} = \rho \end{cases}$

Summary 总结

Chapter V

Title 大数定律与中心极限

Date /

Keywords 关键词

- 大数定律
- 中心极限定律

Notes 笔记

Review 复习记录 / / / / /

▸【Prop】两种收敛

(i) 依概率收敛

$X_1 . X_2 \cdots X_n$ a 常数 if 对 $\forall \varepsilon > 0$ 有 $\lim_{n\to\infty} P\{|X_n - a| < \varepsilon\} = 1$

then 称序列 $X_1 \sim X_n$ 依概率收敛于 a 记为 $X_n \xrightarrow{P} a$

(ii) 依分布收敛

$X_1\ X_2 \cdots X_n$ $F_n(x)$ $F(x)$ if 在 $F(x)$ 连续点处都有 $\lim_{n\to\infty} F_n(x) = F(x)$

then 称 $X_1 \sim X_n$ 依分布收敛于 $F(x)$ 记为 $X_n \xrightarrow{L} F(x)$

§. 大数定律

▸【Theorem】弱大数定律 / 辛钦大数定律

$X_1 - X_n$ 独立 服从同一分布随机变量序列 $E(X_k) = \mu$

$\forall \varepsilon > 0$ 有 $\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu| < \varepsilon\} = 1$

定义 $\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$ $\lim_{n\to\infty} P\{|\bar{X} - \mu| < \varepsilon\} = 1$ $\bar{X} \xrightarrow{P} \mu$

【Theorem】伯努利大数定理

f_A: n次独立重复试验中 A 发生的次数（频数） p.

对 $\forall \varepsilon > 0$ 有 $\lim_{n\to\infty} P\{|\frac{f_A}{n} - p| < \varepsilon\} = 1$ / $\lim_{n\to\infty} P\{|\frac{f_A}{n} - p| \geq \varepsilon\} = 0$

§. 中心极限定理

▸【Theorem】对于相互独立随机变量 随变量数增加 它们的和收敛于正态分布

独立同分布 $X_1 \sim X_n$ 独立 同分布 $E(X_k) = \mu$ $D(X_k) = \sigma^2$

定义 $Y_n = \frac{\sum X_k - E(\sum X_k)}{\sqrt{D(\sum X_k)}} = \frac{\sum X_k - n\mu}{\sqrt{n}\,\sigma}$

Y_n $F_n(y)$ 对 $\forall y$ 有 $\lim_{n\to\infty} F_n(y) = \lim_{n\to\infty} P\{Y_n \leq y\} = \phi(y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$

近似地 $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$ $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

独立不同分布（李雅普诺夫定理） $X_1 \sim X_n$ 独立 $E(X_k) = \mu_k$ $D(X_k) = \sigma_k^2$

定义 $B_n^2 = \sum \sigma_k^2$ if $\exists\, \delta > 0$ s.t. $n \to \infty$ 时 $\frac{1}{B_n^{2+\delta}} \sum_{k=1}^{n} E\{|X_k - \mu_k|^{2+\delta}\} \to 0$

then $Z_n = \frac{\sum X_k - E(\sum X_k)}{\sqrt{D(\sum X_k)}} = \frac{\sum X_k - \sum \mu_k}{B_n}$

Z_n $F_n(z)$ 对 $\forall z$ 有 $\lim_{n\to\infty} F_n(z) = \lim_{n\to\infty} P\{\frac{\sum X_k - \sum \mu_k}{B_n} \leq z\} = \phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$

近似地 $\sum_{k=1}^{n} X_k = B_n Z_n + \sum_{k=1}^{n} \mu_k \sim N(\sum \mu_k, B_n^2)$

棣莫弗-拉普拉斯定理 η_n 服从 n.p 二项分布

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

Summary 总结

Title 概率论总结

Date /

Keywords 关键词 | Notes 笔记 | Review 复习记录 / / / / /

- 集合论
- 随机变量
- 二维随机变量

§. 集合论

▸[Prop] $A\cup(B\cap C)=(A\cup B)\cap(A\cup C)$ $\quad A\cap(B\cup C)=(A\cap B)\cup(A\cap C)$

$\overline{A\cup B}=\bar{A}\cap\bar{B}$ $\quad \overline{A\cap B}=\bar{A}\cup\bar{B}$

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

▸[Prop] $P(B|A)=\dfrac{P(AB)}{P(A)}$

全概率公式 $B_1\sim B_n$ $\quad P(A)=P(A|B_1)P(B_1)+P(A|B_2)P(B_2)+\cdots+P(A|B_n)P(B_n)$

贝叶斯公式 $P(B_i|A)=\dfrac{P(A|B_i)P(B_i)}{P(A)}$ 逆因公式

▸[Def] 独立性 $P(AB)=P(A)P(B)$

§. 随机变量

离散型随机变量

$p_k=P\{X=x_k\}$ $\quad F(x)=P\{X\le x\}$

分布函数性质 (i) $F(x)$ 非减 (ii) $F(-\infty)=0$ $\quad F(\infty)=1$

(iii) $F(x)$ 右连续

▸[0-1分布] $P\{X=1\}=p$ $\quad P\{X=0\}=1-p$ $\quad$ or $\quad p(x)=p^x(1-p)^{1-x}$ $\quad x=0,1$

$E(X)=p$ $\quad D(X)=p(1-p)$

▸[二项分布] $P\{X=k\}=C_n^k p^k(1-p)^{n-k}$ $\quad$ 记号 $X\sim b(n)$

$E(X)=np$ $\quad D(X)=np(1-p)$

▸[泊松分布] $P\{X=k\}=\dfrac{\lambda^k e^{-\lambda}}{k!}$ $\quad$ 记号 $X\sim\pi(\lambda)$ $\quad (\lambda=np)$

$E(X)=\lambda$ $\quad D(X)=\lambda$

连续型随机变量

$F(x)=\int_{-\infty}^{x}f(t)dt$ $\quad f(x)$: 概率密度

▸[均匀分布] $f(x)=\begin{cases}\frac{1}{b-a} & a<x<b\\ 0 & \text{other}\end{cases}$ $\quad F(x)=\dfrac{x-a}{b-a}$ $\quad a<x<b$ $\quad$ 记号 $X\sim U(a,b)$

$E(X)=\dfrac{a+b}{2}$ $\quad D(X)=\dfrac{(b-a)^3}{12}$

▸[指数分布] $f(x)=\begin{cases}\frac{1}{\theta}e^{-\frac{x}{\theta}} & x>0\\ 0 & x\le 0\end{cases}$ $\quad F(x)=1-e^{-\frac{x}{\theta}}$ $\quad x>0$

$E(X)=\theta$ $\quad D(X)=\theta^2$

▸[正态分布] $f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $\quad$ 记号 $X\sim N(\mu,\sigma^2)$

$E(X)=\mu$ $\quad D(X)=\sigma^2$

标准化 $X\sim N(\mu,\sigma^2)$ 则 $Z=\dfrac{X-\mu}{\sigma}\sim N(1,0)$

对独立 $X_i\sim N(\mu_i,\sigma_i^2)$ $\quad \sum_{i=1}^{n}X_i\sim N(\sum_i\mu_i,\sum_i\sigma_i^2)$

▸[Prop] 对于 $X: f_X(x)$ $\quad Y=g(X)$ $\quad g$ 处处可导且单调

则 $f_Y=f_X[h(y)]\,|h'(y)|$ $\quad \alpha<y<\beta$ $\quad$ 其中 $h(y)=g^{-1}(x)$

例如 $V=A\sin\Theta$ $\quad \Theta\sim U(-\frac{\pi}{2},\frac{\pi}{2})$ $\quad V$?

$v=g(\theta)=A\sin\theta$ $\quad h(v)=\theta=\arcsin\frac{\theta}{A}$

$f_V(v)=\dfrac{1}{\pi}\dfrac{1}{\sqrt{A^2-v^2}}$ $\quad v\in(-A,A)$

Keywords 关键词	Notes 笔记

§ 二维随机变量

▸ [Def] 联合分布函数 $F(x,y) = P\{(X \le x) \cap (Y \le y)\}$

(i) $F(x,y)$ 关于 x, y 右连续

(ii) $P\{x_1 < x < x_2,\ y_1 < y < y_2\} = F(x_2,y_2) - F(x_1,y_2) - F(x_2,y_1) + F(x_1,y_1)$

$f(x,y)$ 概率密度 (点密度)

$P\{(X,Y) \in G\} = \iint_G f(u,v)\,du\,dv \qquad f(x,y) = \frac{\partial^2 F}{\partial x \partial y}$

▸ [Def] 边缘分布 $F_X(x) = F(x,\infty) \qquad F_Y(y) = F(\infty,y)$

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

[Remark] 二维正态分布 X Y

记号 $(X,Y) \sim N(u_1, u_2, 6_1^2, 6_2^2, \rho)$

$$\begin{cases} f(x,y) = \frac{1}{2\pi 6_1 6_2 \sqrt{1-\rho^2}} e^{\frac{-1}{2(1-\rho^2)}\left[\frac{(x-u_1)^2}{6_1^2} - 2\rho\frac{(x-u_1)(y-u_2)}{6_1 6_2} + \frac{(y-u_2)^2}{6_2^2}\right]} \\ f_X(x) = \frac{1}{\sqrt{2\pi}6_1} e^{-\frac{(x-u_1)^2}{26_1^2}} \\ f_Y(y) = \frac{1}{\sqrt{2\pi}6_2} e^{-\frac{(y-u_2)^2}{26_2^2}} \end{cases}$$

▸ [Def] 独立性 $F(x,y) = F_X(x) F_Y(y) \quad // \quad f(x,y) = f_X(x) f_Y(y)$

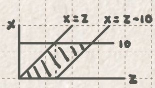
▸ [$Z = X+Y$] $(X, Y) \quad f(x,y) \quad Z = X+Y$

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

$\Vdash\!\!\to$ 当 X Y 独立 $f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\,dy$

例如 $R = R_1 + R_2 \qquad f_{R_1/2}(x) = \frac{10-x}{50} \qquad 0 \le x \le 10$

$f_R(z) = \int_{-\infty}^{+\infty} f(x) f(z-x)\,dx$ 考虑 $\begin{cases} 0 \le x \le 10 \\ 0 \le z-x \le 10 \end{cases} \mapsto \begin{cases} 0 \le x \le 10 \\ z-10 \le x \le z \end{cases}$



$$f_R(z) = \begin{cases} \int_0^z f(x) f(z-x)\,dx & 0 \le z < 10 \\ \int_{z-10}^{10} f(x) f(z-x)\,dx & 10 \le z < 20 \\ 0 & \text{other} \end{cases}$$

▸ [$Z = \frac{Y}{X}$, $Z = XY$]

$f_{Y/X}(z) = \int_{-\infty}^{+\infty} |x| f(x,xz)\,dx \qquad f_{Y/X}(z) = \int_{-\infty}^{\infty} |x| f_X(x) f_Y(xz)\,dx$

$\Vdash\!\!\to$ X, Y 独立

$f_{XY}(z) = \int_{-\infty}^{+\infty} \frac{1}{|x|} f(x, \frac{z}{x})\,dx \qquad f_{XY}(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_X(x) f_Y(\frac{z}{x})\,dx$

▸ [$M = \max\{X,Y\}$, $N = \min\{X,Y\}$]

X, Y 独立 $F_{max}(z) = F_X(z) F_Y(z)$

$F_{min}(z) = 1 - [1-F_X(z)][1-F_Y(z)]$

▸ [Def] 条件分布 $P\{X = x_i \mid Y = y_j\} = \frac{p_{ij}}{p_{\cdot j}}$

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

Summary 总结

Title 概率论总结

Date /

Keywords 关键词 | Notes 笔记 | Review 复习记录 / / / / /

- 数字特征
- 衔接部分

§. 数字特征

[Def] $E(X) = \sum_{k=1}^{\infty} x_k p_k$ // $\int_0^{\infty} x f(x)\,dx$

(i) $E(\alpha X+\beta Y+\gamma) = \alpha E(X) + \beta E(Y) + \gamma$

(ii) $Y = g(X)$ g连续 $E(Y) = \sum_k g(x_k) p_k = \int_{-\infty}^{\infty} g(x) f(x)\,dx$

(iii) X Y 独立 $E(XY) = E(X)E(Y)$

[Def] $D(X) = \sum_k [x_k - E(X)]^2 p_k$ // $\int_{-\infty}^{\infty} [x-E(X)]^2 f(x)\,dx$

$= E(X^2) - [E(X)]^2$

(i) $D(\alpha X + \beta) = \alpha^2 D(X)$

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

[Theorem] 切比雪夫不等式 $E(X)=\mu$ $D(X)=\sigma^2$

$\forall \varepsilon > 0$ $P\{|X-\mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}$

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

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

(i) $Cov(aX, bY) = ab\,Cov(X,Y)$

(ii) $Cov(X_1+X_2, Y) = Cov(X_1, Y) + Cov(X_2, Y)$

(iii) $\rho_{XY} = 1 \rightleftharpoons \exists a>0$ s.t. $Y = aX+b$

$\rho_{XY} = -1 \rightleftharpoons \exists a<0$ s.t. $Y = aX+b$

$\rho_{XY} = 0 \rightleftharpoons$ X Y 不相关

(iv) 独立 一定不相关

[Remark] 二维正态分布中 $(X,Y) \sim N(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \rho)$ $\rho = \rho_{XY}$ $Cov(X,Y) = \rho\sigma_1\sigma_2$

§. 衔接部分

[Theorem] 弱大数定律 $X_1 \sim X_n$ 独立 服从同一分布随机序列 $E(X_k) = \mu$

则 $\overline{X} = \frac{1}{n}\sum_{k}^{n} X_k \longrightarrow \mu$

[Theorem] 伯努利大数定律 f_A 频数 p $\frac{f_A}{n} \longrightarrow p$

[Theorem] 中心极限定理

独立同分布 $X_1 \sim X_n$ $E(X_k) = \mu$ $D(X_k) = \sigma^2$

$\overline{X} = \frac{1}{n}\sum_{k}^{n} X_k \sim N(\mu, \frac{\sigma^2}{n})$ $n \to \infty$

独立不同分布 $X_1 \sim X_n$ $E(X_k) = \mu_k$ $D(X_k) = \sigma_k^2$

$\sum_{k}^{n} X_k \sim N(\sum \mu_k, \sum \sigma_k^2)$ $n \to \infty$

Summary 总结

13

Content
内容

【数理统计】

Review
复习记录

/ / / / /

Content 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

Title 数理统计导论 与参数估计

Date /

Keywords 关键词 | Notes 笔记 | Review 复习记录 / / / / /

- 相关概念
- 抽样分布
- 参数估计

[Def] 总体 所研究问题有关对象的全体所构成的集合

样本 从总体中抽取一部分个体

简单随机样本 代表性 + 独立性

[Theorem] 样本联合分布 $F(x_1, x_2 \cdots x_n) = F(x_1)F(x_2)\cdots F(x_n) = \prod_{i=1}^{n} F(x_i)$

§. 概念

[Def] 统计量 完全由样本决定的量 $g(x_1\ x_2 \cdots x_3)$

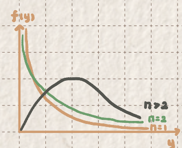
$\Rightarrow$ 平均值 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 方差 $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right)$

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

§. 抽样分布

[Def] 抽样分布 统计量的分布

[χ^2分布]



$X \sim N(0,1)$ $X_1 \sim X_n$ 样本

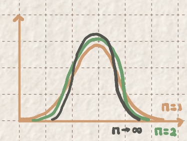
$Y = X_1^2 + X_2^2 + \cdots + X_n^2$ 服从自由度为 n 的 χ^2 分布 $Y \sim \chi^2(n)$

$\Rightarrow$ if $Y \sim \chi^2(n)$ $E(Y) = n$ $D(Y) = 2n$

if $Y_1 \sim \chi^2(n_1)$ $Y_2 \sim \chi^2(n_2)$ $Y_1\ Y_2$ 相互独立 $Y_1 + Y_2 \sim \chi^2(n_1 + n_2)$

上分位点 $\{P\{\chi^2(n) > \chi^2_\alpha(n)\} = \alpha$

[t分布]



$X \sim N(0,1)$ $Y \sim \chi^2(n)$ X.Y 相互独立

$T = \frac{X}{\sqrt{Y/n}}$ 服从自由度为 n 的 t 分布 $T \sim t(n)$

[F分布]

$U \sim \chi^2(n_1)$ $V \sim \chi^2(n_2)$ U V 相互独立

$F = \frac{U/n_1}{V/n_2}$ 服从自由度为 (n_1, n_2) 的 F 分布 $F \sim F(n_1, n_2)$

$\Rightarrow$ if $F \sim F(n_1, n_2)$ $\frac{1}{F} \sim F(n_2, n_1)$

$T \sim t(n)$ $T^2 \sim F(1, n)$

上 $1-\alpha$ 分位点 $F_{1-\alpha}(n_1, n_2) = \frac{1}{F_\alpha(n_1, n_2)}$

[Theorem] $X \sim N(\mu, \sigma^2)$ 样本 $\bar{X}$. S^2 则有

$E(\bar{X}) = \mu$ $D(\bar{X}) = \sigma^2/n$ $E(S^2) = \sigma^2$

由此推出

(i) $\bar{X} \sim N(\mu, \sigma^2/n)$ (ii) $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

(iii) $\bar{X}$ 和 S 相互独立 (iv) $\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$

[Theorem] $X_1 \sim X_{n_1}$ $Y_1 \sim Y_{n_2}$ 分别来自总体 $X \sim N(\mu_1, \sigma_1^2)$ $Y \sim N(\mu_2, \sigma_2^2)$

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

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

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

Content 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

Keywords 关键词 | Notes 笔记

§. 参数估计

点估计

设总体 X 分布函数 $F(x;\theta)$　θ: 待测参数

构造一个适当 $\hat{\theta}(X_1, X_2 \cdots X_n)$ 用其观测值 $\hat{\theta}(x_1, x_2 \cdots x_n)$ 近似取代 θ

[矩估计法]

X 连续性　总体 $f(x;\theta_1\theta_2\cdots\theta_n)$　$X_1, X_2 \cdots X_n$ 为 X 的样本

if ∃ X 前n阶矩　$\mu_l = E(X^l) = \int_{-\infty}^{\infty} x^l f(x;\theta_1\sim\theta_n)\,dx$

then　$A_l = \sum_{i=1}^{n} X_i^l \xrightarrow{P} \mu_l$

[1] k 个未知参数 ↦ k个方程 ⟼ [2] 反解 $\theta_1\sim\theta_k$ ⟼ [3] 以 A_i 代替 u_i

$$\begin{cases} u_1 = u_1(\theta_1\cdots\theta_k) \\ u_2 = u_2(\theta_1\cdots\theta_k) \\ \cdots\cdots \\ u_k = u_k(\theta_1\cdots\theta_k) \end{cases} \qquad \begin{cases} \theta_1 = \theta_1(u_1\cdots u_k) \\ \theta_2 = \theta_2(u_1\cdots u_k) \\ \cdots\cdots \\ \theta_k = \theta_k(u_1\cdots u_k) \end{cases} \qquad \hat{\theta}_i = \theta_i(A_1\cdots A_k)$$

[最大似然估计]

构建似然函数 $L(\theta) = L(x_1\cdots x_n;\theta) = \prod_{i=1}^{n} f(x_i;\theta)$

$L(\hat{\theta}) = \max L(\theta)$　$\hat{\theta}$ 为 θ 最大似然估计值　// $\frac{\partial}{\partial\theta_i}\ln L(\theta) = 0$

[1] 写出分布表达 $p(x_i)$ / $f(x_i)$　[2] 列出似然函数 L / $\ln L$　[3] 对 θ_i 求导为0

▸ **[Prop]** 无偏性、有效性、相合性

(i) 无偏性　$E(\hat{\theta}) = \theta$　衡量系统误差

(ii) 有效性　对两种无偏估计　$\hat{\theta}_1$　$\hat{\theta}_2$　$D(\hat{\theta}_1) < D(\hat{\theta}_2)$　则称 θ_1 比 θ_2 更有效

(iii) 相合性　$\hat{\theta}$　$n \to \infty$ 时　$\hat{\theta} \xrightarrow{P} \theta$

Summary 总结

Keywords 关键词 | Notes 笔记 | Review 复习记录 / / / / /

- 区间估计
- 正态分布
- (0-1)分布
- 单侧置信区间

//区间估计

【Def】置信区间 $X \quad F(x;\theta)$ 含一个未知参数 $\theta\in\Theta$

对给定值 $\alpha\in(0,1)$ if $X_1\sim X_n$ 确定的 2个统计量 $\underline{\theta}=\underline{\theta}(X_1,X_2\cdots X_n)$ $\bar{\theta}=\bar{\theta}(X_1\cdots X_n)$

对 $\forall\,\theta\in\Theta$ 满足 $P\{\underline{\theta}(X_1,X_2\cdots X_n)<\theta<\bar{\theta}(X_1,X_2\cdots X_n)\}\geq 1-\alpha$

则称 $(\underline{\theta},\bar{\theta})$ 为 θ 置信水平为 $1-\alpha$ 的置信区间

§.区间估计

【枢轴变量法】

(i) 寻求样本 $X_1\cdots X_n$ 和 θ 的函数 $W=W(X_1\sim X_n,\theta)$ s.t. W 分布不依赖 θ 及其它参数

则称 W 为枢轴量

(ii) 定出 a,b s.t. $P\{a<W(X_1\sim X_n,\theta)<b\}=1-\alpha$

if $a<W<b \longmapsto \underline{\theta}<\theta<\bar{\theta}$ then $(\underline{\theta},\bar{\theta})$ 为 θ 的一个置信水平为 $1-\alpha$ 的置信区间

【正态总体】

(i) 单个总体 $N(\mu,\sigma^2)$

$1-\alpha$ $X_1\sim X_n$ 为总体 $N(\mu,\sigma^2)$ 的样本 $\bar{X}$ S^2：样本均值 / 方差

[1] 关于 μ

(1) σ^2 已知 $W=\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ $1-\alpha$ 置信水平区间 $(\bar{X}\pm\dfrac{\sigma}{\sqrt{n}}z_{\alpha/2})$ 正态分布上分位点

(2) σ^2 未知 令 $\sigma=S$ $W=\dfrac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$

$1-\alpha$ 置信水平区间 $(\bar{X}\pm\dfrac{S}{\sqrt{n}}t_{\alpha/2}(n-1))$ t分布上分位点

[2] 关于 σ^2

$W=\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$ $1-\alpha$ 置信水平区间 $[\dfrac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)},\dfrac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}]$

(ii) 两个总体 $N(\mu_1,\sigma_1^2)$ $N(\mu_2,\sigma_2^2)$

$1-\alpha$ $X_1\sim X_{n_1}$ $Y_1\sim Y_{n_2}$ 相互独立 $\bar{X}$ $\bar{Y}$ S_1^2 S_2^2

[1] 关于 $\mu_1-\mu_2$

(1) σ^2 已知 $W=\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1)\longmapsto(\bar{X}-\bar{Y}\pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}})$

(2) σ^2 未知但 $\sigma_1^2=\sigma_2^2$ $\longmapsto(\bar{X}-\bar{Y}\pm t_{\alpha/2}(n_1+n_2-2)S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}})$

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

[2] 关于 σ_1^2/σ_2^2

$W=\dfrac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}\sim F(n_1-1,n_2-1)\longmapsto[\dfrac{S_1^2}{S_2^2}\dfrac{1}{F_{\alpha/2}(n_1-1,n_2-1)},\dfrac{S_1^2}{S_2^2}\dfrac{1}{F_{1-\alpha/2}(n_1-1,n_2-1)}]$

【0-1分布】

$f(x;p)=p^x(1-p)^{1-x}$ $x=0,1$

$\mu=p$ $\sigma^2=p(1-p)$

$W=\dfrac{n\bar{X}-np}{\sqrt{np(1-p)}}\sim N(0,1)\longmapsto(p_1,p_2)$ 其中 $\begin{cases}p_1=\frac{1}{2a}(-b-\sqrt{b^2-4ac})\\p_2=\frac{1}{2a}(-b+\sqrt{b^2-4ac})\end{cases}$ $\begin{cases}a=n+z^2_{\alpha/2}\\b=-(2n\bar{X}+z^2_{\alpha/2})\\c=n\bar{X}^2\end{cases}$

§.单侧置信区间

【Def】 $0<\alpha<1$ 由 $X_1\sim X_n$ 确定的 $\underline{\theta}=\underline{\theta}(X_1\cdots X_n)$

对 $\forall\,\theta\in\Theta$ 满足 $P\{\theta>\underline{\theta}\}\geq 1-\alpha$

则称 $(\underline{\theta},\infty)$ 为 θ 置信水平为 $1-\alpha$ 的单侧置信区间 $\underline{\theta}$：单侧置信区间

上限同理

Content 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

Keywords 关键词	Notes 笔记

【正态分布】

$X \sim N(\mu, \sigma^2)$ μ, σ^2 未知 $X_1 \sim X_n$

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

$$\mapsto \left(\bar{X} - \frac{S}{\sqrt{n}} t_\alpha(n-1), \infty\right)$$

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

$$\mapsto \left(0, \frac{(n-1)S^2}{\chi^2_{1-\alpha}(n-1)}\right)$$

Summary 总结

Title 假设检验

Date /

Keywords 关键词 | Notes 笔记 | Review 复习记录 / / / / /

• 正态分布假设检验

▸ [Def] 假设 H_0 数 k 作为检验假设的门槛值 $Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

if $|Z| \geq k$ 则称 $\bar{X}$ 与 μ_0 差异显著 $\longmapsto$ 拒绝 H_0

反之接受 H_0 $k = Z_{\alpha/2}$ α：显著性水平 Z：检验统计量

[Remark] H_0 原假设 对应 H_1 备择假设

§.μ假设检验

▸ [正态总体均值的假设检验]

(i) 单个总体 $N(\mu, \sigma^2)$ 均值 μ 检验

σ^2 已知 Z检验法 令 $Z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

对确定的 α 比较 Z 与 $Z_{\alpha/2}$ $|Z| \geq Z_{\alpha/2}$ 则拒绝

σ^2 未知 t检验法 $H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$ $X_1 \sim X_n$ 样本 α

令 $t = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}$ 比较 t 与 $t_{\alpha/2}(n-1)$ $|t| \geq t_{\alpha/2}(n-1)$ 则拒绝

(ii) 两个正态总体均值差的检验

$X_1 \sim X_{n_1}$ 来自总体 $N(\mu_1, \sigma^2)$ $Y_1 \sim Y_{n_2}$ 来自 $N(\mu_2, \sigma^2)$ $\bar{X}$ $\bar{Y}$ S_1^2 S_2^2

$H_0: \mu_1 - \mu_2 = \delta$ $H_1: \mu_1 - \mu_2 \neq \delta$

令 $t = \dfrac{(\bar{X} - \bar{Y}) - \delta}{S_w\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ $S_w^2 = \dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}$

比较 t 与 $t_{\alpha/2}(n_1 + n_2 - 2)$ $|t| \geq t_{\alpha/2}(n_1 + n_2 - 2)$ 则拒绝

§.σ²假设检验

▸ [正态总体分布方差的假设检验]

(i) 单个总体

$X \sim N(\mu, \sigma^2)$ μ, σ^2 未知 $X_1 \sim X_n$ 来自 X 样本 $H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 \neq \sigma_0^2$

令 $\chi^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$ 比较 χ^2 与 $\chi^2_{1-\alpha/2}(n-1)$ $\chi^2_{\alpha/2}(n-1)$

拒绝域 $\chi^2 \leq \chi^2_{1-\alpha/2}(n-1) \cup \chi^2 \geq \chi^2_{\alpha/2}(n-1)$

▸ [Remark] 单边检验：$H_0: \sigma^2 \leq \sigma_0^2 \longmapsto$ 拒绝域 $\chi^2 \geq \chi^2_{\alpha}(n-1)$

$H_0: \sigma^2 \geq \sigma_0^2 \longmapsto \chi^2 \leq \chi^2_{1-\alpha}(n-1)$

(ii) 两个总体

$X_1 \sim X_{n_1}$ $N(\mu_1, \sigma_1^2)$ $Y_1 \sim Y_{n_2}$ $N(\mu_2, \sigma_2^2)$ S_1^2 S_2^2

$H_0: \sigma_1^2 \leq \sigma_2^2$ $H_1: \sigma_1^2 > \sigma_2^2$

令 $F = \dfrac{S_1^2}{S_2^2}$ 比较 F 与 $F_\alpha(n_1-1, n_2-1)$

拒绝域 $F \geq F_\alpha(n_1-1, n_2-1)$

Summary 总结

Chapter Ⅲ

Title 分布拟合检验 与 p值法检验

Date /

Keywords 关键词 | Notes 笔记 | Review 复习记录 / / / / /

- 分布拟合检验
- p值检验

§. 分布拟合检验

▸【单个分布的 χ^2 拟合检验法】

H_0：总体 X 分布函数 $F(x)$

H_0 为真时，X 全部可能取值 Ω 分为互无交集的子集 $A_1 \sim A_k$

f_i 为 n 次独立试验 x_i 落入 A_i 的次数

则建立统计量 $\sum_{i=1}^{k} C_i \left(\frac{f_i}{n} - p_i\right)^2$

▸【Theorem】当 n 充分大时 ($n \geq 50$, $np_i > 5$) H_0 为真时取 $C_i = n/p_i$

统计量 $\sum_{i=1}^{k} \frac{n}{p_i}\left(\frac{f_i}{n} - p_i\right)^2 = \sum_{i=1}^{k} \frac{f_i^2}{np_i} - n = \chi^2$

【Remark】拒绝域 $\chi^2 \geq \chi^2_\alpha(K-1)$

▸【分布族 χ^2 拟合检验】

H_0：总体 X 分布函数为 $F(x;\theta_1 \cdots \theta_r)$

H_0 为真时，X 全部可能取值 Ω 分为互无交集的子集 $A_1 \sim A_k$ $(k > r+1)$

用最大似然估计 求出 $\hat{\theta} \longmapsto \hat{p}_i = P(A_i)$

建立统计量 $\chi^2 = \sum_{i=1}^{k} \frac{f_i^2}{n\hat{p}_i} - n \sim \chi^2(k-r-1)$

【Remark】拒绝域 $\chi^2 \geq \chi^2_\alpha(k-r-1)$

§. p值法检验

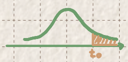
▸【Def】假设检验问题的 p 值是由检验统计量的样本观察值得出的原假设可以被拒绝的最小显著性水平

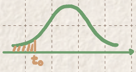
Ex：σ 未知 $N(\mu,\sigma^2)$ $t \sim t(n-1)$

$H_0: \mu \leq \mu_0^2$ $p = P_{\mu_0}\{t \geq t_0\} = t_0$ 右侧尾部面积

$H_0: \mu \geq \mu_0^2$ $p = P_{\mu_0}\{t \leq t_0\} = t_0$ 左侧尾部面积

$H_0: \mu = \mu_0^2$ $p = 2 \times$ (t_0 界定的尾部面积)





【Theorem】$p \leq \alpha$ 拒绝 H_0；$p > \alpha$ 接受 H_0

Summary 总结

Content
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

Title 方差分析 Date /

Keywords 关键词	Notes 笔记 Review 复习记录 / / / / /

- 单因素
- 双因素

§. 单因素方差分析

[Def] 试验指标：考察的指标　因素：影响试验指标的条件

[Prop] Assume 各个水平 A_j 下样本：$X_{1j}\ X_{2j} \cdots X_{n_j j}$ 具有相同方差 σ^2　均值 μ_j

μ_j σ^2 未知　A间相互独立

$X_{ij}-\mu_j \sim N(0,\sigma^2)$　记 $X_{ij}-\mu_j=\varepsilon_{ij}$

$$\left.\begin{array}{l} X_{ij}=\mu_j+\varepsilon_{ij} \\ \varepsilon_{ij}\sim N(0,\sigma^2) \quad \varepsilon_{ij}\text{ 相互独立} \\ i=1,2\cdots n_j \quad j=1,2\cdots s \end{array}\right\} \ (*)\ \text{单因素试验方差分析数学模型}$$

方差分析

[Step1] 检验 s 个总体 $N(\mu_j,\sigma^2)$ 均值是否相等

$H_0: \mu_1=\mu_2=\cdots=\mu_s$

[Step2] 参数估计

记 $\mu=\frac{1}{n}\sum n_j\mu_j$ 总平均　$\delta_j=\mu_j-\mu$　水平 A_j 的效应

则 (*) 改写为：　H_0 等价于：

$$\left.\begin{array}{l} X_{ij}=\mu+\delta_j+\varepsilon_{ij} \\ \varepsilon_{ij}\sim N(0,\sigma^2) \\ i=\cdots \quad j=\cdots \\ \sum_j n_j\delta_j=0 \end{array}\right\} (**) \qquad \delta_1=\delta_2=\cdots=\delta_s$$

[Def] 总偏差平方和 $S_T=\sum_{j=1}^{s}\sum_{i=1}^{n_j}(X_{ij}-\bar{X})^2=\underbrace{\sum_{j=1}^{s}\sum_{i=1}^{n_j}(X_{ij}-\bar{X}_{\cdot j})^2}_{\text{误差平方和}S_E}+\underbrace{\sum_{j=1}^{s}\sum_{i=1}^{n_j}(\bar{X}_{\cdot j}-\bar{X})^2}_{\text{效应平方和}S_A}$

[Theorem] (i) $\frac{S_E}{\sigma^2}\sim\chi^2(n-s)$　(ii) $E(S_E)=(n-s)\sigma^2$

(iii) $E(S_A)=(s-1)\sigma^2+\sum_{j=1}^{s}n_i\delta_j^2$

(iv) if H_0 为真 then $\frac{S_A}{\sigma^2}\sim\chi^2(s-1)$

[Step3] H_0 拒绝域　$F=\frac{S_A/(s-1)}{S_E/(n-s)}\geq F_\alpha(s-1,n-s)$

方差来源	平方和	自由度	均方	F比
因素A	S_A	$s-1$	$\bar{S}_A=\frac{S_A}{s-1}$	$F=\frac{\bar{S}_A}{\bar{S}_E}$
误差	S_E	$n-s$	$\bar{S}_E=\frac{S_E}{n-s}$	
总和	S_T	$n-1$		

记 $T_{\cdot j}=\sum_{i=1}^{n_j}X_{ij}$　$j=1,2,\cdots s$　　$T_{\cdot\cdot}=\sum_{j=1}^{s}\sum_{i=1}^{n_j}X_{ij}$

即有　$S_T=\sum_i\sum_j X_{ij}^2-n\bar{X}^2=\sum\sum X_{ij}^2-\frac{T_{\cdot\cdot}^2}{n}$

$S_A=\sum_j n_j\bar{X}_{\cdot j}^2-n\bar{X}^2=\sum_j\frac{T_{\cdot j}^2}{n_j}-\frac{T_{\cdot\cdot}^2}{n}$

$S_E=S_T-S_A$

[Remark] $\hat{\sigma}^2=\frac{S_E}{n-s}$　$\hat{\mu}=\bar{X}$　$\hat{\mu}_j=\bar{X}_{\cdot j}$　$\hat{\delta}_j=\bar{X}_{\cdot j}-\bar{X}$

对 $N(\mu_j,\sigma^2)$ $N(\mu_k,\sigma^2)$　$j\neq k$

$$\frac{(\bar{X}_{\cdot j}-\bar{X}_{\cdot k})-(\mu_j-\mu_k)}{\sqrt{\bar{S}_E(\frac{1}{n_j}+\frac{1}{n_k})}}=\frac{(\bar{X}_{\cdot j}-\bar{X}_{\cdot k})-(\mu_1-\mu_2)}{\sigma\sqrt{1/n_j+1/n_k}}\Big/\sqrt{S_E/(n-s)\sigma^2}\sim t(n-s)$$

则 $\mu_j-\mu_k=\delta_j-\delta_k$ 置信水平为 $1-\alpha$ 的置信区间

$$\left(\bar{X}_{\cdot j}-\bar{X}_{\cdot k}\pm t_{\alpha/2}(n-s)\sqrt{\bar{S}_E\left(\frac{1}{n_j}+\frac{1}{n_k}\right)}\right)$$

Content 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

Keywords 关键词

§. 双因素方差分析

Notes 笔记

等重复实验

两因素 A . B　　A: $A_1 \cdots A_r$　　B: $B_1 \cdots B_s$

对 $(A_i\ B_j)$ 都作 $t \geq 2$ 次试验　　设 $X_{ijk} \sim N(\mu_{ij}, \sigma^2)$　　$k = 1 \cdots t$　$i = 1 \cdots r$　$j = 1 \cdots s$

记 $X_{ijk} = \mu_{ij} + \varepsilon_{ijk}$　// $\varepsilon_{ijk} \sim N(0, \sigma^2)$

$\mu = \frac{1}{rs} \sum_i \sum_j \mu_{ij}$　　$\mu_{i\cdot} = \frac{1}{s} \sum_j \mu_{ij}$　　$\mu_{\cdot j} = \frac{1}{r} \sum_i \mu_{ij}$

$\alpha_i = \mu_{i\cdot} - \mu$　　$\beta_j = \mu_{\cdot j} - \mu$　// $\sum_{i=1} \alpha_i = 0$　　$\sum_{j=1} \beta_j = 0$

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \underbrace{(\mu_{ij} - \mu_{i\cdot} - \mu_{\cdot j} + \mu)}_{\text{记为 } r_{ij}\quad \text{AB交互效应}}$$

// $\sum_i r_{ij} = 0$　　$\sum_j r_{ij} = 0$

$$= \mu + \alpha_i + \beta_j + r_{ij}$$

$\mapsto$
$$\left.\begin{array}{l} X_{ijk} = \mu + \alpha_i + \beta_j + r_{ij} + \varepsilon_{ijk} \\ \varepsilon_{ijk} \sim (0, \sigma^2) \\ \sum_i \alpha_i = 0 \quad \sum_j \beta_j = 0 \quad \sum_i r_{ij} = 0 \quad \sum_j r_{ij} = 0 \end{array}\right| (**)$$

【三个假设】 $H_{01}: \alpha_1 = \cdots = \alpha_r = 0$　　$H_{02}: \beta_1 = \cdots = \beta_s = 0$　　$H_{03}: r_{11} = \cdots = r_{rs}$

▸【Def】
$$S_T = \sum_i \sum_j \sum_k (X_{ijk} - \bar{X})^2 = \underbrace{\sum_i \sum_j \sum_k (X_{ijk} - \bar{X}_{ij\cdot})^2}_{S_E} + \underbrace{st \sum_i (\bar{X}_{i\cdot\cdot} - \bar{X})^2}_{S_A} + \underbrace{rt \sum_j (\bar{X}_{\cdot j} - \bar{X})^2}_{S_B}$$
$$+ t \sum_i \sum_j (\bar{X}_{ij\cdot} - \bar{X}_{i\cdot\cdot} - \bar{X}_{\cdot j\cdot} + \bar{X})^2 \longrightarrow S_{A\times B}$$

方差来源	平方和	自由度	均方	F比
A	S_A	$r-1$	$\bar{S}_A = \frac{S_A}{r-1}$	$F_A = \frac{\bar{S}_A}{\bar{S}_E}$
B	S_B	$s-1$	$\bar{S}_B = \frac{S_B}{s-1}$	$F_B = \frac{\bar{S}_B}{\bar{S}_E}$
交互作用	$S_{A\times B}$	$(r-1)(s-1)$	$\bar{S}_{A\times B} = \frac{S_{A\times B}}{(r-1)(s-1)}$	$F_{A\times B} = \frac{\bar{S}_{A\times B}}{\bar{S}_E}$
误差	S_E	$rs(t-1)$	$\bar{S}_E = \frac{S_E}{rs(t-1)}$	
总和	S_T	$rst-1$		

无重复实验

方差来源	平方和	自由度	均方	F比
A	S_A	$r-1$	$\bar{S}_A = \frac{S_A}{r-1}$	$F_A = \frac{\bar{S}_A}{\bar{S}_E}$
B	S_B	$s-1$	$\bar{S}_B = \frac{S_B}{s-1}$	$F_B = \frac{\bar{S}_B}{\bar{S}_E}$
误差	S_E	$(r-1)(s-1)$	$\bar{S}_E = \frac{S_E}{(r-1)(s-1)}$	
总和	S_T	$rs-1$		

Summary 总结

Content
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24

Title 数理统计总结

Date /

Keywords 关键词 | Notes 笔记 | Review 复习记录 / / / / /

- 抽样分布
- 参数估计
- 假设检设

§. 抽样分布

▸ [Def] 统计量 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ $\quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right)$

▸ [χ^2分布] $X \sim N(0,1)$ $\quad X_1 \sim X_n$ 为样本

$Y = X_1^2 + X_2^2 + \cdots + X_n^2 \sim \chi^2(n)$ 其中 $E(\chi^2) = n$ $\quad D(\chi^2) = 2n$

▸ [t分布] $X \sim N(0,1)$ $\quad Y \sim \chi^2(n)$

$T = \frac{X}{\sqrt{Y/n}} \sim t(n)$

▸ [F分布] $U \sim \chi^2(n_1)$ $\quad V \sim \chi^2(n_2)$

$F = \frac{U/n_1}{V/n_2} \sim F(n_1, n_2)$

▸ [Prop] $X \sim N(\mu, \sigma^2)$

(i) $E(\bar{X}) = \mu$ $\quad D(\bar{X}) = \sigma^2/n$ $\quad E(S^2) = \sigma^2$ // 不依赖于 $X \sim N(\mu, \sigma^2)$

(ii) $\bar{X} \sim N(\mu, \sigma^2/n)$

(iii) $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ $\quad \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

▸ [Prop] $X \sim N(\mu_1, \sigma_1^2)$ $\quad X_1 \sim X_{n_1}$ $\quad Y \sim N(\mu_2, \sigma_2^2)$ $\quad Y_1 \sim Y_{n_2}$

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

当 $\sigma_1 = \sigma_2 = \sigma$ 时 $T = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$ $\quad S_w^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$

§. 参数估计

点估计

[矩估计法] $X \quad f(x;\theta_1,\theta_2,\theta_3,\cdots\theta_n)$ $\quad A_L = \frac{X_1^L + \cdots + X_n^L}{n}$

X的L阶矩 $\mu_L = E(X^L) = \int x^L f(x,\theta)\,dx$

if 有K个未知参数 $\theta_1 \sim \theta_K$

$\longmapsto$ K个不等式 $u_i = \int x^i f(x,\theta)\,dx$ $\longmapsto$ 反解 $\theta_i = \theta_i(u_1\cdots u_K)$ $\longmapsto$ $u_1\cdots u_K$用统计值$A_1\cdots A_K$代替

[最大似然估计法] 已知X服从某分布 $f(x;\theta)$ 构造 $L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$

$L(\hat{\theta}) = \max L(\theta)$

$\longmapsto \frac{\partial L}{\partial \theta} = 0 \;/\; \frac{\partial \ln L}{\partial \theta} = 0$

[Theorem] 参数估计原则

(i) 无偏性 $E(\hat{\theta}) = \theta$ $\quad$ (ii) 有效性 $D(\hat{\theta})$尽可能小 $\quad$ (iii) 相合性 $\hat{\theta} \xrightarrow{P} \theta$

区间估计

[枢轴变量法] 选定W s.t. W不依赖于θ

[正态分布] (i) 单个总体 $N(\mu, \sigma^2)$

已知σ 求μ $\quad W = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

对置信水平 $1-\alpha$ $\quad$ 置信区间 $\left(\bar{X} \pm \frac{\sigma}{\sqrt{n}} Z_{\alpha/2}\right)$

未知σ 求μ $\quad W = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

对置信水平 $1-\alpha$ $\quad$ 置信区间 $\left(\bar{X} \pm \frac{S}{\sqrt{n}} t_{\alpha/2}(n-1)\right)$

已知μ 求σ $\quad W = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

对置信水平 $1-\alpha$ $\quad$ 置信区间 $\left(\frac{\bar{X}-\mu}{Z_{\alpha/2}/\sqrt{n}}, \frac{\bar{X}-\mu}{Z_{1-\alpha/2}/\sqrt{n}}\right)$

未知μ 求σ $\quad W = \frac{(n-1)S}{\sigma^2} \sim \chi^2(n-1)$

对置信水平 $1-\alpha$ $\quad$ 置信区间 $\left(\frac{(n-1)S}{\chi^2_{\alpha/2}(n-1)}, \frac{(n-1)S}{\chi^2_{1-\alpha/2}(n-1)}\right)$

Keywords 关键词 | Notes 笔记

(iii) 两个总体 $N(\mu_1, \sigma_1^2)$ $N(\mu_2, \sigma_2^2)$

σ^2 已知 求 $\mu_1-\mu_2$ $W = \dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1)$

$\mapsto \left(\bar{X}-\bar{Y} \pm Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$

σ^2 未知但 $\sigma_1^2=\sigma_2^2$ 求 $\mu_1-\mu_2$ $W = \dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{\alpha/2}(n_1+n_2-2)$

$\mapsto \left(\bar{X}-\bar{Y} \pm t_{\alpha/2}(n_1+n_2-2)\, S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$

μ 未知 求 σ_1^2/σ_2^2 $W = \dfrac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$

$\mapsto \left(\frac{S_1^2}{S_2^2}\cdot\frac{1}{F_{\alpha/2}(n_1-1,n_2-1)},\ \frac{S_1^2}{S_2^2}\cdot\frac{1}{F_{1-\alpha/2}(n_1-1,n_2-1)}\right)$

§. 假设检验

▸【μ假设检验】

σ^2 已知 Z检验法 $Z = \dfrac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ 对比 z

σ^2 未知 t检验法 $T = \dfrac{\bar{X}-\mu_0}{S/\sqrt{n}}$ 对比 $t(n-1)$

两正态总体分布 $\sigma_1=\sigma_2$ 对 $\mu_1-\mu_2$ t检验法 $T = \dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ 对比 $t(n_1+n_2-2)$

▸【σ^2 假设检验】

χ^2 检验法 $\chi^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$ 对比 $\chi^2(n-1)$

两正态总体分布 比较 σ_1^2 σ_2^2 大小 F检验法 $F = \dfrac{S_1^2}{S_2^2}$ 对比 $F(n_1-1, n_2-1)$

▸【分布拟合检验】

针对 H_0: 总体X服从分布 $F(x)$

X 全部可能取值 Ω 分为 $A_1 \sim A_k$ f_i 为 A_i 频数 p_i 为 $F(x)$ 对应概率

当 $n \ge 50$ $np_i > 5$ 时 $\chi^2 = \sum_{i=1}^{k}\dfrac{f_i^2}{np_i} - n$ 对比 $\chi^2(k-1)$

针对 H_0: 总体X服从分布 $F(x;\theta_1\cdots\theta_r)$

Ω 分为 $A_1 \sim A_k$ $(k > r+1)$ 先用最大似然估计 $\hat{\theta}$ $\mapsto$ 求出 $\hat{p}_i$

$\chi^2 = \sum_{i=1}^{k}\dfrac{f_i^2}{n\hat{p}_i} - n$ 对比 $\chi^2(k-r-1)$

【p值法检验】

▸【Def】p值是由检验统计量的样本观察值得出的原假设可被拒绝的最小显著性水平

$p \le \alpha$ 拒绝 H_0 ; $p > \alpha$ 接受 H_0 即 p 代表最小 α 对应最大 $1-\alpha$

以Z检验法为例，观察值 z_0 $\frac{1}{2}p = P\{|Z| > z_0\}$

Summary 总结

▸【关键】抽样定理

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$

则 if $X \sim N(\mu, \sigma^2)$ then $E(\bar{X}) = \mu$ $D(\bar{X}) = \dfrac{\sigma^2}{n}$

$\mapsto \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ $\dfrac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$ $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

if $X_1\cdots X_{n_1}$ 来自 $X \sim N(\mu_1, \sigma_1^2)$ $Y_1\cdots Y_{n_2}$ 来自 $Y \sim N(\mu_2, \sigma_2^2)$

$\mapsto \dfrac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$ $\sigma_1=\sigma_2=\sigma$ 时 $\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$

Chapter V

Title 数理统计总结　　Date /

Keywords 关键词　　Notes 笔记　　Review 复习记录 / / / / /

- 方差分析

§ 5. 方差分析

单因素方差分析

$H_0: \mu_1 = \mu_2 = \cdots = \mu_s$　(对应 $A_1 \cdots A_s$)

方差来源	平方和	自由度	均方	F比
因素A	S_A	$s-1$	$\bar{S}_A = \frac{S_A}{s-1}$	$F = \frac{\bar{S}_A}{\bar{S}_E}$
误差	S_E	$n-s$	$\bar{S}_E = \frac{S_E}{n-s}$	
总和	S_T	$n-1$		

$$S_T = \sum_{j=1}^{s}\sum_{i=1}^{n_j}(X_{ij} - \bar{X})^2$$

$$S_A = \sum_{j=1}^{s}\sum_{i=1}^{n_j}(\bar{X}_{\cdot j} - \bar{X})^2$$

$$S_E = S_T - S_A$$

拒绝域：$F = \frac{S_A/s-1}{S_E/n-s} \geq F_\alpha(s-1, n-s)$

双因素方差分析

重复实验：

方差来源	平方和	自由度	均方	F比
因素A	S_A	$r-1$	$\bar{S}_A = \frac{S_A}{r-1}$	$F_0 = \frac{\bar{S}_0}{\bar{S}_E}$
因素B	S_B	$s-1$	$\bar{S}_B = \frac{S_B}{s-1}$	
交互作用	$S_{A\times B}$	$(r-1)(s-1)$	$\bar{S}_{A\times B} = \frac{S_{A\times B}}{(r-1)(s-1)}$	
误差	S_E	$rs(t-1)$	$\bar{S}_E = \frac{S_E}{rs(t-1)}$	
总和	S_T	$rst-1$		

无重复实验：

方差来源	平方和	自由度	均方	F比
因素A	S_A	$r-1$	$\bar{S}_A = \frac{S_A}{r-1}$	$F_0 = \frac{\bar{S}_0}{\bar{S}_E}$
因素B	S_B	$s-1$	$\bar{S}_B = \frac{S_B}{s-1}$	
误差	S_E	$(r-1)(s-1)$	$\bar{S}_E = \frac{S_E}{(r-1)(s-1)}$	
总和	S_T	$rs-1$		

Summary 总结