概率论与数理统计

第一章 随机事件与概率

1. 如果A与B有且仅有一个发生，即同时成立 $A\cup B=\Omega$，$A\cap B=\phi$，则称A与B为对立事件（互逆），记为 $\overline{A}=B$ 或 $\overline{B}=A$

2. $A-B=A\overline{B}$ 而非…

3. 德摩根律：$\overline{\bigcup_{i=1}^{n}A_i}=\bigcap_{i=1}^{n}\overline{A_i}$；$\overline{\bigcap_{i=1}^{n}A_i}=\bigcup_{i=1}^{n}\overline{A_i}$

4. 条件概率：设A，B为两事件，且 $P(A)>0$，称 $P(B|A)=\frac{P(AB)}{P(A)}$

5. 独立性：$P(\bigcap_{i\in S}A_i)=\prod_{i\in S}P(A_i)$ 对∀可能的S集合成立.

∴三事件A，B，C相互独立，除 $P(AB)=P(A)P(B)$，$P(AC)=P(A)P(C)$
$P(BC)=P(B)P(C)$，还需 $P(ABC)=P(A)P(B)P(C)$

6. A与B相互独立的充要条件是A与 $\overline{B}$ 或 $\overline{A}$ 与B或 $\overline{A}$ 与 $\overline{B}$ 相互独立.

如：$P(A\overline{B})=P(A)P(\overline{B})\overset{?}{\Longrightarrow}P(AB)=P(A)P(B)$

$P(AB)+P(A\overline{B})=P(A)\Rightarrow P(AB)=P(A)-P(A\overline{B})=P(A)(1-P(\overline{B}))$
$=P(A)P(B)\Rightarrow A.B$ 独立

7. 当 $0<P(A)<1$ 时，A与B独立等价于 $P(B|A)=\frac{P(AB)}{P(A)}=P(B)$ 或
$P(B|A)=P(B|\overline{A})$ 成立.

8. 计算相互独立事件的概率时，常将事件之间的"并"、"差"转化为"交"计算。因为事件的独立性是用事件之交的概率来定义的.

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

$P(A-B)=P(A\overline{B})=P(A)P(\overline{B})$

9. 只给出概率的条件是得不出事件的结论的。⇐因为是离散概型还是连续概型是未知的。

D 例：已知随机事件A，B满足条件 $P(AB\cup\overline{A}\,\overline{B})=0$，则有：

(A) A，B为对立事件　　(B) A，B为互斥，但不对立

(C) $P(A)=P(B)$　　(D) $P(A)=P(\overline{B})$

利用上述结论，A，B肯定不对。

$P(AB\cup\overline{A}\,\overline{B})=P(AB)+P(\overline{A}\,\overline{B})-P(\underbrace{AB\cap\overline{A}\,\overline{B}}_{=\phi})=0$

$\Rightarrow P(AB)=P(\overline{A}\,\overline{B})=0$，又 $P(\overline{A}\,\overline{B})=1-P(A\cup B)=1-[P(A)+P(B)-P(AB)]$
$=1-P(A)-P(B)=0\Rightarrow P(A)=P(\overline{B})$

第二章 随机变量及其概率分布

1. 分布函数的性质:

<1> $0 \le F(x) \le 1$; $\lim_{x \to -\infty} F(x) = 0$, 记为 $F(-\infty)=0$; $\lim_{x \to +\infty} F(x) = 1$, 记为 $F(+\infty)=1$

<2> $F(x)$ 是单调非减函数, 即当 $x_1 < x_2$ 时, $F(x_1) \le F(x_2)$

<3> $F(x)$ 右连续, $F(x+0)=F(x)$

<4> 对 $\forall x_1 < x_2$, 有 $P\{x_1 < X \le x_2\} = F(x_2) - F(x_1)$

<5> 对 $\forall x$, $P\{X=x\} = F(x) - F(x-0)$ 单点概率, 用于计算间断点

<1><2><3> 简化: $F(-\infty)=0$, $F(+\infty)=1$, 右连续, 单调非减

<5>: 若 $F(x)$ 在 x 处连续, 则 $F(x) = F(x-0)$, 有 $P\{X=x\}=0$

例: 设随机变量 X 的 CDF 为 $F(x) = \begin{cases} a + \frac{b}{1+x}, & x \ge 0 \\ c, & x < 0 \end{cases}$, 其中 $a.b.c$ 为常数, 则 b 的取值范围为 $\underline{[-1,0]}$

$F(+\infty) = a = 1$, $F(-\infty) = c = 0$, $F'(x) = -\frac{b}{(1+x)^2} \ge 0$, $\underline{b \le 0}$ ($x \ge 0$ 时)

$F(0) = a+b = 1+b \ge 0$, $b \ge -1$

2. 设 X 是一个随机变量, 则它的函数 $Y=g(X)$ 也是随机变量、先求 Y 的分布函数 $F_Y(y)$:

反解或直接积分

$F_Y(y) = P\{Y \le y\} = P\{g(X) \le y\} = \int_{g(x) \le y} f_X(x)\,dx$.

注意先观察 y 的范围

第三章 多维随机变量及其分布

1. 独立性：如果对任意的 x, y 都有 $P\{X \le x, Y \le y\} = P\{X \le x\} P\{Y \le y\}$

$$F(x,y) = F_X(x) F_Y(y)$$

$$\begin{cases} \text{离散型：} P\{X = x_i, Y = y_j\} = P\{X = x_i\} P\{Y = y_j\} \\ \text{连续型：} f(x,y) = f_X(x) f_Y(y) \end{cases}$$

2. 二维均匀分布：如果二维连续型随机变量 (X, Y) 的概率密度为 $f(x,y) = \begin{cases} \frac{1}{A}, & (x,y) \in G \\ 0, & \text{其他} \end{cases}$ （A为平面有界区域G的面积、）

则称 (X, Y) 服从区域 G 上的均匀分布.

No.
DATE / /

第四章 随机变量的数字特征

离散 $E(X)=\sum_{k=1}^{+\infty} x_k p_k$

$\Updownarrow$

连续 $E(X)=\int_{-\infty}^{+\infty} x f(x)\,dx$

1. 期望运算：$E(cX)=cE(X)$，$E(X\pm Y)=E(X)\pm E(Y)$
   若 X,Y 相互独立/ X,Y 不相关（线性相关，$Y=\sin X$），$E(XY)=E(X)E(Y)$
2. 方差运算：$D(X)=E[X-EX]^2$，$D(aX+b)=a^2D(X)$
   $D(X\pm Y)=D(X)+D(Y)$ 并不总成立，注意二者独立性/相关性
   若 X,Y 不线性相关，$D(X\pm Y)=D(X)+D(Y)$
   $D(X)=E(X^2)-(EX)^2\geq 0 \Rightarrow E(X^2)\geq (EX)^2$
3. Γ 函数：
   实数域，$\Gamma(s)=\int_0^{+\infty} x^{s-1}e^{-x}dx \quad (s>0)$
   换元，$\Gamma(s)=2\int_0^{+\infty} x^{2s-1}e^{-x^2}dx \quad (s>0)$
   分部积分：$\Gamma(s+1)=s\Gamma(s) \Rightarrow \Gamma(n+1)=n!$

   对 $\frac{1}{1-x}$ 进行离散与连续展开：

$$\frac{1}{1-x}=\sum_{k=0}^{\infty}x^k=\int_0^{+\infty}e^{-(1-x)t}dt=\int_0^{+\infty}e^{-t}\sum_{k=0}^{\infty}\frac{(xt)^k}{k!}dt$$

$$=\sum_{k=0}^{\infty}\frac{\int_0^{+\infty}e^{-t}t^k dt}{k!}x^k \Rightarrow \frac{\int_0^{+\infty}e^{-t}t^k dt}{k!}=1 \Rightarrow \Gamma(k+1)=k!$$ 余元+换元

   余元公式：$\Gamma(1-x)\Gamma(x)=\frac{\pi}{\sin\pi x}$，$\Gamma(\frac{1}{2})=\sqrt{\pi}=2\int_0^{+\infty}e^{-x^2}dx$，$\int_0^{+\infty}e^{-x^2}dx=\frac{\sqrt{\pi}}{2}$

4. 常用随机变量的数学期望和方差.
   ① 泊松分布，$X\sim P(\lambda)$ $\quad P\{X=k\}=\frac{\lambda^k}{k!}e^{-\lambda}$（记参数分布）
   $E(X)=\lambda$，$D(X)=\lambda$
   ② 指数分布，$X\sim E(\lambda)$ $\quad f(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0\\ 0, & x\leq 0\end{cases}$，$\lambda>0$
   $E(X)=\frac{1}{\lambda}$，$D(X)=\frac{1}{\lambda^2}$

5. 协方差：$Cov(X,Y)=E[X-EX][Y-EY]=E(XY)-E(X)E(Y)$
   $Cov(X,Y)=0$，则称 X,Y 不相关

No.
DATE / /

X, Y若相互独立，$E(XY)=E(X)E(Y)$

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

6. 全期望定理：设$A_1, A_2, \cdots, A_n$为互不相容的n个事件，对每个i，$P(A_i)>0$，并且这些事件形成样本空间的一个分割。则

$$E(X) = \sum_{i=1}^{n} E(X|A_i)P(A_i)$$

相似地，$E(X) = \int_{-\infty}^{+\infty} E[X|Y=y] f_Y(y)\,dy$.

证明：
$$= \int_{-\infty}^{+\infty}\left[\int_{-\infty}^{+\infty} x f_{X|Y}(x|y)\,dx\right] f_Y(y)\,dy.$$

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

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

$$= \int_{-\infty}^{+\infty} x f_X(x)\,dx$$

$$= E(X)$$

$\Longrightarrow$ 重期望法则：$E[E(X|Y)] = E[X]$.

☆一个随机变量X的条件期望$E(X|Y=y)$的值，依赖于Y的值y。因为$E(X|Y=y)$是y的函数，所以$E(X|Y)$是Y的函数，因此也成为一个随机变量，它的分布依赖于Y的分布。$E(X|Y)$的期望和方差对估计和统计推断都很重要。

7. 全方差法则：$Var(X) = E[Var(X|Y)] + Var[E(X|Y)]$

来源：$Var(X) = Var(\underset{\text{估计量}}{\underset{\uparrow}{\hat{X}}}) + Var(\underset{\text{估计误差}}{\underset{\uparrow}{\tilde{X}}})$

8. 矩母函数：一个与随机变量X相关的矩母函数是参数s的函数$M_X(s)$，定义如下：

$$M_X(s) = E[e^{sX}] = \int_{-\infty}^{+\infty} e^{sx} f_X(x)\,dx \text{（连续）}$$

"矩母函数"这一名称是由于随机变量的矩可以通过矩母函数的公式计算。$\frac{d}{ds}M(s)\Big|_{s=0} = E(X)$，$\frac{d^n}{ds^n}M(s)\Big|_{s=0} = E[X^n]$.

第五章 大数定律

1. 极限理论

$$Z_n = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$$

中心极限定理研究Z_n的分布的渐进性质，并且得出结论：当n充分大的时候，Z_n的分布接近标准正态分布。

意义：(a) 从理论上看，极限理论为期望（或概率）和独立同分布试验序列之间的联系提供了合理解释

(b) 提供了S_n等随机变量序列当样本量n充分大时的渐近性质。

(c) 大量观测数据集下，在统计推断中发挥重要作用

2. 马尔可夫不等式：

设RV（随机变量）X只取非负值，则对任意$a>0$

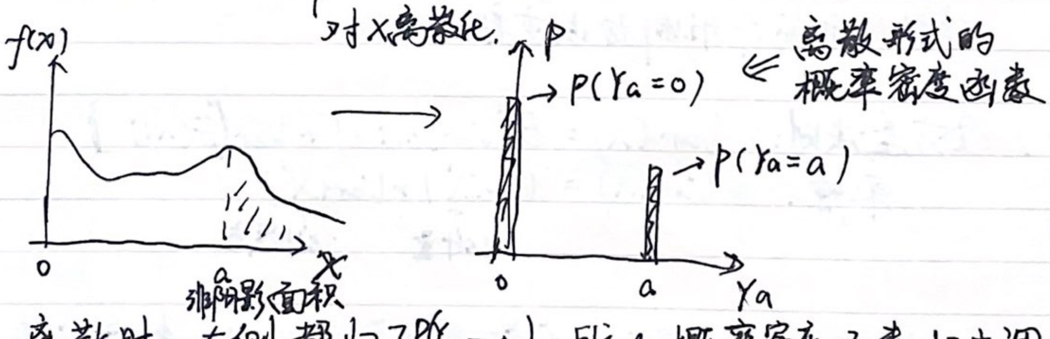
$$P(X \geq a) \leq \frac{E(X)}{a}$$

含义：

一个非负RV如果均值很小，则该RV取大值的概率也非常小

证明是设计了一个新的RV，固定正数a，定义RV Y_a

$$Y_a = \begin{cases} 0, & 若 X<a \\ a, & 若 X \geq a \end{cases}$$



离散时，0右侧都归了$P(Y_a=0)$，所以概率密度函数相当调整了质量分布，左边重，质量左移，期望减少

$$E(Y_a) \leq E(X) \Longrightarrow aP(Y_a=a) = aP(X \geq a) \leq E(X)$$

$$\Longrightarrow P(X \geq a) \leq \frac{E(X)}{a}$$

3. 切比雪夫不等式

设RV X的均值为 μ，方差为 σ^2，则对 $\forall c>0$，

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

含义：

如果一个R.V.的方差非常小的话，那么R.V.的取值远离（偏离）均值 μ 的概率也会非常小

证明．考虑非负R.V. $(X-\mu)^2$．令 $a=c^2$，使用马尔可夫不等式

$P((X-\mu)^2 \geq c^2) \leq \frac{E[(X-\mu)^2]}{c^2}$ （事件变成 $|X-\mu| \geq c$

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

4. 弱大数定律：

$\rightarrow E(\bar{X})=\mu$，$D(\bar{X})=\frac{\sigma^2}{n}$

$P(|\bar{X}-\mu| \geq \varepsilon) \leq \frac{\sigma^2/n}{\varepsilon} = \frac{\sigma^2}{n\varepsilon}$ 对 $\forall \varepsilon>0$ 成立．

对任意固定的 $\varepsilon>0$，上面不等式的右边在 $n\to\infty$ 时趋于0，得弱大数定律．

定义：设 $X_1, X_2, \cdots$ 独立同分布，其公共分布的均值为 μ，则对 $\forall \varepsilon>0$，当 $n\to\infty$ 时，

$$P(|\overset{\rightarrow \bar{X}}{M_n}-\mu| \geq \varepsilon) = P(|\bar{X}-\mu| \geq \varepsilon) \to 0$$

5. 依概率收敛：

设 $Y_1, Y_2, \cdots$ 是R.V.序列（不必相互独立），a 为一实数．如果对 $\forall \varepsilon>0$

都有 $\lim\limits_{n\to\infty} P(|Y_n-a| \geq \varepsilon)=0$，

则称 Y_n 依概率收敛于 a．

$\Rightarrow$ 依概率收敛的定义也可以这样描述，对 $\forall \varepsilon>0$ 和 $\delta>0$，$\exists n_0$，使得对所有的 $n \geq n_0$，都有 $P(|Y_n-a| \geq \underset{\text{精度}}{\varepsilon}) < \underset{\text{置信水平}}{\delta}$

所以（弱）大数定律就是 $\bar{X}$ 依概率收敛于 μ．

6. 中心极限定理.

设 $X_1, X_2, \cdots$ 是独立同分布的R.V.序列,序列的每一项的均值为 μ,方差为 σ^2,记

$$Z_n = \frac{X_1+X_2+\cdots+X_n-n\mu}{\sqrt{n}\sigma} = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$$

则 Z_n 的分布函数的极限分布为标准正态分布函数.

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz.$$

即 $\lim\limits_{n\to\infty} P(Z_n \le x) = \Phi(x)$ 对 $\forall x$ 成立.

证明会用到矩母函数,然后在 $s=0$ 处对其二阶泰勒展开

$n\to\infty$ 后,矩母函数变为标准正态的矩母函数.

7. 棣莫弗-拉普拉斯定理.

η_n 服从参数为 $n, p\ (0<p<1)$ 的二项分布

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

该定理表明:二项分布的极限分布是标准正态分布.

8. 强大数定律:(约束更强,偏离只发生有限次,导出频率就是概率)

设 $X_1, X_2, \cdots$ 是均值为 μ 的独立同分布R.V.序列,则样本均值 $M_n(\overline{X})$ 以概率1收敛于 μ,即

$$P\left(\lim_{n\to\infty} \overline{X} = \mu\right) = 1$$

9. 伯努利大数定理: $\lim\limits_{n\to\infty} P\left\{\left|\frac{f_A}{n} - p\right| \ge \varepsilon\right\} = 0$ (基于弱)

频率 概率

实际推断原理,在实际应用中,当试验次数很大时便可用事件的频率来代替事件的概率.

DATE / /

No.

第六章 样本及抽样分布

1. 所谓从总体抽取一个个体，就是对总体X进行一次观察并记录其结果。我们在相同的条件下对总体X进行n次重复的、独立的观察。$\Rightarrow X_1, X_2, \cdots, X_n$是独立同分布的，同期望、同方差。

统计学：样本 $\xrightarrow{\text{探索}}$ 分布，

2. 离群值检测：第一四分位数Q_1与第三四分位数Q_3之间的距离：$Q_3 - Q_1 \stackrel{def}{=} IQR$，称为四分位数间距，若数据小于$Q_1 - 1.5IQR$或大于$Q_3 + 1.5IQR$，就认为它是疑似异常值。

3. 统计量：设$X_1, X_2, \cdots X_n$是来自总体X的一个样本，$g(X_1, X_2, \cdots, X_n)$是$X_1, X_2, \cdots, X_n$的函数，若g中不含未知参数，则称$g(X_1, X_2, \cdots, X_n)$是一统计量。

☆统计量一定是随机变量，随机变量不一定是统计量。

如：

样本均值 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

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

$= \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - 2\overline{X}\,\overline{X}\cdot n + n\overline{X}^2\right) = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right)$

标准差：S

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

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

4. 若总体X的k阶矩$E(X^k) \stackrel{def}{=} \mu_k$存在，则当$n \to \infty$时，$A_k \xrightarrow[\text{独立值}]{P} \mu_k$，$k \in N_+$.

这是因为$X_1, X_2, \cdots, X_n$独立且与X同分布，所以$X_1^k, X_2^k, \cdots X_n^k$与$X^k$同分布，故有

$$E(X_1^k) = E(X_2^k) = \cdots = E(X_n^k) = \mu_k$$

从而由强大数定理可知 $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{P} \mu_k$，$k \in N_+$.

这就是矩估计法的理论依据．样本k阶原点矩依概率收敛总体k阶矩．

5. 经验分布函数：

设$X_1, X_2, \cdots, X_n$是总体F的一个样本，用$S(x)$，$-\infty < x < +\infty$表示$X_1, X_2, \cdots, X_n$中不大于x的随机变量的个数．定义经验分布函

$F_n(x)$
数为 $F_n = \frac{1}{n}S(x), -\infty < x < +\infty$.

若对样本进行排序：$x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$

则 $F_n(x)$ 的观察值为

$$F_n(x) = \begin{cases} 0, & 若 x < x_{(1)} \\ \frac{k}{n}, & 若 x_{(k)} \le x < x_{(k+1)},\ k=1,2,\cdots,n-1 \\ 1, & 若 x \ge x_{(n)} \end{cases}$$

☆对于经验分布函数 $F_n(x)$，格里汶科证明，对于任一实数 x，当 $n\to\infty$ 时，$F_n(x)$ 以概率 I 一致收敛于 $F(x)$（分布函数）

6. 正态总体的几个常用统计量的分布 $\begin{cases} \chi^2 分布 \\ t 分布 \\ F 分布 \end{cases}$ $\boxed{抽样分布}$.

7. 设 $X_1, X_2, \cdots, X_n$ 是来自总体 $N(0,1)$ 的样本，则称统计量

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

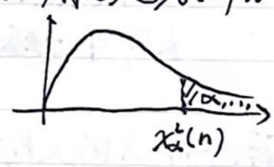
① 可加性：$\chi_1^2 \sim \chi^2(n_1)$，$\chi_2^2 \sim \chi^2(n_2)$，并且 χ_1^2, χ_2^2 相互独立

则有 $\chi_1^2 + \chi_2^2 \sim \chi^2(n_1+n_2)$

② χ^2 分布的数学期望和方差：若 $\chi^2 \sim \chi^2(n)$，则有 $E(\chi^2) = n$，$D(\chi^2) = 2n$

③ 上 α 分位点：对于给定的正数 α，$0<\alpha<1$，满足条件

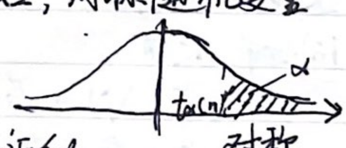
$P\{\chi^2 > \chi^2_\alpha(n)\} = \alpha$，则点 $\chi^2_\alpha(n)$ 就是 $\chi^2(n)$ 分布的上 α 分位点



8. t 分布.

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

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



$n > 45$ 时，对于常用 α 的值，可用正态分布近似

$t_\alpha(n) \approx z_\alpha$

9. F 分布.

设 $U \sim \chi^2(n_1)$，$V \sim \chi^2(n_2)$，且 U, V 相互独立，则称随机变量 $F = \frac{U/n_1}{V/n_2}$ 服从自由度为 (n_1, n_2) 的 F 分布，记为 $F \sim F(n_1, n_2)$

No.
DATE / /

① 倒数：若 $F\sim F(n_1,n_2)$，则 $\frac{1}{F}\sim F(n_2,n_1)$
② 分位点：$F_{1-\alpha}(n_1,n_2)=\frac{1}{F_\alpha(n_2,n_1)}$

10. 正态总体的样本均值与样本方差的分布.
设总体 X（不管服从什么分布，只要均值和方差存在）的均值为 μ，方差为 σ^2，$X_1,X_2,\cdots,X_n$ 是来自 X 的一个样本，$\overline{X}$，S^2 分别是样本均值和样本方差，则有 $E(\overline{X})=\mu$，$D(\overline{X})=\sigma^2/n$（基本运算公式）

$$E(S^2)=E\left[\frac{1}{n-1}\left(\sum_{i=1}^n X_i^2-n\overline{X}^2\right)\right]=\frac{1}{n-1}\left[\sum_{i=1}^n E(X_i^2)-nE(\overline{X}^2)\right]$$

又 $E(X_i^2)=D(X_i)+[E(X_i)]^2=\sigma^2+\mu^2$，$E(\overline{X}^2)=D(\overline{X})+[E(\overline{X})]^2=\frac{\sigma^2}{n}+\mu^2$

$$\therefore E(S^2)=\frac{1}{n-1}[n(\sigma^2+\mu^2)-\sigma^2-n\mu^2]=\sigma^2$$

进而，设 $X\sim N(\mu,\sigma^2)$，由正态分布可加性，$\overline{X}=\frac{1}{n}\sum_{i=1}^n X_i$ 服从正态分布.

① 定理1：设 $X_1,X_2,\cdots,X_n$ 是来自正态总体 $N(\mu,\sigma^2)$ 的样本，$\overline{X}$ 是样本均值，则有 $\overline{X}\sim N(\mu,\sigma^2/n)$

② 定理2：设 $X_1,X_2,\cdots,X_n$ 是来自总体 $N(\mu,\sigma^2)$ 的样本，$\overline{X}$，S^2 分别为样本均值和方差，则有.
1° $\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$ 2° $\overline{X}$ 与 S^2 相互独立.

③ 定理3：$X_1,X_2,\cdots,X_n$ 来自 $N(\mu,\sigma^2)$，

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

证：
$$\frac{\overline{X}-\mu}{S/\sqrt{n}}=\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\cdot\frac{\sigma}{S}=\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\Big/\frac{S\sqrt{n-1}}{\sigma\sqrt{n-1}}=\underbrace{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}_{\to N(0,1)}\Big/\sqrt{\frac{S^2(n-1)}{\sigma^2(n-1)}}\to\chi^2(n-1)$$

④ 定理4：$X_1,X_2,\cdots X_n$ 来自 $N(\mu_1,\sigma_1^2)$；$Y_1,Y_2,\cdots,Y_n$ 来自 $N(\mu_2,\sigma_2^2)$
且两个样本独立.

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

2° 当 $\sigma_1^2=\sigma_2^2=\sigma^2$ 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 $S_w^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$，$S_w=\sqrt{S_w^2}$

No

DATE / /

3° 方差非齐：

$$v=\frac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1-1}+\frac{(S_2^2/n_2)^2}{n_2-1}}$$

均值之差.

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}\sim t(v)$$

第七章 参数估计

1. 统计推断的基本问题可以分为两大类，一类是估计问题，另一类是假设检验问题。

2. 矩估计：样本矩依概率收敛于相应的总体矩 μ_l $(l=1,2,\cdots,k)$，样本矩的连续函数依概率收敛于相应的总体矩的连续函数

$$\begin{cases}\mu_1=\mu_1(\theta_1,\theta_2,\cdots,\theta_k)\\ \mu_2=\mu_2(\theta_1,\theta_2,\cdots,\theta_k)\\ \vdots\\ \mu_k=\mu_k(\theta_1,\theta_2,\cdots,\theta_k)\end{cases}\xrightarrow{\text{反解}}\begin{cases}\theta_1=\theta_1(\mu_1,\mu_2,\cdots,\mu_k)\\ \theta_2=\theta_2(\mu_1,\mu_2,\cdots,\mu_k)\\ \vdots\\ \theta_k=\theta_k(\mu_1,\mu_2,\cdots,\mu_k)\end{cases}$$ 矩估计量.

※ 用 A_i 分别代替上式中的 μ_i

（样本矩代替总体矩，估计未知参数）

3. MAP： $P(\Theta|X)=\frac{P(\Theta,X)}{P(X)}\propto P(\Theta,X)=P(X|\Theta)\,P(\Theta)$

后验. $P(X)$：对所有 Θ 分母都一样；$P(X|\Theta)$：信息、观测向量；$P(\Theta)$：先验.

$$\boxed{\begin{aligned}&\text{最大后验(MAP)：}\ \hat{\theta}=\arg\max_{\theta}P_{\Theta|X}(\theta|x)\\ &\text{或}\ \hat{\theta}=\arg\max_{\theta}f_{\Theta|X}(\theta|x)\end{aligned}}$$

4. MLE（最大似然估计）：固定样本观察值 $x_1,x_2,\cdots,x_n$，在 θ 可能取值范围 Θ 内挑选使得似然函数 $L(x_1,x_2,\cdots,x_n;\hat{\theta})=\max\limits_{\theta\in\Theta}L(x_1,x_2,\cdots,x_n;\theta)$，这样得到最大的参数值 $\hat{\theta}$，作为参数 θ 的估计值.

在很多情形下，$p(x;\theta)$ 和 $f(x;\theta)$ 关于 θ 可微，这时 $\hat{\theta}$ 常可从方程 $\frac{d}{d\theta}L(\theta)=0 \Longleftrightarrow \frac{d}{d\theta}\ln L(\theta)=0$ 解得。

↳ 对数似然方程.

5. 估计量的评选标准

① 无偏性：若估计量 $\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$ 的数学期望 $E(\hat{\theta})$ 存在，且对 $\forall\theta\in\Theta$ 有 $E(\hat{\theta})=\theta$，则称 $\hat{\theta}$ 是 θ 的无偏估计量.

② 有效性：$D(\hat{\theta}_1)\le D(\hat{\theta}_2)$ 且至少对于某一个 $\theta\in\Theta$ 上式中的不等号成立，则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 有效

③ 相合性：若对任意 $\theta\in\Theta$ 都满足：对于 $\forall\varepsilon>0$，有 $\lim\limits_{n\to\infty}P\{|\hat{\theta}-\theta|<\varepsilon\}=1$

即估计量 $\hat{\theta}$ 依概率收敛于 θ.

6. 区间估计.

背景或意义：① 我们希望得到估计的精确程度.

② 我们希望估计出一个范围，并希望知道这个范围包含参数 θ 真值的可信程度.

定义：设总体 X 的分布函数 $F(x;\theta)$ 含有一个未知参数 θ，$\theta\in\Theta$（Θ 是 θ 可能取值的范围），对于给定值 $\alpha\,(0<\alpha<1)$，若由来自 X 的样本 $X_1,X_2,\cdots,X_n$ 确定的两个统计量 $\underline{\theta}=\underline{\theta}(X_1,X_2,\cdots,X_n)$ 和 $\overline{\theta}=\overline{\theta}(X_1,X_2,\cdots,X_n)$，$(\underline{\theta}<\overline{\theta})$，对于 $\forall\theta\in\Theta$ 满足 $P\{\underline{\theta}(X_1,X_2,\cdots,X_n)<\theta<\overline{\theta}(X_1,X_2,\cdots,X_n)\}\ge1-\alpha$，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，$\underline{\theta}$ 和 $\overline{\theta}$ 分别称为置信水平为 $1-\alpha$ 的双侧置信区间置信下限、置信上限，$1-\alpha$ 称为置信水平.

☆ 对 $P\{\underline{\theta}(X_1,X_2,\cdots,X_n)<\theta<\overline{\theta}(X_1,X_2,\cdots,X_n)\}\ge1-\alpha$ 的理解：

① 若反复抽样多次（各次得到的样本的容量相等，都是 n）。每次得到的样本值确定一个区间 $(\underline{\theta},\overline{\theta})$，每个这样的区间要么包含 θ 的真值，要么不包含 θ 的真值. 按伯努利大数定律，包含 θ 真值的约占 $100(1-\alpha)\%$，不包含 θ 真值的约仅占 $100\alpha\%$。例如，若 $\alpha=0.01$，反复抽样1000次，则得到的1000个区间中不包含 θ 真值的约仅为10个。

② 一次抽样，计算出置信区间上下限得一 $(\underline{\theta}^*,\overline{\theta}^*)$，置信水平为 $1-\alpha$.

则正确的理解是 θ 真值落在该区间 $(\underline{\theta}^*,\overline{\theta}^*)$ 的概率为 $1-\alpha$.

$\underline{\theta}(X_1,X_2,\cdots,X_n)<\theta<\overline{\theta}(X_1,X_2,\cdots,X_n)$ 是一个事件，θ 是真值，不是随机变量.

当当当→最下面一横应该包住当
解：牟 解

7. 置信区间的计算，与前面介绍的抽样分布相结合，且主要集中于正态总体，见书P172.

第八章 假设检验

1. 假设检验的步骤：

① 根据实际问题的要求，提出原假设H_0及备择假设H_1

② 给定显著性水平α以及样本容量n

③ 确定检验统计量以及拒绝域的形式

④ 按P{当H_0为真拒绝H_0}$\le\alpha$求出拒绝域

⑤ 取样，根据样本观察值作出决策，是接受H_0还是拒绝H_1。

2. 显著性水平α的理解 →弃真

书上是这样说的：P{当H_0为真拒绝H_0}，我们无法排除这类错误的可能性，因此自然希望将犯这类错误限制在一定的限度之内，即给出一个较小的数α（$0<\alpha<1$），使犯这类错误的概率不超过α.

我的理解：当原假设为真时你拒绝的风险大小，（能接受的最大风险）因为在假设检验这个模型中，一般经计算的检验统计量落入拒绝域时，我们会拒绝原假设；但是若在重复的实验中，仍然有α的可能性这个拒绝是错误的。

3. 第二类错误β的理解：

① 施行特征函数/OC函数：$\beta(\theta)=P_\theta$（接受H_0）

当真值$\theta\in H_0$时，$\beta(\theta)$就是作出正确判断（即H_0为真时接受H_0）的概率故此时$\beta(\theta)\ge1-\alpha$；

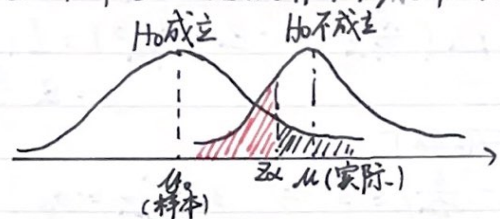
而当真值$\theta\in H_1$时，$\beta(\theta)$就是犯型Ⅱ错误的概率，而$1-\beta(\theta)$是作出正确判断的概率。函数$1-\beta(\theta)$称为检验法C的功效函数.

② 以Z检验为例，

右边检验问题：$H_0:\mu\le\mu_0$，$H_1:\mu>\mu_0$.

$\beta(\mu)=P_\mu$（接受H_0）$=P_\mu\left\{\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\le z_\alpha\right\}$.

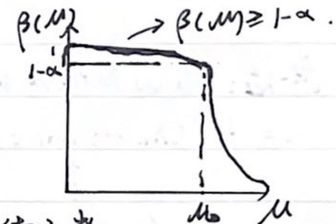
但计算型Ⅱ错误并不在原假设的分布中，而是在实际的分布中。



即应该计算红色（H_0为假但接受H_0，即检验统计量$\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \le z_\alpha$的概率）面积。此时抽样分布实际反应的是右侧的真实总体，

$$\beta(\mu) = P_\mu\Big(\underbrace{\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}}_{\text{中心为}\mu_0} \le z_\alpha\Big) = P_\mu\Big(\underbrace{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}_{\text{(中心转移到}\mu)} \le z_\alpha - \frac{\mu-\mu_0}{\sigma/\sqrt{n}}\Big) = \Phi(z_\alpha - \lambda),$$

$\lambda = \frac{\mu-\mu_0}{\sigma/\sqrt{n}}$。



③根据①②描述，$\beta(\mu)$图形：

此OC函数$\beta(\mu)$具有以下性质：

1° 它是$\lambda = \frac{\mu-\mu_0}{\sigma/\sqrt{n}}$的单调递减连续函数

2° $\lim\limits_{\mu\to\mu_0^+} \beta(\mu) = 1-\alpha$. $\lim\limits_{\mu\to\infty} \beta(\mu) = 0$

Φ单增，但λ系数为负，故单减

由$\beta(\mu)$的连续性可知，当参数的真值μ($\mu>\mu_0$)在μ_0附近时，检验的功效很低，即$\beta(\mu)$的值很大，亦即犯Ⅱ型错误的概率很大。因为α通常较小，不管σ多小，n多大，只要n给定，总存在μ_0附近的点μ($\mu>\mu_0$)使$\beta(\mu)$几乎等于$1-\alpha$.

④但③中的问题可以解决，使用OC函数($\beta(\mu)$)我们可以确定合适的样本容量n，使当真值$\mu \ge \mu_0+\delta$($\delta>0$为取定的值)时，犯型Ⅱ错误的概率不超过给定的β。这是由于$\beta(\mu)$是μ的递减函数，故当$\mu \ge \mu_0+\delta$时有

$$\beta(\mu_0+\delta) \ge \beta(\mu)$$

于是只要：

$$\beta(\mu_0+\delta) = \Phi\left(z_\alpha - \frac{\delta}{\sigma/\sqrt{n}}\right) \le \beta$$

$$\Rightarrow z_\alpha - \frac{\sqrt{n}\delta}{\sigma} \le -z_\beta \to \text{下}\beta$$

$$\Rightarrow \sqrt{n} \ge \frac{(z_\alpha+z_\beta)\sigma}{\delta}$$

就能使当$\mu \in H_1$且$\mu \ge \mu_0+\delta$时犯型Ⅱ错误的概率不大于β.

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

⑤ 左边检验问题: $H_0: \mu \geq \mu_0$, $H_1: \mu < \mu_0$ (真值在μ_0左边)

$$\beta(\mu) = P_\mu\left(\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \geq -z_\alpha\right) = P_\mu\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \geq -z_\alpha - \frac{\mu-\mu_0}{\sigma/\sqrt{n}}\right) = \Phi\left(z_\alpha + \frac{\mu-\mu_0}{\sigma/\sqrt{n}}\right)$$

$$\lambda = \frac{\mu-\mu_0}{\sigma/\sqrt{n}}$$

$\mu \in H_1, \mu \leq \mu_0-\delta$ 时 $\beta(\mu)$ 单↗, $\beta(\mu) \leq \beta(\mu_0-\delta)$

$$\beta(\mu_0-\delta) = \Phi\left(z_\alpha - \frac{\delta}{\sigma/\sqrt{n}}\right) \leq \beta \Rightarrow z_\alpha - \frac{\delta}{\sigma/\sqrt{n}} \leq -z_\beta, \quad \sqrt{n} \geq \frac{(z_\alpha+z_\beta)\sigma}{\delta}$$

⑥ 双边检验问题: $\sqrt{n} \geq \frac{(z_{\frac{\alpha}{2}}+z_\beta)\sigma}{\delta}$

⑦ 综合④⑤⑥,我们可以用增大样本容量的方法降低型II错误.

4. 总结: 记W为拒绝域,对于样本$x_1, x_2, \cdots, x_n$来讲:

$\alpha = P\{(x_1, \cdots, x_n) \in W \mid H_0\} \Rightarrow$ 拒真错误

作出的判断, 实际上θ的取值

$\beta = P\{\underbrace{(x_1, \cdots, x_n)}_{一个检验统计量} \in \overline{W} \mid H_1\} \Rightarrow$ 取伪错误

$$\alpha(\theta) = P\{(x_1, \cdots, x_n) \in W \mid \theta \in \Theta_0\}$$
$$\beta(\theta) = P\{(x_1, \cdots, x_n) \in \overline{W} \mid \theta \in \Theta_1\}$$

功效函数: $g(\theta) = P\{(x_1, \cdots, x_n) \in \overset{拒绝域}{W} \mid \theta\}$,

当$\theta \in \Theta_0$时, $g(\theta)=\alpha$ (计算$g(\theta)$使用的分布与原假设成立时的抽样分布一致)

当$\theta \in \Theta_1$时, 计算$g(\theta)$要在实际的抽样分布,具体的例子见前面Z检验。

5. 分布拟合检验.

① 单个分布的χ^2拟合检验法.

设总体X的分布未知, $x_1, x_2, \cdots, x_n$是来自X的样本值,假设检验构造如下:

H_0: 总体X的分布函数为$F(x)$

H_1: 总体X的分布函数不是$F(x)$.

我们使用统计量 $\sum_{i=1}^{k} C_i\left(\frac{f_i}{n}-p_i\right)^2$ 来度量样本与H_0中所

No.
DATE / /

假设的分布的吻合程度，令 $C_i = n/p_i$，则下述统计量

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

在 n 充分大 ($n \geq 50$) 时，且 H_0 为真时近似服从 $\chi^2(k-1)$ 分布.

② 一个很重要的应用：我们可以用 χ^2 齐性检验，检验样本性状是否符合遗传比例.（分布 $F(x)$）

6. p值的定义：假设检验问题的 p-value 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值的作用是给出统计量的位置.

第九章 ANOVA 及回归分析

1. ANOVA 模型的本质是一种方差分解技巧。"方差分析"事实上不是真正分析方差，而是分析用偏差平方和度量的数据的变异。方差是数据信息的载体，没有(数据)变异的数据是没有意义的。
Snedecor 说过："它是从可比组的数据中分解出可追溯到某些指定来源的变异的一种技巧."

假设检验，设计 $\frac{\text{组间方差}}{\text{组内方差}} = \frac{S_A}{S_E}$ 这个统计量 → F统计量. 当实验组得到的数据使得该统计量足够大时，那么因素对试验的指标影响很大.

数据间总的变异/离散程度　　随机误差.　　各水平数据的差异.

组内　　组间

总偏差平方和分解：$S_T = \sum_{j=1}^{s}\sum_{i=1}^{n_j}(x_{ij} - \bar{x})^2 = S_E + S_A$

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

即 $\sum_{j=1}^{s} n_j(\bar{x}_{\cdot j} - \bar{x})^2$

2. 回归分析中的F检验.
$SST = SSR + SSE \rightarrow R^2 = \frac{SSR}{SST}$
也是一种方差的分解.

证明Cauchy-Schwarz不等式

今天早上听了我院董朝华教授作的《计量经济学与课本上的那些事儿》报告，收获颇多。在感叹老师精彩讲述的同时，我也对之前被我忽视的Cauchy-Schwarz不等式有了新的认识。

报告开始，老师给我们这些学员提出了两点寄语或者是期望。第一，是知识可以忘记，但要记得反复回味，这样我们才能将知识转化为生产力，更好地投入到学术研究；第二，不问收获，只问耕耘，在做学问上，从来就没有坦途，储备知识是研究生涯的第一要务。

接着，老师为我们提纲挈领地总结了大学阶段学习的数理框架：数学分析、泛函分析等是分析方法；代数、矩阵分析等是代数方法；数理统计、概率论等是应用数学。在数理金融的学海里徜徉时，要温故知新，及时回味这些基本功。

①对Cauchy-Schwarz不等式，我们先给出的是它在函数空间中的定义：

如果 $\int f^2(x)dx<\infty$，$\int g^2(x)dx<\infty$，$\left|\int f(x)g(x)dx\right| \le \left(\int f^2(x)dx\right)^{\frac{1}{2}}\left(\int g^2(x)dx\right)^{\frac{1}{2}}$.

它的证明并不复杂，从必然成立的条件出发：

对 $\forall\alpha$，$0 \le \int (f(x)+\alpha g(x))^2 dx$，展开：

$$0 \le \int f^2(x)dx + 2\alpha\int f(x)g(x)dx + \alpha^2\int g^2(x)dx.$$

所以，关于 α 的二次三项式应有：$\Delta = 4\left(\int f(x)g(x)dx\right)^2 - 4\int f^2(x)dx\cdot\int g^2(x)dx \le 0$

即有 $\left|\int f(x)g(x)dx\right| \le \left(\int f^2(x)dx\right)^{\frac{1}{2}}\left(\int g^2(x)dx\right)^{\frac{1}{2}}$，得证。

可以注意到 $\left(\int f^2(x)dx\right)^{\frac{1}{2}}$、$\left(\int g^2(x)dx\right)^{\frac{1}{2}}$ 是 f、g 的模，简记为 $\|f\|$, $\|g\|$.

当然董老师介绍了更为巧妙的证法：

从必然成立的条件出发，$0 \le \int (af(x)+bg(x))^2 dx$.

展开：$0 \le a^2\int f^2(x)dx + 2ab\int f(x)\cdot g(x)dx + b^2\int g^2(x)dx$，显然这是一个二次型，引入矩阵、向量记法：

$$\forall a,b\in R,\quad 0 \le (a\ \ b)\begin{pmatrix}\int f^2(x)dx & \int f(x)g(x)dx \\ \int f(x)g(x)dx & \int g^2(x)dx\end{pmatrix}\begin{pmatrix}a \\ b\end{pmatrix}$$

No.

DATE / /

而对于一个距离的二次型，且$\forall a,b\in R$，有距离≥ 0，则分解出来的对称矩阵满足半正定，即任意阶子式≥ 0。

一阶必满足，

二阶则有 $\begin{vmatrix} \int f^2(x)dx & \int f(x)g(x)dx \\ \int f(x)g(x)dx & \int g^2(x)dx \end{vmatrix} = \int f^2(x)dx\int g^2(x)dx - \left(\int f(x)g(x)dx\right)^2 \geq 0.$

即有 $\left|\int f(x)g(x)dx\right| \leq \left(\int f^2(x)dx\right)^{\frac{1}{2}}\left(\int g^2(x)dx\right)^{\frac{1}{2}}$，得证。

随之，董老师将该问题推广至向量空间和概率空间。

②对于向量空间，Cauchy-Schwarz不等式可表示为：对于2个vector：$\vec{a},\vec{b}\in R^n$，有 $|\langle\vec{a},\vec{b}\rangle| \leq \|\vec{a}\|\|\vec{b}\|$，证明类似：

由 $0\leq(\vec{a}+\alpha\vec{b})^2$ 对$\forall\alpha$成立，有 $0\leq\vec{a}^2+\alpha^2\vec{b}^2+2\alpha\langle\vec{a},\vec{b}\rangle$

$\Delta=4\langle\vec{a},\vec{b}\rangle^2-4\vec{a}^2\vec{b}^2\leq 0$，有$|\langle\vec{a},\vec{b}\rangle|\leq\|\vec{a}\|\cdot\|\vec{b}\|$

取"="条件是$\vec{a}$，$\vec{b}$线性相关，即$\cos\theta=1$，$\theta=0$。

③对于概率空间，Cauchy-Schwarz不等式可表示为：对于2个随机变量X，Y，若$E[X^2]<\infty$，$E[Y^2]<\infty$，则$|E[XY]|\leq\sqrt{E[X^2]}\sqrt{E[Y^2]}$

证明思路依旧：

$$E\left[(X+\alpha Y)^2\right]\geq 0，展开：E\left[X^2+2\alpha XY+\alpha^2Y^2\right]$$

$$=E(X^2)+2\alpha E(XY)+\alpha^2E(Y^2)\geq 0 \text{ 对 } \forall\alpha \text{ 成立}.$$

$$\Delta=4E^2(XY)-4E(X^2)E(Y^2)\leq 0，得证.$$

④最后，老师将推导的一致性用泛函分析的方法做了解释。在任意内积空间V中，有 $|\langle x,y\rangle|^2\leq\langle x,x\rangle\langle y,y\rangle$，$\forall x,y\in V$。

有一位同学指出函数空间中的证明要用绝对值不等式进行补充，非常全面。

最后，作为本校生，祝祝咱们院夏令营越办越好！

中南财经政法大学 大数据1901

苏烨