内容简介

本书记录了作者在山大数院的三年所学的数学知识，简明扼要地列举出每个学科所必须掌握的定理等重要知识点。此书不宜作为初学某一科目的参考资料，而适合已学完部分内容者查漏补缺。

根据山大本科开课顺序以及个人自学进度，大致收录下列学科：

大一上；数学分析 1，高等代数 1，解析几何

大一下：数学分析 2，高等代数 2

大二上：数学分析 3，复变函数，常微分方程

大二下：实变函数，偏微分方程，概率论

大三上：机器学习、数字图像处理、数理统计、矩阵论

大三下：时间序列分析、数据库系统、数据结构

目　录

第一章　概率论与数理统计

Mathematical Analysis

概率论与数理统计是由数分高代派生出来的应用学科，用于刻画日常生活中随机发生的事件，具有很高的应用价值．其中，概率论主要研究随机变量的分布与特征，而数理统计主要研究通过样本对未知分布进行估计．

概率论的重点：概率的定义，条件概率与独立性，一元或多元随机变量分布，常用分布函数，随机变量的特征数，大数定律和中心极限定理

数理统计的重点：基本概念与三大分布，参数估计，假设检验，方差分析，回归分析

1.1　随机事件与概率

之前数学分析研究的内容都是具有确定解析式或约束条件的函数, 但概率论引进了随机因素, 即实验和结果并不是一一对应的, 一次实验可能会出多种结果. 这一部分的任务是使用概率这一量化方式, 将随机性规范化.

1.1.1　随机事件

1. 随机现象: 重复实验会出现不同结果的现象.
2. 样本空间: 随机现象可能出现的结果组成的集合.
3. 随机事件: 样本空间的子集. 当实验结果属于此子集时, 称随机事件发生.
4. 随机变量: 用于描述随机事件的人为设定变量 (非正式定义).
5. 事件的运算: 和集合一致, 有交并补余四大运算. 有两个公式很重要.

   (1) 集合减法公式: $A - B = A \cap \bar{B}$.

   (2) 德摩根律: $\overline{A \cup B} = \bar{A} \cap \bar{B}$; $\overline{A \cap B} = \bar{A} \cup \bar{B}$.

6. 事件域: 令 Ω 为样本空间, 定义事件域 $\mathscr{F}$ 符合下列性质:

   (1) $\Omega \in \mathscr{F}$; (2)$A \in \mathscr{F} \Rightarrow \bar{A} \in \mathscr{F}$; (3)$A_n \in \mathscr{F} \Rightarrow \bigcup\limits_{n=1}^{\infty} A_n \in \mathscr{F}$.

1.1.2　概率

1. 公理化定义: 在事件域 $(\Omega, \mathscr{F})$ 上定义可测函数 $P(A)$ 满足:

   (1) 非负性: $P(A) \geq 0$; (2) 正则性: $P(\Omega) = 1$;

   (3) 可列可加性: 事件 $A_1, \cdots, A_n$ 互不相容时, $P(\bigcup\limits_{n=1}^{\infty} A_n) = \sum\limits_{n=1}^{\infty} P(A_i)$.

2. 用频率定义概率: 令 $n(A)$ 为事件 A 发生的频数, 则可用大量重复事件的频率表示概率: $P(A) = \lim\limits_{n \to \infty} \frac{n(A)}{n}$.

3. 古典概型: 若样本空间有 n 个等可能发生的样本点, 则事件 A 包含 k 个样本点时, $P(A) = \frac{k}{n}$.

4. 几何概型: 若样本空间 Ω 的面积测度为 S_n, 事件 A 包含其中面积为 S_A 的一部分, 则 $P(A) = \frac{S_A}{S_n}$. (蒙特卡罗法的理论依据)

5. 贝叶斯概率: 对事件发生可能性的主观预测, 在机器学习中使用频率很高.

1.1.3 概率的性质

1. 有限可加性: 若 $A_1, \cdots, A_n$ 互不相容, 则 $P(\bigcup\limits_{i=1}^{n} A_i) = \sum\limits_{i=1}^{n} P(A_i)$.
2. 单调性: 若 $A \subset B$, 则 $P(A) \leq P(B)$.
3. 加法公式: $P(A \bigcup B) = P(A) + P(B) - P(AB)$.

1.1.4 条件概率

1. 定义: $P(A|B)$ 表示已知 B 发生的条件下 A 发生的概率. $P(A|B) = \frac{P(AB)}{P(B)}$.
2. 乘法公式: $P(AB) = P(B)P(A|B)$, 即定义式的变种.
3. 全概率公式: 若 B_i 互不相容, 且 $\bigcup\limits_{i=1^n} B_i = \Omega$, 则

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

.

4. 贝叶斯公式: 用先验概率推后验概率. 若 B_i 互不相容, 且 $\bigcup\limits_{i=1^n} B_i = \Omega$, 则

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum\limits_{k=1}^{n} P(A|B_k)P(B_k)}$$

.

5. 独立性: 若 $P(A|B) = P(A)$, 即 $P(AB) = P(A)P(B)$, 则称事件 A 和 B 相互独立.

1.2　随机变量及其分布

用概率描述随机事件发生可能性的大小后, 为了更充分认识随机事件, 我们引入随机变量来刻画随机事件, 如抽奖是随机事件, 在此基础上可以定义随机变量” 是否中奖”, 这是一个二值随机变量 (0/1).

使用随机变量来描述随机事件, 能更方便地研究随机事件中我们感兴趣的性质, 比如随机变量” 灯泡坏掉的个数” 能帮助我们衡量灯泡的寿命. 这些随机变量取值的规律可以用分布来描述, 离散随机变量和连续随机变量的刻画方式略有区别.

1.2.1　随机变量

1. 定义: 样本空间 Ω 上的实值函数 $X(\omega)$.

2. 离散随机变量的确定: 使用分布列描述.

X	X_1	X_2	$\cdots$	X_n
P	p_1	p_2	$\cdots$	p_n

其中 p_i 表示随机变量 X 取值 X_i 的概率, $\sum\limits_{i=1}^{n} p_i = 1$.

3. 连续随机变量的描述: 使用分布函数与概率密度函数.

(1) 分布函数 $F(x)$: $F(x) = P(X \leq x)$, 是单调递增的右连续函数, 且 $F(+\infty) = 1, F(-\infty) = 0$.

(2) p.d.f 概率密度函数 $p(x)$: $p(x) = F'(x)$, 是非负函数且 $\int_{-\infty}^{+\infty} p(x)\mathrm{d}x = 1$.

1.2.2 常用分布及概率密度函数

1. 离散分布

(1) 泊松分布: $P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$, 用于计数过程, 记作 $X\sim P(\lambda)$.

(2) 伯努利分布: $P(X=1)=p, P(X=0)=1-p$, 又称两点分布.

(3) 二项分布: $P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$, 即 n 重伯努利分布中事件发生的次数, 记作 $X\sim b(n,p)$.

(4) 几何分布: $P(X=k)=(1-p)^{k-1}p$, 具有无记忆性.

2. 连续分布

(1) 正态分布: $p(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{1\frac{(x-\mu)^2}{2\sigma^2}}$, 是最常用的分布. 记作 $X\sim N(\mu,\sigma^2)$, 标准正态分布即 $N(0,1)$.

(2) 均匀分布: $p(x)=\frac{1}{b-a}$, 其中 $x\in(a,b)$, 记作 $X\sim U(a,b)$.

(3) 指数分布: $p(x)=\lambda e^{-\lambda x}$, 其中 $x\geq 0$, 记作 $X\sim\epsilon(\lambda)$, 具有无记忆性.

(4) 伽马分布: $p(x)=\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}$, 其中 $x\geq 0$, 记作 $X\sim Ga(\alpha,\lambda)$. 特殊地, $Ga(\frac{n}{2},\frac{1}{2})=\chi^2(n)$ 为卡方分布, 统计中常用.

1.2.3 数字特征

1. 数学期望: X 在不同取值数按概率的加权平均数, 是消除随机性的主要手段, 记作 Ex. 在离散场合, $EX=\sum\limits_{n=1}^{\infty}p_ix_i$. 在连续场合, $EX=\int_{-\infty}^{+\infty}xp(x)\mathrm{d}x$.

2. 方差: $DX=E[(X-EX)^2]$, 也记作 $\mathrm{Var}(X)$, 用于衡量数据的集中程度. 常用的计算公式为

$$DX=E(X^2)-(EX)^2$$

.

3. 标准差: $\sigma(x)=\sqrt{DX}$, 也记作 $Std(X)$, 好处是与 X 的量纲一致.

4. 切比雪夫不等式: $P(|X-EX|\geq textVarepsilon)\leq\frac{DX}{textVarepsilon^2}$.

1.2.4 随机变量函数的分布

1. 离散情形: 先求各项的像 $g(x_1),\cdots,g(x_n)$, $g(x_i$ 对应概率仍为 p_i, 再合并相同项.

2. 连续情形: 若 $Y=g(x)$ 严格单调, 反函数为 $x=h(y)$, X 的概率密度函数为 $p(x)$, 则 Y 的概率密度函数为 $p_Y(y)=p_X(h(y))\cdot|h'(y)|$. 一般情况下, 需要根据 $P(g(x)\leq y)$ 反解出 x 的范围, 再利用 X 的分布函数求解.

1.3 多元随机变量及其分布

若样本点含有不止一个我们感兴趣的属性, 如身体指标包含身高和体重, 则可定义多元随机变量来刻画这些指标的分布. 研究多元随机变量, 除了明确各分量的分布外, 还需要研究各分量间的相关关系, 以及给定某条件后的分布情况.

事实上, 只要给定多元随机变量的联合分布, 就能得到所有信息, 该部分的目的就是掌握将信息从联合分布中提取出来的方法.

1.3.1 多元随机变量

1. 定义: 样本空间 Ω 上的向量值函数 $X(\omega)=(X_1(\omega),\cdots,X_n(\omega))$.

2. 联合分布函数: $F(x_1,\cdots,x_n)=P(X_1\leq x_1,\cdots,X_n\leq x_n)$

3. 离散情形的联合分布列: 仅用于二元分布 (X,Y), 用 i 行 j 列元素 p_{ij} 表示 $X=X_i, Y=Y_j$ 的概率, 其中 $\sum\limits_{i,j} p_{ij}=1$.

4. 连续情形的联合密度函数: $p(x_1,\cdots,x_n)=\frac{\partial^n F(x_1,\cdots,x_n)}{\partial x_1\partial x_2\cdots\partial x_n}$.

5. 多元正态分布: 最重要的多元连续分布. 令 $x=(x_1,\cdots,x_n)$, 均值向量为 μ, 协方差矩阵为 Γ, 则 n 元正态分布的联合密度函数

$$p(x_1,\cdots,x_n)=\frac{1}{(2\pi)^{\frac{n}{2}}|\Gamma|^{\frac{1}{2}}}e^{-\frac{1}{2}(x-\mu)^T\Gamma^{-1}(x-\mu)}$$

特殊地, 当 $n=2$ 时, 二元正态分布为

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

记作 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$.

1.3.2　边缘分布

1. 边缘分布函数: $F_x(x)=F(x,\infty)$, $F_y(y)=F(\infty,y)$.

2. 离散情形的边缘分布列: $P(X=X_i)=\sum\limits_j P(X=X_i,Y=Y_j)$; $P(Y=Y_j)=\sum\limits_i P(X=X_i,Y=Y_j)$

3. 连续情形的边缘密度函数: $p_X(x)=\int_{-\infty}^{\infty}p(x,y)\mathrm{d}y$; $p_Y(y)=\int_{-\infty}^{\infty}p(x,y)\mathrm{d}x$.

4. 随机变量的独立性: 若 $\prod\limits_{i=1}^{n}p_i(X_i)=p(x_1,\cdots,x_n)$, 即联合密度函数为边缘密度函数之积, 则称 $X_1,\cdots,X_n$ 相互独立.

1.3.3　多元随机变量函数的分布

1. $Z = X + Y$ 的分布: 可用后面提到的特征函数法, 也可用卷积公式, 即 $p_Z(z) = \int_{-\infty}^{\infty} p_X(x)p_Y(z-x)\mathrm{d}x$.

2. 次序统计量分布: 若 $X_{(1)}, \cdots, X_{(n)}$ 独立同分布且升序排列, 则第 k 个次序统计量 $X_{(k)}$ 的概率密度函数为

$$p_{(k)}(x) = \frac{n!}{(k-1)!(n-k)!}[F(x)]^{k-1}p(x)[1-F(x)]^{n-k}$$

特殊地, $\min X$ 即 $X_{(1)}$ 的概率密度函数为 $n[1-F(x)]^{n-1}p(x)$; $\max X$ 即 $X_{(n)}$ 的概率密度函数为 $n[F(x)]^{n-1}p(x)$.

3. 变量变换法: 令 $u = u(x,y), v = v(x,y)$, 从中反解出 $x = x(u,v), y = y(u,v)$, 则 $p(u,v) = p(x,y)\left|\frac{\partial(x,y)}{\partial(u,v)}\right|$.

1.3.4　多元随机变量的特征数

1. 数学期望: $g(x,y)$ 的期望为 $\int_R g(x,y)p(x,y)\mathrm{d}x\mathrm{d}y$.

2. 方差: 定义不变, 仍有 $\mathrm{Var}(x) = E(X-EX), \mathrm{Var}(y) = E(Y-EY)$.

3. 协方差: $\mathrm{Cov}(X,Y) = E[(X-EX)(Y-EY)] = E(XY) - EXEY$, 用于刻画两变量的相关程度.

4. 相关系数: $\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}$. 当 $\mathrm{Corr}(x) \in (0,1]$ 时, 称 X 和 Y 正相关; $\mathrm{Corr}(x) \in [-1,0)$ 时, 称 X 和 Y 负相关; $\mathrm{Corr}(x) = 0$ 时, 称 X 和 Y 不相关.

5. 方差运算性质: $\mathrm{Var}(X \pm Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) \pm 2\mathrm{Cov}(X,Y)$

6. n 元随机变量的协方差矩阵:

$$\Gamma=\begin{pmatrix}\mathrm{Cov}(X_1,X_1) & \mathrm{Cov}(X_1,X_2) & \cdots & \mathrm{Cov}(X_1,X_n)\\ \mathrm{Cov}(X_2,X_1) & \mathrm{Cov}(X_2,X_2) & \cdots & \mathrm{Cov}(X_2,X_n)\\ \vdots & \vdots & \vdots & \vdots\\ \mathrm{Cov}(X_n,X_1) & \mathrm{Cov}(X_n,X_2) & \cdots & \mathrm{Cov}(X_n,X_n)\end{pmatrix}$$

用于刻画各分量之间的总体相关性.

1.3.5 条件分布与条件期望

1. 离散条件分布: $P_{i|j}=P(X=X_i|Y=y_j)=\frac{p_{ij}}{\sum\limits_i p_{ij}}$.

2. 连续条件分布: $p(y|x)=\frac{p(x,y)}{p_X(x)}$; $p(x|y)=\frac{p(x,y)}{p_Y(y)}$

3. 全概率公式: $p_Y(y)=\int_{-\infty}^{\infty}p_X(x)p(y|x)\mathrm{d}x$.

4. 贝叶斯公式:

$$p(x|y)=\frac{p(y|x)p_X(x)}{\int_{-\infty}^{\infty}p(y|x)p_X(x)\mathrm{d}x}$$

.

5. 条件数学期望: $E(X|Y=y)=\int_{-\infty}^{\infty}xp(x|y)\mathrm{d}x$.

6. 重期望公式: $E[E(X|Y)]=EX$.

1.4　大数定律和中心极限定理

这部分首先将傅里叶变换引入概率密度函数的求解中, 得到特征函数这个很好用的工具, 再借助特征函数推导大数定律和中心极限定理的一般结论, 为数理统计的展开做好铺垫.

大数定律的内容很简单, 就是抽样次数足够大时, 频率近似于概率, 均值近似于数学期望, 这给大样本统计提供了理论依据. 中心极限定理说明多个独立同分布随机变量之和近似于正态分布, 这鼓励我们在大样本统计中使用正态分布进行统计推断.

1.4.1　随机变量的特征函数

1. 定义: $\varphi(t)=E(e^{itX})$. 离散情形下, $\varphi(t)=\sum\limits_{k=1}^{\infty}p_k e^{itX_k}$; 连续情形下, $\varphi(t)=\int_{-\infty}^{\infty}p(x)e^{itx}\mathrm{d}x$.

2. 性质: (1) 若 X 与 Y 独立, $Z=X+Y$, 则 $\varphi_Z(t)=\varphi_X(t)\cdot\varphi_Y(t)$.

(2) 求各阶矩的方式: $\varphi^{(k)}(0)=i^kE(X^k)$.

(3) 唯一性定理: 分布函数由特征函数唯一确定.

3. 逆转公式: $F(x_2)-F(x_1)=\lim\limits_{T\to+\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-itx_1}-e^{-itx_2}}{it}\varphi(t)\mathrm{d}t$.

4. 连续随机变量的逆变换公式: $p(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}\varphi(t)\mathrm{d}t$.

1.4.2　大数定律

1. 一般形式: 对任意 $\varepsilon>0$, 有

$$\lim_{n\to+\infty}P\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}EX_i\right|<\varepsilon\right)=1$$

.

2. 伯努利大数定律: 令 S_n 为 n 重伯努利试验中事件发生的次数, p 为事件发生的概率, 则对任意 $\varepsilon > 0$, 有

$$\lim_{n\to+\infty} P\left(\left|\frac{S_n}{n} - p\right| < \varepsilon\right) = 1$$

.

3. 切比雪夫大数定律: 当 $\{X_n\}$ 两两不相关且 $Var(X_i)$ 有界时, 大数定律成立.

4. 辛钦大数定律: 当 $X_1, \cdots, X_n$ 独立同分布且 EX_i 存在时, 大数定律成立.

5. 辛钦大数定律的证明: 令 $\varphi(t)$ 为 X_i 共同的特征函数, 数学期望为 a, 将 $\varphi(t)$ 在 $t=0$ 处泰勒展开: $\varphi(t) = 1 + iat + o(t)$. 故 $\varphi_{\frac{1}{n}\sum\limits_{i=1}^{n} X_i}(t) = \left[\varphi\left(\frac{t}{n}\right)\right]^n \sim e^{iat}$, 恰是退化分布的特征函数.

1.4.3　中心极限定理

1. 林德伯格-莱维中心极限定理: 令 X_n 独立同分布, $EX_i = \mu$, $DX_i = \sigma^2$, 则 $n \to \infty$ 时, $\frac{\sum\limits_{i-1}^{n} X_i - n\mu}{\sigma\sqrt{n}}$ 的分布弱收敛于标准正态分布.

2. 蒂莫夫-拉普拉斯中心极限定理: 令 S_n 为 n 重伯努利试验中事件发生的次数, p 为事件发生的概率, 则 $\frac{S_n - np}{\sqrt{np(1-p)}}$ 在 $n \to \infty$ 时的分布弱收敛于标准正态分布.

3. 中心极限定理证明: 将 X_i 标准化: $Y_i = \frac{X_i - \mu}{\sigma}$, 则 $EY_i = 0$ 且 $DY_i = 1$. 令 Y_N 的特征函数为 $\varphi(t)$, 故 $\frac{1}{\sqrt{n}}\sum\limits_{i=1}^{n} Y_i$ 的特征函数为 $\left[\varphi\left(\frac{t}{\sqrt{n}}\right)\right]^n$. 由泰勒展开: $\varphi\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + o(t^2)$, 在 $n \to \infty$ 时, $\frac{1}{\sqrt{n}}\sum\limits_{i=1}^{n} Y_i$ 的特征函数趋近于 $e^{-\frac{t^2}{2}}$, 恰为 $N(0,1)$ 的特征函数.