基础概率论

致敬

本书作者：李贤平先生

目录

第一章：事件与概率

1.随机现象与统计规律性

1. 随机现象
    1. 必然事件：在一定条件下，必然会发生的事情。
    2. 不可能事件：在一定条件下，不然不会发生的事情。
    3. 随机现象：在基本条件不变的情况下，一系列试验或观察会得到不同的结果的现象。
    4. 随机事件：随机现象出现的结果。
2. 频率稳定性
    1. 频率：对于随机事件A，若在N次试验中出现了n次，则称$F_N(A)=\frac{n}{N}$为随机事件A在N次试验中出现的频率。
    2. 概率：对于随机事件A，表示该事件发生的可能性大小的数$P(A)$称为该事件的概率。
3. 频率与概率
    1. $F_N(A)\geq 0$
    2. 若记必然事件为Ω，则有$F_N(\Omega)=1$
    3. 若$A\cap B=\emptyset$，则$F_N(A\cup B)=F_N(A)+F_N(B)$

2.样本空间与事件

1. 样本空间
    1. 样本点：随机试验可能出现的结果称为样本点，记作ω。
    2. 样本空间：样本点全体构成样本空间，记作Ω。
    3. 样本空间的类型
        1. 有限个样本点
        2. 无穷可列个样本点
        3. 无穷不可列个样本点
        4. 三维空间
        5. 函数空间
2. 事件
    1. 事件：样本点的某个集合。
    2. 属于：$\omega\in S$
    3. 不属于：$\omega\overline{\in}S$
    4. 空集：$\emptyset$
3. 事件的运算
    1. 运算
        1. 包含：$A\subset B$或$B\supset A$
        2. 等价：$A=B$，当且仅当$A\subset B$与$B\subset A$同时成立
        3. 逆：$\overline{A}=\Omega-A$
        4. 交：$A\cap B$
        5. 并：$A\cup B$
        6. 差：$A-B$
    2. 性质
        1. 交换律：$A\cup B=B\cup A$，即$AB=BA$
        2. 结合律：$(A\cup B)\cup C=A\cup(B\cup C)$，即$(AB)C=A(BC)$

3. 分配律：$\begin{aligned}(A \cup B) \cap C = (A \cap C) \cup (B \cap C) \\ (A \cap B) \cup C = (A \cup C) \cap (B \cup C)\end{aligned}$

3. De Morgan定律（对偶原理）（适用于可列个事件）
    1. $\overline{A \cup B} = \overline{A} \cap \overline{B}$
    2. $\overline{A \cap B} = \overline{A} \cup \overline{B}$
4. > 本质上只需要"交与逆"或"并与逆"两种运算
   >
   > 1. $A \cup B = \overline{\overline{A} \cap \overline{B}}$
   > 2. $A \cap B = \overline{\overline{A} \cup \overline{B}}$
   > 3. $A - B = A \cap \overline{B} = \overline{\overline{A} \cup B}$

4. 有限样本空间
    1. $P(\omega_1) + \cdots + P(\omega_n) = 1$
    2. $P(\Omega) = 1$
    3. $0 \leq P(A) \leq 1$

3.古典概型

1. 模型与计算公式
    1. $P(\omega_k) = \frac{1}{n}$，其中$k = 1, \cdots, n$
    2. 若$A = \omega_{k_1} + \cdots + \omega_{k_m}$，则$P(A) = \frac{m}{n}$。
2. 基本的组合分析公式
    1. 两条原理：乘法原理与加法原理
    2. 排列
        1. 有放回的选取并排列：n^r
        2. 不放回的选取并排列：$A_n^r = n(n-1)\cdots(n-r+1) = \frac{n!}{(n-r)!}$
        3. 全排列：$n!$
        4. 圆排列：$(n-1)!$
    3. 组合
        1. 不放回的选取：$C_n^r = \binom{n}{r} = \frac{n(n-1)\cdots(n-r+1)}{r!} = \frac{n!}{r!(n-r)!}$
        2. 分为多个部分：$\frac{n!}{r_1!\cdots r_k!}$
        3. 有放回的选取：$\binom{n+r-1}{r}$
        4. 全错排：$a_n = n!(\frac{1}{0!} - \frac{1}{1!} + \frac{1}{2!} - \ldots + (-1)^n \frac{1}{n!})$
    4. 组合公式
        1. $\binom{n}{k} = \binom{n}{n-k}$
        2. $\binom{n}{0} + \cdots + \binom{n}{n} = 2^n$
        3. $\binom{a}{0}\binom{b}{n} + \cdots + \binom{a}{n}\binom{b}{0} = \binom{a+b}{n}$
        4. $\binom{-\alpha}{k} = (-1)^k \binom{\alpha+k-1}{k}$
3. 最大似然估计法：把概率$p(n)$看作未知参数n的函数，称为似然函数，再通过求其最大值而得到n的估计。
4. 概率的基本性质
    1. 非负性：$P(A) \geq 0$
    2. 规范性：$P(\Omega) = 1$
    3. 有限可加性：若$A_1, \cdots, A_n$两两不相容，则
       $P(A_1 + \cdots + A_n) = P(A_1) + \cdots + P(A_n)$

4.几何概率

1. $P(A_g)=\frac{g\text{的测度}}{\Omega\text{的测度}}$
2. 性质
    1. 非负性：$P(A)\geq 0$
    2. 规范性：$P(\Omega)=1$
    3. 可列可加性：若$A_1,A_2,\cdots$两两不相容，则$P(\sum_{n=1}^{\infty}A_n)=\sum_{n=1}^{\infty}P(A_n)$

5.概率空间

1. 事件域
    1. σ域：称Ω的一些满足以下性质的子集构成的集类$\mathscr{F}$为σ域。
        1. $\Omega\in\mathscr{F}$
        2. 若$A\in\mathscr{F}$，则$\overline{A}\in\mathscr{F}$。
        3. 若$A_n\in\mathscr{F},n=1,2,\cdots$，则$\bigcup_{n=1}^{\infty}A_n\in\mathscr{F}$。
    2. 事件域：若$\mathscr{F}$是由样本空间Ω的一些元素构成的一个σ域，则称之为事件域。
        1. $\mathscr{F}$中的元素称为事件。
        2. Ω称为必然事件。
        3. $\emptyset$称为不可能事件。
    3. 最小σ域：若给定Ω的一个非空集类$\mathscr{F}$，必存在唯一的满足以下性质的Ω上的σ域$\mathfrak{m}(\mathscr{F})$，称之为包含$\mathscr{F}$的最小$\sigma$域，亦称由$\mathscr{F}$产生的$\sigma$域。
        1. $\mathscr{F}\subset\mathfrak{m}(\mathscr{F})$
        2. 若存在σ域$\mathscr{G}$满足$\mathscr{F}\subset\mathscr{G}$，则$\mathfrak{m}(\mathscr{F})\subset\mathscr{G}$。
    4. 一维Borel点集：由一切形为$[a,b)$的有界左闭右开区间构成的集类所产生的σ域为一维Borel σ域，记作$\mathscr{B}_1$，称$\mathscr{B}_1$中的集为一维Borel点集。
2. 概率
    1. 概率：称定义在事件域$\mathscr{F}$上的满足以下性质的集合函数P称为概率。
        1. 非负性：对于任意$A\in\mathscr{F}$，$P(A)\geq 0$。
        2. 规范性：$P(\Omega)=1$
        3. 可列可加性或完全可加性：若$A_n\in\mathscr{F},n=1,2,\cdots$且两两不相容，则$P(\sum_{n=1}^{\infty}A_n)=\sum_{n=1}^{\infty}P(A_n)$。
    2. 性质
        1. $P(\emptyset)=0$
        2. 有限可加性：若$A_iA_j=\emptyset$，其中$i,j=1,\cdots,n$且$i\neq j$，则$P(\sum_{k=1}^{n}A_k)=\sum_{k=1}^{k}P(A_k)$。
        3. 对于任意$A\in\mathscr{F}$，$P(\overline{A})=1-P(A)$。
        4. 若$B\subset A$，则$P(A-B)=P(A)-P(B)$。
        5. $P(A\cup B)=P(A)+P(B)-P(A\cap B)$
        6. > Boole不等式：$P(A\cup B)\leq P(A)+P(B)$
           >
           > Bonferroni不等式：$P(A\cap B)\geq P(A)+P(B)-1$
3. 可列可加性与连续性
    1. 下连续：称$\mathscr{F}$上的集合函数P为下连续的，如果对于任何单调不减的集序列S_n成立$\lim_{n\to\infty}P(S_n)=P(\lim_{n\to\infty}S_n)$。
    2. 若P是$\mathscr{F}$上满足$P(\Omega)=1$的非负集合函数，则其具有可列可加性的充分必要条件是
        1. P是有限可加的。
        2. P是下连续的。
    3. > 概率是上连续且下连续的。

4. > $P(\sum_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} P(A_n)$

4. 概率空间：三元总体$(\Omega, \mathscr{F}, P)$为概率空间。

第二章：条件概率与统计独立性

1.条件概率，全概率公式，Bayes公式

1. 条件概率：
    1. 条件概率：设$(\Omega,\mathscr{F},P)$是一个概率空间，$B\in\mathscr{F}$，且$P(B)>0$，则对于任意$A\in\mathscr{F}$，记

    $$P(A|B)=\frac{P(AB)}{P(B)} \tag{1}$$

    并称$P(A|B)$为在事件B发生的条件下事件A发生的条件概率。
    2. 性质
        1. 非负性：$P(A|B)\geq 0$
        2. 规范性：$P(\Omega|B)=1$
        3. 可列可加性：若$A_1,A_2,\cdots$两两不相容，则$P(\sum_{k=1}^{\infty}A_k|B)=\sum_{k=1}^{\infty}P(A_k|B)$
        4. 乘法公式（乘法原理）：

        $$P(AB)=P(B)P(A|B) \tag{2}$$

        5. 推广的乘法公式：

        $$P(A_1\cdots A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1}) \tag{3}$$

2. 全概率公式：设事件$A_1,A_2,\cdots$是样本空间Ω的一个分割，亦称完备事件组，即$A_1,A_2,\cdots$两两不相容，且$\sum_{k=1}^{\infty}A_k=\Omega$，则对于事件$B=\sum_{k=1}^{\infty}A_kB$，存在全概率公式

$$P(B)=\sum_{k=1}^{\infty}P(A_k)P(B|A_k) \tag{4}$$

3. Bayes公式：若事件$A_1,A_2,\cdots$两两不相容，且对于事件$B=\sum_{k=1}^{\infty}A_kB$，存在Bayes公式

$$P(A_k|B)=\frac{P(A_k)P(B|A_k)}{\sum_{i=1}^{\infty}P(A_i)P(B|A_i)} \tag{5}$$

2.事件独立性

1. 两个事件的独立性：称事件A和B是统计独立的，如果满足

$$P(AB)=P(A)P(B) \tag{6}$$

2. 三个事件的独立性：称事件A,B,C是统计独立的，如果同时满足

$$\begin{cases}P(AB)=P(A)P(B)\\P(BC)=P(B)P(C)\\P(CA)=P(C)P(A)\\P(ABC)=P(A)P(B)P(C)\end{cases} \tag{7}$$

3. 推论
    1. 若事件A,B独立，且$P(B)>0$，则$P(A|B)=P(A)$。
    2. 若事件A,B独立，则事件$A,\overline{B}$独立。

4. 试验的独立性：记$\mathscr{A}_k$为第k次试验有关的事件全体。若对于任意的$A^{(k)} \in \mathscr{A}_k, k=1,\cdots,n$，称试验$E_1,\cdots.E_n$是相互独立的，如果成立

$$P(A^{(1)}\cdots P^{(n)}) = P(A^{(1)})\cdots P(A^{(n)}) \tag{8}$$

3.Bernoulli试验与直线上的随机游动

1. Bernoulli试验：只存在两种可能结果的试验称为Bernoulli试验。
2. n重Bernoulli试验满足
    1. 每次实验之多出现两个可能结果之一：A和$\overline{A}$。
    2. A在每次试验中出现的概率p保持不变。
    3. 各次试验相互独立。
3. 几何分布的无记忆性：在Bernoulli试验中，已知在前$n \in N$次试验中没有成功，则首次成功所在需要的时间满足几何分布。在离散型分布中，仅有几何分布满足此性质。
4. > 分赌注问题：甲、乙两个赌徒中止赌博，若甲再胜n场则可赢得赌注，乙再胜m场则可赢得赌注。甲在每局获胜的概率为p，则甲赢得赌注的概率为
   >
   > $$\begin{aligned} p_{\text{甲}} &= \sum_{k=0}^{m-1}\binom{n+k-1}{k}p^n(1-p)^k \qquad (9)\\ &= \sum_{k=n}^{\infty}\binom{m+k-1}{k}p^k(1-p)^m \qquad (10)\\ &= \sum_{k=n}^{n+m-1}\binom{n+m-1}{k}p^k(1-p)^{n+m-1-k} \qquad (11)\end{aligned}$$
5. > 直线上的随机游动：考虑x轴上的一个质点，规定其只能位于整数点，在$t=0$时刻，位于初始位置$x=a\in Z$，以后每隔单位时间，分别以概率p及$1-p$向正的或负的方向移动一个单位。
   >
   > 1. 无限制随机游动：若质点在$t=0$时刻从原点出发，质点在$t=n$时刻位于k的概率为
   >
   > $$p(n,k) = \begin{cases} \binom{n}{\frac{n+k}{2}}p^{\frac{n+k}{2}}(1-p)^{\frac{n-k}{2}}, & 2|n=2|k \\ 0, & 2|n \neq 2|k \end{cases} \tag{12}$$
   >
   > 2. 两端带有吸收壁的随机游动：若质点在$t=0$时刻从$x=n\in(a,b)$出发，而在$x=a\in Z$和$x=b\in Z$处各有一个吸收壁，质点碰到吸收壁后将不再运动，则质点最终在a点被吸收的概率为
   >
   > $$f_a(n) = \begin{cases} \frac{(\frac{1-p}{p})^n-(\frac{1-p}{p})^b}{(\frac{1-p}{p})^a-(\frac{1-p}{p})^b}, & p\neq\frac{1}{2} \\ \frac{n-b}{a-b}, & p=\frac{1}{2} \end{cases} \tag{13}$$
   >
   > 质点最终在b点被吸收的概率为
   >
   > $$f_b(n) = \begin{cases} \frac{(\frac{1-p}{p})^n-(\frac{1-p}{p})^a}{(\frac{1-p}{p})^b-(\frac{1-p}{p})^a}, & p\neq\frac{1}{2} \\ \frac{n-a}{b-a}, & p=\frac{1}{2} \end{cases} \tag{14}$$
6. 多项分布：n次重复独立试验且每次试验出现的可能结果为$A_1,\cdots,A_r$，而$P(A_k)=p_k\geq 0, k=1,\cdots,r$，且$p_1+\cdots+p_r=1$，则$n$次试验中$A_k$出现$n_k$次的概率为

$$P = \frac{n!}{n_1!\cdots n_r!}p_1^{n_1}\cdots p_r^{n_r} \tag{15}$$

其中$n_k\geq 0, k=1,\cdots,r$，且$n_1+\cdots+n_r=n$。

4.二项分布与Poisson分布

1. 函数$b(k;n,p)$关于k先递增后递减，且

$$b_{\max}(k;n,p)=b([(n+1)p];n,p) \tag{16}$$

2. Poisson定理：在独立实验中，以与n有关的常数p_n代表事件在实验中出现的概率。若$np_n\to\lambda$，则当$n\to\infty$时，

$$b(k;n,p)\to\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda} \tag{17}$$

第三章：随机变量与分布函数

1.随机变量与其分布

1. 随机变量
    1. 随机变量：设$\xi(\omega)$是定义在概率空间$(\Omega,\mathscr{F},P)$上的单值实函数，如果对于直线上任一Borel点集B，有

    $$\{\omega:\xi(\omega)\in B\}\in\mathscr{F} \tag{18}$$

    则称$\xi(\omega)$为随机变量。
    2. 概率分布：称$P\{\xi(\omega)\in B\}$为随机变量$\xi(\omega)$的概率分布。
    3. 分布函数：称

    $$F(x)=P\{\xi(\omega)<x\},-\infty<x<\infty \tag{19}$$

    为随机变量$\xi(\omega)$的分布函数。
    4. 分布函数的性质
        1. 单调性：若$a<b$，则$F(a)\le F(b)$。
        2. $\lim_{x\to-\infty}F(x)=0$
           $\lim_{x\to\infty}F(x)=1$
        3. 左连续性：$F(x^-)=F(x)$
2. 离散型随机变量
    1. 概率分布：设$\{x_k\}$为离散型随机变量ξ的所有可能值，$p(x_k)$是ξ取x_k的概率，即

    $$P\{\xi=x_k\}=p(x_k),k=1,2,\cdots \tag{20}$$

    $\{p(x_k),k=1,2,\cdots\}$称为随机变量ξ的概率分布，并满足性质

    $$p(x_k)\ge 0,k=1,2,\cdots \tag{21}$$

    $$\sum_{k=1}^{\infty}p(x_k)=1 \tag{22}$$

    2. 分布函数：

    $$F(x)=\sum_{x_k<x}p(x_k) \tag{23}$$

3. 连续型随机变量
    1. （分布）密度函数：

    $$p(x),x\in[a,b]\text{或}(-\infty,\infty) \tag{24}$$

    满足在区间$[a,b]$或$(-\infty,\infty)$上可积。
    2. 分布函数：

    $$F(x)=\int_{-\infty}^{x}p(t)\mathrm{d}t \tag{25}$$

    3. 性质
        1. $p(x)\ge 0$
        2. $\int_{-\infty}^{\infty}p(x)\mathrm{d}x=1$
        3. $P\{a\le\xi<b\}=F(b)-F(a)=\int_a^b p(x)\mathrm{d}x$
        4. $P\{\xi=c\}=0$

4. 正态分布$N(\mu,\sigma^2)$
    1. 密度函数：

$$p(x)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}},-\infty<x<\infty \tag{26}$$

    其中$\sigma>0$，μ与μ均为常数。

    2. 分布函数：

$$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}\mathrm{e}^{-\frac{(t-\mu)^2}{2\sigma^2}}\mathrm{d}t,-\infty<x<\infty \tag{27}$$

    3. 标准正态分布：$\mu=0,\sigma=1$，此时相应的分布密度函数以及分布函数分别记为$\varphi(x)$和$\Phi(x)$，即

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}},-\infty<x<\infty \tag{28}$$

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t,-\infty<x<\infty \tag{29}$$

    于是显然有

$$F(x)=\Phi(\frac{x-\mu}{\sigma}) \tag{30}$$

4. > 分布函数的性质
   >
   > 1. 分布函数至多仅有可列个不连续点。
   > 2. 对于分布函数$F(x)$存在Lebesgue分解
   >
   > $$F(x)=c_1F_1(x)+c_2F_2(x)+c_3F_3(x) \tag{31}$$
   >
   > 其中$F_1(x)$为跳跃函数，$F_1(x)$为绝对连续函数，$F_1(x)$为奇异函数。

2.随机向量，随机变量的独立性

1. 随机向量及其分布
    1. n维随机向量：若随机变量$\xi_1(\omega),\cdots,\xi_n(\omega)$定义在同一概率空间$(\Omega,\mathscr{F},P)$上，则称

$$\vec{\xi}(\omega)=(\xi_1(\omega),\cdots,\xi_n(\omega)) \tag{32}$$

    为n维随机向量。

    2. （联合）分布函数：称n元函数

$$F(x_1,\cdots,x_n)=P\{\xi_1(\omega)<x_1,\cdots,\xi_n(\omega)<x_n\} \tag{33}$$

    为随机向量$\vec{\xi}(\omega)=(\xi_1(\omega),\cdots,\xi_n(\omega))$的（联合）分布函数。

    3. 多元分布函数的性质
        1. 单调性：关于每个变元是单调不减函数。
        2. $F(x_1,\cdots,-\infty,\cdots,x_n)=0$
           $F(+\infty,\cdots,+\infty)=1$
        3. 左连续性：关于每个变元左连续。
    4. 密度函数：在连续型场合，存在函数$p(x_1,\cdots,x_n)$，使得成立

$$F(x_1,\cdots,x_n)=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n}p(t_1,\cdots,t_n)\mathrm{d}t_1\cdots\mathrm{d}t_n \tag{34}$$

这里的$p(x_1,\cdots,x_n)$称为（多元分布）密度函数，满足如下性质

$$p(x_1,\cdots,x_n)\geq 0 \tag{35}$$

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}p(t_1,\cdots,t_n)\mathrm{d}t_1\cdots\mathrm{d}t_n=1 \tag{36}$$

5. 多项分布

$$P\{\xi_1=k_1,\cdots,\xi_r=k_r\}=\frac{n!}{k_1!\cdots k_r!}p_1^{k_1}\cdots p_r^{k_r} \tag{37}$$

其中$p_1+\cdots+p_r=1$且$k_1+\cdots+k_r=n$。

6. 多元超几何分布

$$P\{\xi_1=n_1,\cdots,\xi_r=n_r\}=\frac{\binom{N_1}{n_1}\cdots\binom{N_r}{n_r}}{\binom{N}{n}} \tag{38}$$

其中$n_1+\cdots+n_r=n$。

7. 均匀分布

$$p(x_1,\cdots,x_n)=\begin{cases}\frac{1}{S}, & (x_1,\cdots,x_n)\in G\\ 0, & (x_1,\cdots,x_n)\overline{\in} G\end{cases} \tag{39}$$

8. 多元正态分布$N(\vec{\mu},\Sigma)$

$$p(\vec{x})=\frac{1}{(2\pi)^{\frac{n}{2}}(\det\Sigma)^{\frac{1}{2}}}\mathrm{e}^{-\frac{1}{2}(\vec{x}-\vec{\mu})\Sigma^{-1}(\vec{x}-\vec{\mu})^T} \tag{40}$$

其中Σ为n阶正定对称矩阵，$\vec{\mu}$为n阶实值行向量。

2. 边际分布（仅讨论二维场合）

1. 边际分布（边缘分布）：考虑二维随机向量(ξ,η)，设ξ取值$x_1,x_2,\cdots$；η取值$y_1,y_2,\cdots$。记

$$P\{\xi=x_i,\eta=y_j\}=p(x_i,y_j),\ i,j=1,2,\cdots \tag{41}$$

$$P\{\xi=x_i\}=p_1(x_i),\ i=1,2,\cdots \tag{42}$$

$$P\{\eta=y_j\}=p_2(y_j),\ j=1,2,\cdots \tag{43}$$

显然

$$p(x_i,y_j)\geq 0,\sum_{i,j}p(x_i,y_j)=1 \tag{44}$$

此外，

$$\sum_j p(x_i,y_j)=P\{\xi=x_i\}=p_1(x_i),\ i=1,2,\cdots \tag{45}$$

$$\sum_i p(x_i,y_j)=P\{\eta=y_j\}=p_2(y_j),\ j=1,2,\cdots \tag{46}$$

这里$\{p_1(x_i),i=1,2,\cdots\}$与$\{p_2(y_j),j=1,2,\cdots\}$称为$\{p(x_i,y_j),i,j=1,2,\cdots\}$的边际分布或边缘分布。

2. 边际分布函数：考虑二维随机向量(ξ,η)，其分布函数为$F(x,y)$，则

$$F_1(x)=F(x,+\infty) \tag{47}$$

$$F_2(y)=F(+\infty,y) \tag{48}$$

$F_1(x)$与$F_2(y)$称为$F(x,y)$的边际分布函数。

3. 边际（分布）密度函数：若$F(x,y)$为连续型分布函数，有密度函数$p(x,y)$，则

$$p_1(x)=\int_{-\infty}^{\infty}p(x,y)\mathrm{d}y \tag{49}$$

$$p_2(y)=\int_{-\infty}^{\infty}p(x,y)\mathrm{d}x \tag{50}$$

$p_1(x)$与$p_2(y)$称为$p(x,y)$的边际（分布）密度函数。

4. 二元正态分布$N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$

$$p(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\mathrm{e}^{-\frac{1}{2(1-\rho^2)}(\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2})} \tag{51}$$

其中μ_1,μ_2为两个边际分布的数学期望，σ_1,σ_2为两个边际分布的标准差,ρ为二元正态分布的相关系数，且构成协方差矩阵

$$\Sigma=\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix} \tag{52}$$

5. 二元正态分布密度函数的典型分解：二元正态分布$N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$存在如下两个分解

$$p(x,y)=\frac{1}{\sqrt{2\pi}\sigma_1}\mathrm{e}^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\cdot\frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}\mathrm{e}^{-\frac{(y-(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)))^2}{2\sigma_2^2(1-\rho^2)}} \tag{53}$$

$$=\frac{1}{\sqrt{2\pi}\sigma_2}\mathrm{e}^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}\cdot\frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}\mathrm{e}^{-\frac{(x-(\mu_1+\rho\frac{\sigma_1}{\sigma_2}(x-\mu_2)))^2}{2\sigma_1^2(1-\rho^2)}} \tag{54}$$

显然，第一式的第一部分为$N(\mu_1,\sigma_1)$的密度函数，第二部分为$N(\mu_2+\rho\frac{\sigma_1}{\sigma_2}(x-\mu_1),\sigma_2^2(1-\rho^2))$的密度函数；第二式的第一部分为$N(\mu_2,\sigma_2)$的密度函数，第二部分为$N(\mu_1+\rho\frac{\sigma_2}{\sigma_1}(y-\mu_2),\sigma_1^2(1-\rho^2))$的密度函数。

6. 二元正态分布的边际分布：对于二元正态分布(80)，其边际（分布）密度函数为

$$p_1(x)=\frac{1}{\sqrt{2\pi}\sigma_1}\mathrm{e}^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \tag{55}$$

$$p_2(y)=\frac{1}{\sqrt{2\pi}\sigma_2}\mathrm{e}^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}} \tag{56}$$

因此二元正态分布的边际分布仍未正态分布。

3. 条件分布（仅讨论二维场合）

1. 离散型随机变量：考虑二维随机向量(ξ,η)，设ξ取值$x_1,x_2,\cdots$；η取值$y_1,y_2,\cdots$，则随机变量η关于随机变量ξ的条件分布为

$$P\{\eta=y_j|\xi=x_i\}=\frac{p(x_i,y_j)}{p_1(x_i)} \tag{57}$$

2. 连续型随机变量：考虑二维随机向量(ξ,η)，其分布函数为$F(x,y)$，则随机变量η关于随机变量ξ的条件分布为

$$P\{\eta<y|\xi=x\}=\lim_{\Delta x\to 0}\frac{F(x+\Delta x,y)-F(x,y)}{F(x+\Delta x,\infty)-F(x,\infty)} \tag{58}$$

$$=\lim_{\Delta x\to 0}\frac{\int_x^{x+\Delta x}\int_{-\infty}^{y}p(u,v)\mathrm{d}u\mathrm{d}v}{\int_x^{x+\Delta x}\int_{-\infty}^{\infty}p(u,v)\mathrm{d}u\mathrm{d}v} \tag{59}$$

若$p_1(x)\neq 0$，则

$$p(y|x)=\frac{p(x,y)}{p_1(x)} \tag{60}$$

4. 随机变量的独立性
   1. 离散型随机变量：设$\xi_1,\cdots,\xi_n$为n个随机变量，若对于任意的$x_1,\cdots,x_n$，成立

      $$P\{\xi_1=x_1,\cdots,\xi_n=x_n\}=P\{\xi_1=x_1\}\cdots P\{\xi_n=x_n\} \tag{61}$$

      则称$\xi_1,\cdots,\xi_n$是相互独立的。
   2. 设$\xi_1,\cdots,\xi_n$为n个随机变量，若对于任意的$x_1,\cdots,x_n$，成立

      $$P\{\xi_1<x_1,\cdots,\xi_n<x_n\}=P\{\xi_1<x_1\}\cdots P\{\xi_n<x_n\} \tag{62}$$

      则称$\xi_1,\cdots,\xi_n$是相互独立的。若ξ_k的分布函数为$F_k(x),k=1,\cdots,n$，则(91)等价于

      $$F(x_1,\cdots,x_n)=F_1(x_1)\cdots F_n(x_n) \tag{63}$$

3.随机变量的函数及其分布

1. Borel函数与随机变量的函数
   1. n元Borel（可测）函数：设$y=g(x_1,\cdots,x_n)$是$\mathbb{R}^n\to\mathbb{R}^1$映射，若对于任意Borel点集$B_1\subset\mathbb{R}^1$均有

      $$\{(x_1,\cdots,x_n):g(x_1,\cdots,x_n)\in B_1\}\in\mathscr{B}_n \tag{64}$$

      其中$\mathscr{B}_n$为$\mathbb{R}^n$上Borelσ域，则称$g(x_1,\cdots,x_n)$为n元Borel（可测）函数。同时，若$(\xi_1,\cdots,\xi_n)$是$(\Omega,\mathscr{F},P)$上的随机变量，而$g(x_1,\cdots,x_n)$是n元Borel（可测）函数，则$g(\xi_1,\cdots,\xi_n)$是$(\Omega,\mathscr{F},P)$上的随机变量。
   2. 离散卷积公式：

      $$c_n=\sum_{k=0}^{n}a_kb_{n-k} \tag{65}$$

2. 单个随机变量的函数的分布律
   1. 目标：已知随机变量ξ的分布函数$F(x)$或密度函数$p(x)$，求解$\eta=g(\xi)$的分布函数$G(y)$或密度函数$q(y)$，即

      $$G(y)=\int_{g(x)<y}p(x)\mathrm{d}x \tag{66}$$

   2. 若$g(x)$严格单调，其反函数$g^{-1}(y)$存在连续导函数，则$\eta=g(\xi)$具有密度函数

      $$q(y)=p(g^{-1}(y))|(g^{-1}(y))'| \tag{67}$$

   3. > 随机变量的存在性定理：若$F(x)$是左连续的单调不减函数，且$F(-\infty)=0,F(\infty)=1$，则存在概率空间$(\Omega,\mathscr{F},P)$及其上的随机变量$\xi(\omega)$，使$\xi(\omega)$的分布函数恰好为$F(x)$。

   4. 倍数分布：随机变量ξ的分布函数为$F(x)$，密度函数为$p(x)$，则$\eta=c\xi,c\neq0$的分布函数为

      $$G(y)=\begin{cases}F(\frac{y}{c}), & c>0\\ 1-F(\frac{y}{c}), & c<0\end{cases} \tag{68}$$

      密度函数为

      $$q(y)=\frac{p(\frac{y}{c})}{|c|} \tag{69}$$

5. 平方分布：随机变量ξ的分布函数为$F(x)$，密度函数为$p(x)$，则$\eta=\xi^2$的分布函数为

$$G(y)=\begin{cases}F(\sqrt{y})-F(-\sqrt{y}), & y>0\\ 0, & y\le 0\end{cases} \tag{70}$$

密度函数为

$$q(y)=\begin{cases}\frac{p(\sqrt{y})+p(-\sqrt{y})}{2\sqrt{y}}, & y>0\\ 0, & y<0\end{cases} \tag{71}$$

3. 随机变量的函数的分布律

1. 若$\eta=g(\xi_1,\cdots,\xi_n)$，而$(\xi_1,\cdots,\xi_n)$的密度函数为$p(x_1,\cdots,x_n)$，则

$$G(y)=\int\cdots\int_{g(x_1,\cdots,x_n)<y}p(x_1,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n \tag{72}$$

2. 和的分布：若$\eta=\xi_1+\xi_2$，(ξ_1,ξ_2)的密度函数为$p(x_1,x_2)$，则

$$G(y)=\int_{-\infty}^{\infty}\int_{-\infty}^{y-x_1}p(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2 \tag{73}$$

即

$$G(y)=\int_{-\infty}^{y}\int_{-\infty}^{\infty}p(u,v-u)\mathrm{d}u\mathrm{d}v \tag{74}$$

$$=\int_{-\infty}^{y}\int_{-\infty}^{\infty}p(v-u,u)\mathrm{d}u\mathrm{d}v \tag{75}$$

因此η的密度函数为

$$q(y)=\int_{-\infty}^{\infty}p(u,y-u)\mathrm{d}u \tag{76}$$

$$=\int_{-\infty}^{\infty}p(y-u,u)\mathrm{d}u \tag{77}$$

此为卷积公式。

3. 商的分布：若$\eta=\frac{\xi_1}{\xi_2}$，(ξ_1,ξ_2)的密度函数为$p(x_1,x_2)$，则

$$G(x)=\int_0^{\infty}(\int_{-\infty}^{zx}p(y,z)\mathrm{d}y)\mathrm{d}z+\int_{-\infty}^{0}(\int_{zx}^{\infty}p(y,z)\mathrm{d}y)\mathrm{d}z \tag{78}$$

因此η的密度函数为

$$q(x)=\int_{-\infty}^{\infty}|z|p(zx,z)\mathrm{d}z \tag{79}$$

4. 顺序统计量的分布：若$\xi_1,\cdots,\xi_n$是相互独立的随机变量，具有相同的分布函数和密度函数$p(x)$，记其极小值为ξ_1^*，极大值为ξ_n^*

    1. 极小值ξ_1^*

        1. 密度函数：

$$q(x)=np(x)(1-F(x))^{n-1} \tag{80}$$

        2. 分布函数：

$$G(x)=1-(1-F(x))^n \tag{81}$$

    2. 极大值ξ_n^*

        1. 密度函数：

$$q(x)=np(x)(F(x))^{n-1} \tag{82}$$

2. 分布函数:

$$G(x)=(F(x))^n \tag{83}$$

3. 联合分布

1. 密度函数:

$$q(x,y)=\begin{cases}0, & x\geq y\\ n(n-1)q(x)q(y)(F(y)-F(x))^{n-2}, & x<y\end{cases} \tag{84}$$

2. 分布函数:

$$G(x,y)=\begin{cases}(F(y))^n, & x\geq y\\ (F(y))^n-(F(y)-F(x))^n, & x<y\end{cases} \tag{85}$$

4. 随机变量的变换若$(\xi_1,\cdots,\xi_n)$的密度函数为$p(x_1,\cdots,x_n)$，变量$\eta_1,\cdots,\eta_n$满足$\eta_1=g_1(\xi_1,\cdots,\xi_n),\cdots,\eta_n=g_n(\xi_1,\cdots,\xi_n)$，且对于$y_k=g_k(x_1,\cdots,x_n),k=1,\cdots,n$存在唯一的反函数$x_k=x_k(y_1,\cdots,y_n),k=1,\cdots,n$，同时$(\eta_1,\cdots,\eta_n)$的密度函数为$q(y_1,\cdots,y_n)$，则

$$q(y_1,\cdots,y_n)=p(x_1(y_1,\cdots,y_n),\cdots,x_n(y_1,\cdots,y_n))|J| \tag{86}$$

其中J为Jacobi行列式

$$J=\begin{vmatrix}\frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n}\\ \vdots & \ddots & \vdots\\ \frac{\partial x_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n}\end{vmatrix} \tag{87}$$

5. 随机变量的函数的独立性：若$\xi_1,\cdots,\xi_n$是相互独立的随机变量，则$f_1(\xi_1),\cdots,f_n(\xi_n)$也是相互独立的，这里$f_k,k=1,\cdots,n$是任意的一元Borel函数。

第四章：数字特征与特征函数

1.数学期望

1. 平均值与加权平均值
    1. 平均值：
    $$\overline{x}=\frac{\sum_{k=1}^{n}x_k}{n} \tag{88}$$
    2. 加权平均值：
    $$\overline{x}_w=\sum_{k=1}^{n}w_kx_k \tag{89}$$
    其中$\omega_k\geq 0,k=1,\cdots,n$，且$\sum_{k=1}^{n}w_k=1$
2. 离散型随机变量的数学期望：设ξ为离散型随机变量，且$P\{\xi=x_k\}=p_k,k=1,2,\cdots$，若级数
$$\sum_{k=1}^{\infty}x_kp_k \tag{90}$$
绝对收敛，则称其为ξ的数学期望，记作$E\xi$。
3. 连续型随机变量的数学期望：设ξ为具有密度函数$p(x),-\infty<x<\infty$的连续型随机变量，若积分
$$\int_{-\infty}^{\infty}xp(x)\mathrm{d}x \tag{91}$$
绝对收敛，则称其为ξ的数学期望，记作$E\xi$。
4. 数学期望
    1. 若ξ的分布函数为$F(x)$，且积分
    $$\int_{-\infty}^{\infty}x\mathrm{d}F(x) \tag{92}$$
    绝对收敛，则定义
    $$E\xi=\int_{-\infty}^{\infty}x\mathrm{d}F(x) \tag{93}$$
    为ξ的数学期望。
    2. Stieltjes积分：
    $$I=\int_{-\infty}^{\infty}g(x)\mathrm{d}F(x) \tag{94}$$
    性质
        1. 当$F(x)$为跳跃函数，且在$x_k(k=1,2,\cdots)$具有跃度p_k时，积分化为无穷级数
        $$I=\sum_{k}g(x_k)p_k \tag{95}$$
        2. 当$F(x)$存在导数$p(x)$时，积分化为
        $$I=\int_{-\infty}^{\infty}g(x)p(x)\mathrm{d}x \tag{96}$$

3. 线性性质

$$\int_{-\infty}^{\infty}(ag_1(x)+bg_2(x))\mathrm{d}F(x)=a\int_{-\infty}^{\infty}g_1(x)\mathrm{d}F(x)+b\int_{-\infty}^{\infty}g_2(x)\mathrm{d}F(x) \tag{97}$$

$$\int_{-\infty}^{\infty}g(x)\mathrm{d}(aF_1(x)+bF_2(x))=a\int_{-\infty}^{\infty}g(x)\mathrm{d}F_1(x)+b\int_{-\infty}^{\infty}g(x)\mathrm{d}F_2(x) \tag{98}$$

4. $$\int_a^b g(x)\mathrm{d}F(x)=\int_a^c g(x)\mathrm{d}F(x)+\int_c^b g(x)\mathrm{d}F(x),a\le c\le b \tag{99}$$

5. 若$g(x)\ge 0$，$F(x)$单调不减，$b>a$，则

$$\int_a^b g(x)\mathrm{d}F(x)\ge 0 \tag{100}$$

5. 随机变量函数的数学期望
   1. 若ξ是分布函数为$F(x)$的随机变量，$g(x)$是一元Borel函数，则随机变量$\eta=g(\xi)$的数学期望为

$$E\eta=\int_{-\infty}^{\infty}g(x)\mathrm{d}F(x) \tag{101}$$

   2. 若ξ是分布函数为$F_\xi(x)$的随机变量，$\eta=g(\xi)$是分布函数为$F_\eta(x)$的随机变量，其中$g(x)$是一元Borel函数，则成立

$$\int_{-\infty}^{\infty}y\mathrm{d}F_\eta(y)=\int_{-\infty}^{\infty}g(x)\mathrm{d}F_\xi(x) \tag{102}$$

6. 随机向量的数学期望
   1. 随机向量$(\xi_1,\cdots,\xi_n)$的数学期望为$(E\xi_1,\cdots,E\xi_n)$，其中

$$E\xi_k=\int_{-\infty}^{\infty}x_k\mathrm{d}F_k(x_k),k=1,\cdots,n \tag{103}$$

   这里$F_k(x_k)$是ξ_k的分布函数。
   2. 若随机向量$(\xi_1,\cdots,\xi_n)$的的分布函数为$F(x_1,\cdots,x_n)$，$g(x_1,\cdots,x_n)$为n元Borel函数，则

$$Eg(\xi_1,\cdots,\xi_n)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}g(x_1,\cdots,x_n)\mathrm{d}F(x_1,\cdots,x_n) \tag{104}$$

7. 数学期望的性质
   1. 若$a\le\xi\le b$，则$a\le E\xi\le b$。
   2. 线性性质：

$$E(c)=c \tag{105}$$
$$E(c\xi)=cE\xi \tag{106}$$
$$E(\xi+\eta)=E\xi+E\eta \tag{107}$$

2.方差，相关系数，矩

1. 方差
   1. 方差：对于随机变量ξ，若$E(\xi-E\xi)^2$存在，则称其为随机变量ξ的方差，记为$D\xi$。
   2. 标准差：随机变量的方差的方根，即$\sqrt{D\xi}$。
   3. $$D\xi=E\xi^2-(E\xi)^2 \tag{108}$$
   4. 标准化随机变量：对于随机变量ξ，若其数学期望$E\xi$及方差$D\xi$均存在，且$D\xi>0$，则可标准化为

$$\xi^* = \frac{\xi - E\xi}{D\xi} \tag{109}$$

此时，$E\xi^* = 0$，$D\xi^* = 1$。

5. 性质
    1. 非线性性质

$$D(c) = 0 \tag{110}$$
$$D(c\xi) = c^2 D\xi \tag{111}$$
$$D(\xi + \eta) = D\xi + D\eta + 2(E\xi\eta - E\xi \cdot E\eta) \tag{112}$$

    2. $D\xi \le E(\xi - c)^2$,当且仅当$E\xi = c$时等号成立。
6. Чебышёв不等式：若随机变量ξ存在数学期望$E\xi$及方差$D\xi$，则成立不等式

$$P\{|\xi - E\xi| \ge \varepsilon\} \le \frac{D\xi}{\varepsilon^2} \tag{113}$$

其中ε为任一正数。

2. 相关系数
    1. 协方差
        1. 协方差：对于随机变量ξ和η，其协方差为

$$\operatorname{cov}(\xi, \eta) = E((\xi - E\xi)(\eta - E\eta)) \tag{114}$$

可计算为

$$\operatorname{cov}(\xi, \eta) = E\xi\eta - E\xi \cdot E\eta \tag{115}$$

        2. 双线性性质

$$\operatorname{cov}(\xi, \eta) = \operatorname{cov}(\eta, \xi) \tag{116}$$
$$\operatorname{cov}(\xi, c) = 0 \tag{117}$$
$$\operatorname{cov}(\xi, c\eta) = c\operatorname{cov}(\xi, \eta) \tag{118}$$
$$\operatorname{cov}(\xi, \eta + \zeta) = \operatorname{cov}(\xi, \eta) + \operatorname{cov}(\xi, \zeta) \tag{119}$$

        3. 协方差矩阵：记$\sigma_{ij} = \operatorname{cov}(\xi_i, \xi_j), i, j = 1, \cdots, n$，则随机向量$\xi = (\xi_1, \cdots, \xi_n)$的协方差矩阵为

$$\Sigma = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{pmatrix} \tag{120}$$

记作$D\xi$。显然此为非负定对称矩阵。

    2. 相关系数：称

$$\rho = \begin{cases} \frac{\operatorname{cov}(\xi,\eta)}{\sqrt{D\xi}\sqrt{D\eta}}, & \xi\text{与}\eta\text{均不为常数} \\ 0, & \xi\text{或}\eta\text{为常数} \end{cases} \tag{121}$$

        2. > 对于事件A和B，定义其相关系数为
           >
           > $$\rho = \frac{P(AB) - P(A)P(B)}{\sqrt{P(A)P(\overline{A})P(B)P(\overline{B})}} \tag{122}$$

        3. > 相关系数是随机变量间线性关系的度量。

        4. 性质

$$\rho(\xi, \eta) = \rho(\eta, \xi) \tag{123}$$
$$\rho(\xi, c\eta) = \operatorname{sgn}(c)\rho(\xi, \eta) \tag{124}$$

3. Cauchy-Schwarz不等式：对任意随机变量ξ和η均有

$$(E\xi\eta)^2 \leq E\xi^2 \cdot E\eta^2 \tag{125}$$

当且仅当

$$P\{a\xi = b\eta\} = 1 \tag{126}$$

时等号成立，其中a, b为常数。

4. 相关性
    1. 记随机变量ξ和η的相关系数为ρ
        1. $\rho > 0$：正相关
        2. $\rho < 0$：负相关
        3. $\rho = 1$：完全正相关
        4. $\rho = -1$：完全负相关
        5. $\rho = 0$：不相关
    2. 对于相关系数ρ，成立

    $$|\rho| \leq 1 \tag{127}$$

    并且，$\rho = 1$当且仅当

    $$P\{\frac{\xi - E\xi}{\sqrt{D\xi}} = \frac{\eta - E\eta}{\sqrt{D\eta}}\} = 1 \tag{128}$$

    $\rho = -1$当且仅当

    $$P\{\frac{\xi - E\xi}{\sqrt{D\xi}} + \frac{\eta - E\eta}{\sqrt{D\eta}} = 0\} = 1 \tag{129}$$

    3. 对于随机变量ξ和η，以下等价
        1. ξ和η不相关。
        2. $\rho = 0$
        3. $\mathrm{cov}(\xi, \eta) = 0$
        4. $E\xi\eta = E\xi \cdot E\eta$
        5. $D(\xi + \eta) = D\xi + D\eta$
    4. 若随机变量ξ和η独立，则ξ和η不相关；反之不然。
    5. 对于二元正态分布，不相关性与独立性是等价的。
    6. 对于二值随机变量，不相关性与独立性是等价的。

3. 矩
    1. 原点矩：对于$n \in N^*$，称

    $$m_n = E\xi^n \tag{130}$$

    为n阶原点矩。

    2. 中心矩：对于$n \in N^*$，称

    $$c_n = E(\xi - E\xi)^n \tag{131}$$

    为n阶中心矩。

    3. 由于$|\xi|^{n-1} \leq 1 + |\xi|^n$，则若$n$阶矩存在，则$n-1$阶矩存在。
    4. $$c_n = \sum_{k=0}^{n} \binom{n}{k} (-m_1)^{n-k} m_k \tag{132}$$
    5. $$m_n = \sum_{k=0}^{n} \binom{n}{k} c_{n-k} m_1^k \tag{133}$$

4. 分位数
    1. p位分数：对于$0 < p < 1$，若
    $$F(x_p) \leq p \leq F(x_p^+) \tag{134}$$
    则称x_p为分布函数$F(x)$的p位分数。
    2. 中位数：$x_{0.5}$成为中位数。
5. 条件数学期望，最佳线性预测
    1. 条件数学期望：在$\xi = x$的条件下，η的条件数学期望定义为
    $$E\{\eta|\xi = x\} = \int_{-\infty}^{\infty} yp(y|x)\mathrm{d}y \tag{135}$$
    2. 最小二乘法：对于相依的随机变量ξ和η，要求寻找两者的函数关系$\eta = h(\xi)$，使得均方误差
    $$E(\eta - h(\xi)) \tag{136}$$
    达到最小，此时取
    $$h(x) = E\{\eta|\xi = x\} \tag{137}$$
    3. 回归：称
    $$y = E\{\eta|\xi = x\} \tag{138}$$
    为η关于ξ的回归。
    4. 重期望公式：
    $$E\eta = E(E\eta|\xi) \tag{139}$$
    5. 线性回归：若ξ和η的数学期望分别为μ_1和μ_2，标准差分别为σ_1和σ_2及相关系数为ρ，η关于ξ的线性回归为
    $$L(x) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1) \tag{140}$$
    其均方误差为
    $$E(\eta - L(\xi)^2) = \sigma_2^2(1 - \rho^2) \tag{141}$$
    6. > 预测值$\hat{\eta} = L(\xi)$与残差$\eta - \hat{\eta}$是不相关的。

3.熵与信息

1. 不确定性与熵
    1. Shannon公式：
    $$H(p_1, \cdots, p_n) = -C\sum_{k=1}^{n} p_k \ln p_k \tag{142}$$
    其中C为正常数且$p_1 + \cdots + p_n = 1$。
    2. 熵：对于随机试验α，其产生的不相容的结果的概率为$p_1, \cdots, p_n$，满足$p_1 + \cdots + p_n = 1$，则其熵为
    $$H(\alpha) = -\sum_{k=1}^{n} p_k \ln p_k \tag{143}$$
2. 熵的基本性质

1. Jensen不等式：若$f(x)$为$[a,b]$熵的上凸函数，对于任意的$x_1,\cdots,x_n\in[a,b]$及正数$\lambda_1,\cdots,\lambda_n$且$\lambda_1+\cdots+\lambda_n=1$，则成立不等式

$$\sum_{k=1}^{n}\lambda_k f(x_k)\le f(\sum_{k=1}^{n}\lambda_k x_k) \tag{144}$$

当且仅当$x_1=\cdots=x_n$时等号成立。

2. $$-\sum_{k=1}^{n}p_k\ln p_k\ge 0 \tag{145}$$

当且仅当存在k使得$p_k=1$时等号成立。

3. $$-\sum_{k=1}^{n}p_k\ln p_k\le \ln n \tag{146}$$

当且仅当$p_1=\cdots=p_n=\frac{1}{n}$时等号成立。

4. 若试验α与β独立，则

$$H(\alpha\beta)=H(\alpha)H(\beta) \tag{147}$$

3. 条件熵与信息量
   1. 条件熵：在试验α实现的条件下试验β的条件熵为

$$H_\alpha(\beta)=\sum_{k=1}^{n}p_k H_k(\beta) \tag{148}$$

   2. 信息量：试验α中的有关试验β的信息量为

$$I(\alpha,\beta)=H(\beta)-H(\alpha) \tag{149}$$

   3. 性质
      1. 熵的加法法则：$H(\alpha\beta)=H(\alpha)+H_\alpha(\beta)$
      2. 熵的非负性：$H_\alpha(\beta)\ge 0$
      3. $H_\alpha(\beta)\le H(\beta)$
4. 连续型分布的熵
   1. 对于随机变量α和β，其密度函数分别为$p(x)$和$q(y)$，且其联合密度函数为$f(x,y)$，则可定义

$$H(\alpha)=-\int_{-\infty}^{\infty}p(x)\ln p(x)\mathrm{d}x \tag{150}$$

$$H(\beta)=-\int_{-\infty}^{\infty}q(y)\ln q(y)\mathrm{d}y \tag{151}$$

$$H(\alpha\beta)=-\iint f(x,y)\ln f(x,y)\mathrm{d}x\mathrm{d}y \tag{152}$$

$$H_\alpha(\beta)=-\iint f(x,y)\ln\frac{f(x,y)}{p(x)}\mathrm{d}x\mathrm{d}y \tag{153}$$

$$H_\beta(\alpha)=-\iint f(x,y)\ln\frac{f(x,y)}{q(y)}\mathrm{d}x\mathrm{d}y \tag{154}$$

   2. 性质
      1. 若α限制于V，则$H(\alpha)\le\ln|V|$，当且仅当α为V中的均匀分布，其中$|V|$为V的测度。
      2. $H(\alpha\beta)=H(\alpha)+H_\alpha(\beta)=H(\beta)+H_\beta(\alpha)$
      3. $H(\alpha\beta)\le H(\alpha)+H(\beta)$，当且仅当试验$\alpha$与$\beta$独立时等号成立。

4. 设$p(x)$为一元密度函数，其标准差为σ，则其熵$H(x)$满足$H(x)\le \ln\sqrt{2\pi \mathrm{e}}\sigma$，当且仅当$p(x)$为正态分布时等号成立。
5. 设$p(x)$为一元密度函数，且当$x\le 0$时，$p(x)=0$，并且其数学期望为λ，则其熵$H(x)$满足$H(x)\le \ln \mathrm{e}\lambda$，当且仅当$p(x)$为指数分布$\mathrm{Exp}(\frac{1}{\lambda})$时等号成立。

4.母函数

1. 母函数：对于随机变量ξ，其概率分布为$P\{\xi=n\}=p_n, n=0,1,\cdots$，则$\xi$的母函数定义为

$$P(s)=\sum_{n=0}^{\infty} p_n s^n \tag{155}$$

2. 母函数的性质
   1. 唯一性：母函数与概率分布函数一一对应。
   2. 数字特征：

$$E\xi = P'(1) \tag{156}$$

$$D\xi = P''(1)+P'(1)-(P'(1))^2 \tag{157}$$

3. 独立随机变量之和的母函数：若随机变量ξ和η相互独立，其相应的母函数分别为$A(s)$和$B(s)$，则随机变量$\xi+\eta$的母函数$C(s)$满足

$$C(s)=A(s)B(s) \tag{158}$$

4. 随机个随机变量之和的母函数：若$\xi_1,\xi_2,\cdots$是相互独立的具有相同概率分布$P\{\xi_i=j\}=f_j$的随机变量，其母函数为

$$F(s)=\sum_{n=0}^{\infty} f_n s^n \tag{159}$$

随机变量ν是取整数值的，且$P\{\nu=n\}=g_n$，其母函数为

$$G(s)=\sum_{n=0}^{\infty} g_n s^n \tag{160}$$

若$\{\xi_n\}$与ν独立，则随机变量$\eta=\xi_1+\cdots+\xi_\nu$的母函数为

$$G(F(s)) \tag{161}$$

5.特征函数

1. 定义
   1. 复随机变量：若ξ和η均为概率空间$(\Omega,\mathscr{F},P)$熵的实值随机变量，则称$\zeta=\xi+i\eta$为复随机变量。
   2. 特征函数：若随机变量ξ的分布函数为$F(x)$，则其特征函数为

$$f(t)=E\mathrm{e}^{it\xi}=\int_{-\infty}^{\infty}\mathrm{e}^{itx}\mathrm{d}F(x) \tag{162}$$

   3. 对于连续型随机变量，若其分布密度函数为$p(x)$，则其特征函数为

$$f(t)=\int_{-\infty}^{\infty}\mathrm{e}^{itx}p(x)\mathrm{d}x \tag{163}$$

   此时，特征函数为密度函数$p(x)$的Fourier变换。

2. 性质

   1. $f(0)=1$

      $|f(t)|\leq f(0)$

      $f(-t)=\overline{f(t)}$

   2. 特征函数$f(t)$在$(-\infty,\infty)$上一致连续。

   3. 非负定性：对于任意$n\in N^*$以及任意实数$t_1,\cdots,t_n$及复数$\lambda_1,\cdots,\lambda_n$，成立

$$\sum_{i,j=1}^{n}f(t_i-t_j)\lambda_i\overline{\lambda_j}\geq 0 \tag{164}$$

   4. 若随机变量ξ和η相互独立，且对应的特征函数分别为$f_\xi(t)$和$f_\eta(t)$，则其和$\xi+\eta$的特征函数$f_{\xi+\eta}(t)$满足

$$f_{\xi+\eta}(t)=f_\xi(t)f_\eta(t) \tag{165}$$

   5. 若随机变量ξ存在n阶矩，则其特征函数可微分n次，且当$k\leq n$时，成立

$$f^{(k)}(0)=i^kE\xi^k \tag{166}$$

   6. 若$\eta=a\xi+b$，则

$$f_\eta(t)=\mathrm{e}^{ibt}f_\xi(at) \tag{167}$$

3. 逆转公式与唯一性定理

   1. 逆转公式：若分布函数$F(x)$的特征函数为$f(t)$，又x_1,x_2是$F(x)$的连续点，则成立逆转公式

$$F(x_2)-F(x_1)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\mathrm{e}^{-itx_1}-\mathrm{e}^{-itx_2}}{it}f(t)\mathrm{d}t \tag{168}$$

   2. 唯一性定理：分布函数$F(x)$由特征函数$f(t)$唯一确定，且

$$F(x)=\frac{1}{2\pi}\lim_{y\to\infty}\int_{-\infty}^{\infty}\frac{\mathrm{e}^{-ity}-\mathrm{e}^{-itx}}{it}f(t)\mathrm{d}t \tag{169}$$

      特别的，若$f(t)$绝对可积，则相应的分布函数$F(x)$的存在并连续，且

$$F'(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{e}^{-itx}f(t)\mathrm{d}t \tag{170}$$

4. 分布函数的可再生性

5. 多元特征函数

   1. 多元特征函数：若随机向量$(\xi_1,\cdots,\xi_n)$的分布函数为$F(x_1,\cdots,x_n)$，则其特征函数定义为

$$f(t_1,\cdots,t_n)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\mathrm{e}^{i(t_1x_1+\cdots+t_nx_n)}\mathrm{d}F(x_1,\cdots,x_n) \tag{171}$$

   2. 逆转公式：若随机向量$(\xi_1,\cdots,\xi_n)$的特征函数为$f(t_1,\cdots,t_n)$，且$F(x_1,\cdots,x_n)$为其分布函数，则

$$P\{a_k\leq\xi_k<b_k,k=1,\cdots,n\}=\frac{1}{(2\pi)^n}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\prod_{k=1}^{n}\frac{\mathrm{e}^{-ia_kt_k}-\mathrm{e}^{-ib_kt_k}}{it_k} \tag{172}$$

      其中a_k与b_k均为任意实数，满足$(\xi_1,\cdots,\xi_n)$落在几何体

$$a_k\leq x_k<b_k,k=1,\cdots,n \tag{173}$$

      的表面上的概率为零。

3. 唯一性定理：分布函数$F(x_1,\cdots,x_n)$由其特征函数$f(t_1,\cdots,t_n)$唯一确定。
4. 连续性定理：若特征函数列$\{f_k(t_1,\cdots,t_n)\}$收敛于一个连续函数$f(t_1,\cdots,t_n)$，则存在某分布函数使其特征函数为$f(t_1,\cdots,t_n)$。
5. 性质
    1. $f(t_1,\cdots,t_n)$在$\mathbb{R}^n$中一直连续，且

$$|f(t_1,\cdots,t_n)|\le f(0,\cdots,0)=1 \tag{174}$$

$$f(-t_1,\cdots,-t_n)=\overline{f(t_1,\cdots,t_n)} \tag{175}$$

    2. 若$f(t_1,\cdots,t_n)$为随机向量$(\xi_1,\cdots,\xi_n)$的特征函数，则随机变量$\eta=a_1t_1+\cdots+a_nt_n$的特征函数为

$$f_\eta(t)=f(a_1t,\cdots,a_nt) \tag{176}$$

    3. 若矩$E\xi_1^{k_1}\cdots\xi_n^{k_n}$存在，则

$$E\xi_1^{k_1}\cdots\xi_n^{k_n}=i^{-(k_1+\cdots+k_n)}\frac{\partial^{k_1+\cdots+k_n}f(t_1,\cdots,t_n)}{\partial t_1^{k_1}\cdots\partial t_n^{k_n}}|_{t_1=\cdots=t_n=0} \tag{177}$$

    4. 若随机向量$(\xi_1,\cdots,\xi_n)$的特征函数为$f(t_1,\cdots,t_n)$，则$k<n$维向量$(\xi_1,\cdots,\xi_k)$的特征函数为

$$f_{1,\cdots,k}(t_1,\cdots,t_k)=f(t_1,\cdots,t_k,0,\cdots,0) \tag{178}$$

    5. 若随机向量$(\xi_1,\cdots,\xi_n)$的特征函数为$f(t_1,\cdots,t_n)$，而ξ_k的特征函数为$f_k(t),k=1,\cdots,n$，则随机变量$\xi_1,\cdots,\xi_n$相互独立的充分必要条件是

$$f(t_1,\cdots,t_n)=f_1(t_1)\cdots f_n(t_n) \tag{179}$$

    6. 若以$f_1(t_1,\cdots,t_n),f_2(s_1,\cdots,s_m)$及$f(t_1,\cdots,t_n,s_1,\cdots,s_n)$分别记作为随机向量$(\xi_1,\cdots,\xi_n),(\eta_1,\cdots,\eta_m)$及$(\xi_1,\cdots,\xi_n,\eta_1,\cdots,\eta_m)$特征函数，则$(\xi_1,\cdots,\xi_n)$与$(\eta_1,\cdots,\eta_m)$独立的充分必要条件是对于任意实数$t_1,\cdots,t_n,s_1,\cdots,s_m$成立

$$f(t_1,\cdots,t_n,s_1,\cdots,s_n)=f_1(t_1,\cdots,t_n)f_2(s_1,\cdots,s_m) \tag{180}$$

6.多元正态分布

1. 密度函数与特征函数
    1. n元正态分布：若

$$\vec{\mu}=\begin{pmatrix}\mu_1\\ \vdots\\ \mu_n\end{pmatrix} \tag{181}$$

为n维实向量，

$$\Sigma=\begin{pmatrix}\sigma_{11} & \cdots & \sigma_{1n}\\ \vdots & \ddots & \vdots\\ \sigma_{n1} & \cdots & \sigma_{nn}\end{pmatrix} \tag{182}$$

为n阶非负定对称矩阵，则以

$$f(\vec{t})=\mathrm{e}^{i\vec{\mu}^T\vec{t}-\frac{1}{2}\vec{t}^T\Sigma\vec{t}} \tag{183}$$

为特征函数的分布，或以

$$p(\vec{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det\Sigma}}\mathrm{e}^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T\Sigma^{-1}(\vec{x}-\vec{\mu})} \tag{184}$$

为密度函数的分布为n元正态分布，记为$N(\vec{\mu},\Sigma)$。若n维随机向量

$$\vec{\xi} = \begin{pmatrix}\xi_1\\ \vdots\\ \xi_n\end{pmatrix} \tag{185}$$

满足n元正态分布，记为$\vec{\xi} \sim N(\vec{\mu},\Sigma)$。

2. $\vec{\xi}$的任一子向量

$$\tilde{\vec{\xi}}_m = \begin{pmatrix}\xi_{k_1}\\ \vdots\\ \xi_{k_m}\end{pmatrix}, m \le n \tag{186}$$

服从正态分布$N(\tilde{\vec{m}}u_m,\tilde{\Sigma}_m)$，其中

$$\tilde{\vec{m}}u_m = \begin{pmatrix}\mu_{k_1}\\ \vdots\\ \mu_{k_m}\end{pmatrix} \tag{187}$$

$\tilde{\Sigma}_m$为Σ保留第$k_1,\cdots,k_m$行及列所得的m阶矩阵。

3. $\vec{\mu}$及Σ分别为随机向量$\vec{\xi}$的数学期望矩阵及协方差矩阵，即

$$\mu_k = E\xi_k, k = 1,\cdots,n \tag{188}$$

$$\sigma_{ij} = E(xi_i - E\xi_i)(xi_j - E\xi_j), i,j = 1,\cdots,n \tag{189}$$

2. 独立性

1. 若$\vec{\xi} \sim N(\vec{\mu},\Sigma)$，则$\xi_1,\cdots,\xi_n$相互独立的充分必要条件是其两两不相关。
2. 若

$$\vec{\xi} = \begin{pmatrix}\vec{\xi}_1\\ \vec{\xi}_2\end{pmatrix} \tag{190}$$

其中$\vec{\xi}_1$与$\vec{\xi}_2$为$\vec{\xi}$的子向量，记

$$\Sigma = \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix} \tag{191}$$

其中Σ_{11}及Σ_{22}分别为$\vec{\xi}_1$及$\vec{\xi}_2$的协方差矩阵，$\Sigma_{12} = {\Sigma_{21}}^T$为由$\vec{\xi}_1$与$\vec{\xi}_2$的相应分量的协方差构成的相互协方差矩阵，则$\vec{\xi}_1$与$\vec{\xi}_2$独立的充分必要条件是$\Sigma_{12} = \Sigma_{21} = \vec{0}$。

3. 线性变换

1. $\vec{\xi} \sim N(\vec{\mu},\Sigma) \Leftrightarrow \vec{l}^T\vec{\xi} \sim N(\vec{l}^T\vec{\mu},\vec{l}^T\Sigma\vec{l})$，其中$\vec{l}$为任意实向量。
2. 正态分布的线性变换不变性：若$\vec{\xi} \sim N(\vec{\mu},\Sigma)$，则对于任意矩阵$C$，成立$C\vec{\xi} \sim N(C\vec{\mu},C\Sigma CT^))$。
3. 推论
   1. 若$\vec{\xi} \sim N(\vec{\mu},\Sigma)$，则存在正交矩阵$U$，使得$\vec{\eta} = U\vec{\xi}$为具有独立正态分布的随机变量，且$E\vec{\eta} = U\vec{\mu}$，$D\vec{\eta}$的方差分量为$\Sigma$的特征值。
   2. 正交变换下，多维正态变量保持其独立、同方差性不变。
   3. 若$\vec{\xi} \sim N(\vec{\mu},\Sigma)$，且$\Sigma$为$n$阶正定矩阵，则

$$(\vec{\xi}-\vec{\mu})^T\Sigma^{-1}(\vec{\xi}-\vec{\mu})\sim\chi_n^2 \tag{192}$$

4. 条件分布：若$\vec{\xi}\sim N(\vec{\mu},\Sigma)$，记

$$\vec{\xi}=\begin{pmatrix}\vec{\xi}_1\\ \vec{\xi}_2\end{pmatrix} \tag{193}$$

其中$\vec{\xi}_1$与$\vec{\xi}_2$为$\vec{\xi}$的子向量，且$E\vec{\xi}_1=\vec{\mu}_1,E\vec{\xi}_2=\vec{\mu}_2$；记

$$\Sigma=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix} \tag{194}$$

其中Σ_{11}及Σ_{22}分别为$\vec{\xi}_1$及$\vec{\xi}_2$的协方差矩阵，$\Sigma_{12}=\Sigma_{21}{}^T$为由$\vec{\xi}_1$与$\vec{\xi}_2$的相应分量的协方差构成的相互协方差矩阵。则在给定$\vec{\xi}_1=\vec{x}_1$时，$\vec{\xi}_2$的条件分布亦为正态分布，其条件数学期望为

$$m\vec{u}_{2\cdot 1}=\vec{\mu}_2+\Sigma_{21}\Sigma_{11}^{-1}(\vec{x}_1-\vec{\mu}_1) \tag{195}$$

条件方差为

$$\Sigma_{22\cdot 1}=\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \tag{196}$$

此时$E(x\vec{i}_2|\vec{\xi}_1=\vec{x}_1)$成为$\vec{\xi}_2$关于$\vec{\xi}_1$的回归，且其为$\vec{x}_1$的线性函数，又条件方差$\Sigma_{22\cdot 1}$与$\vec{x}_1$无关。

第五章：极限定理

1.Bernoulli试验场合的极限定理

1. 大数定律与中心极限定理
    1. 大数定律：若$\xi_1, \xi_2, \cdots$是随机变量序列，令

    $$\eta_n = \frac{\xi_1 + \cdots + \xi_n}{n} \tag{197}$$

    若存在常数序列$a_1, a_2, \cdots$，对于任意的$\varepsilon > 0$，恒成立

    $$\lim_{n\to\infty} P\{|\eta_n - a_n| < \varepsilon\} = 1 \tag{198}$$

    则称序列$\{\eta_n\}$服从大数定律。
    2. 中心极限定理：对于独立随机变量$\xi_1, \xi_2, \cdots$，若$E\xi_k$及$D\xi_k$存在，对随机变量进行标准化

    $$\zeta_n = \frac{\sum_{k=1}^n \xi_k - \sum_{k=1}^n E\xi_k}{\sqrt{\sum_{k=1}^n D\xi_k}} \tag{199}$$

    若$\{\zeta_n\}$满足

    $$\lim_{n\to\infty} P\{\zeta_n < x\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-\frac{t^2}{2}} \mathrm{d}t \tag{200}$$

    则称序列$\{\zeta_n\}$服从中心极限定理。
2. 大数定律
    1. Чебышёв大数定律：若$\xi_1, \xi_2, \cdots$是两两不相关的随机变量所构成的序列，每一随机变量都存在有限的方差，且有公共上界

    $$D\xi_n \le C, n = 1, 2, \cdots \tag{201}$$

    则对于任意的$\varepsilon > 0$，恒成立

    $$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{k=1}^n \xi_k - \frac{1}{n}\sum_{k=1}^n E\xi_k| \le \varepsilon\} = 1 \tag{202}$$

    2. Марков大数定律：对于随机变量$\xi_1, \xi_2, \cdots$，若满足Марков条件，即当$n \to \infty$时

    $$\frac{1}{n^2} D(\sum_{k=1}^n \xi_k) \to 0 \tag{203}$$

    则对于任意的$\varepsilon > 0$，恒成立

    $$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{k=1}^n \xi_k - \frac{1}{n}\sum_{k=1}^n E\xi_k| \le \varepsilon\} = 1 \tag{204}$$

    3. Bernoulli大数定律：若μ_n是n次Bernoulli试验中某事件出现的次数，而p是该事件在每次试验中出现的概率，则对于任意的$\varepsilon > 0$，恒成立

    $$\lim_{n\to\infty} P\{|\frac{\mu_n}{n} - p| < \varepsilon\} = 1 \tag{205}$$

    4. Poisson大数定律：若在一个独立实验序列中，某事件在第k次试验中出现的概率为p_k，以μ_n记前n次试验中该事件出现的次数，则对于任意的$\varepsilon > 0$，恒成立

$$\lim_{n\to\infty} P\{|\frac{\mu_n}{n} - \frac{p_1+\cdots+p_n}{n}| < \varepsilon\} = 1 \tag{206}$$

3. De Moivre—Laplace极限定理

   1. 局部极限定理：若μ_n是n次Bernoulli试验中某事件出现的次数，$0<p<1$，$[a,b]$为任意有限区间，记$x_k=\frac{k-np}{\sqrt{np(1-p)}}$，且满足$a\le x_k\le b$，则当$n\to\infty$时

   $$P(\mu_n=k)\sim\frac{1}{\sqrt{np(1-p)}}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x_k^2}{2}} \tag{207}$$

   2. 积分极限定理：若μ_n是n次Bernoulli试验中某事件出现的次数，$0<p<1$，则对任意有限区间$[a,b]$，当$n\to\infty$时

   $$P\{a\le\frac{\mu_n-np}{\sqrt{np(1-p)}}\le b\}\to\frac{1}{\sqrt{2\pi}}\int_a^b\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x \tag{208}$$

   3. 应用

      1. 用频率估计概率：

      $$P\{|\frac{\mu_n}{n}-p|<\varepsilon\}\approx\frac{2}{\sqrt{2\pi}}\int_{-\infty}^{\varepsilon\sqrt{\frac{n}{p(1-p)}}}\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x-1 \tag{209}$$

      2. 二项式分布计算：

      $$\binom{n}{k}p^k(1-p)^{n-k}\approx\frac{1}{\sqrt{np(1-p)}}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}(\frac{k-np}{\sqrt{np(1-p)}})^2} \tag{210}$$

      3. 二项分布计算：

      $$P\{a\le\mu_n\le b\}\approx\frac{1}{\sqrt{2\pi}}\int_{\frac{a-np}{\sqrt{np(1-p)}}}^{\frac{b-np}{\sqrt{np(1-p)}}}\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x \tag{211}$$

2.收敛性

1. 分布函数弱收敛

   1. 弱收敛：对于分布函数列$\{F_n(x)\}$，若存在单调非减函数$F(x)$使

   $$\lim_{n\to\infty}F_n(x)=F(x) \tag{212}$$

   在$F(x)$的每一连续点上都成立，则称$F_n(x)$弱收敛于$F(x)$，记作$F_n(x)\xrightarrow{\mathrm{W}}F(x)$。

   2. Helly第一定律：任一一致有界的单调非减函数列$\{F_n(x)\}$存在子序列$\{F_{n_k}(x)\}$弱收敛于一有界的单调非减函数$F(x)$。
   3. Helly第二定律：若$f(x)$是$[a,b]$上的连续函数，又$\{F_n(x)\}$是在$[a,b]$上弱收敛于函数$F(x)$的一致有界单调非减函数序列，且a和b是$F(x)$的连续点，则

   $$\lim_{n\to\infty}\int_a^b f(x)\mathrm{d}F_n(x)=\int_a^b f(x)\mathrm{d}F(x) \tag{213}$$

   4. 拓广的Helly第二定律：若$f(x)$是$(-\infty,\infty)$上的连续有界函数，又$\{F_n(x)\}$是在$(-\infty,\infty)$上弱收敛于函数$F(x)$的一致有界单调非减函数序列，且

   $$\lim_{n\to\infty}F_n(-\infty)=F(-\infty),\lim_{n\to\infty}F_n(\infty)=F(\infty) \tag{214}$$

   则

$$\lim_{n\to\infty}\int_{-\infty}^{\infty}f(x)\mathrm{d}F_n(x)=\int_{-\infty}^{\infty}f(x)\mathrm{d}F(x) \tag{215}$$

2. Lévy—Cramer定理（连续性定理）
    1. 正极限定理：若分布函数列$\{F_n(x)\}$弱收敛于分布函数$F(x)$，则相应的特征函数列$\{f_n(t)\}$收敛于特征函数$f(t)$，且在t的任一有限区间内收敛是一致的。
    2. 逆极限定理：若特征函数列$\{f_n(t)\}$收敛于函数$f(t)$，且$f(t)$在$t=0$处连续，则相应的分布函数列$\{F_n(x)\}$弱收敛于以$f(t)$为特征函数的分布函数$F(x)$。
3. 随机变量的收敛性
    1. 依分布收敛：对于随机变量$\xi_n(\omega)$和$\xi(\omega)$的分布函数分别为$F_n(x)$和$F(x)$，若$F_n(x)\xrightarrow{\mathrm{W}}F(x)$，则称$\{\xi_n(\omega)\}$依分布收敛于$\xi(\omega)$，记作$\xi_n(\omega)\xrightarrow{\mathrm{L}}\xi(\omega)$。
    2. 依概率收敛：对于随机变量$\xi_n(\omega)$和$\xi(\omega)$，若对于任意的$\varepsilon>0$，恒成立
    $$\lim_{n\to\infty}P\{|\xi_n(\omega)-\xi(\omega)|\geq\varepsilon\}=0 \tag{216}$$
    则称$\{\xi_n(\omega)\}$依概率收敛于$\xi(\omega)$，记作$\xi_n(\omega)\xrightarrow{\mathrm{P}}\xi(\omega)$。
    3. r阶收敛：对于随机变量$\xi_n(\omega)$和$\xi(\omega)$，存在有限的$E|\xi_n|^r$和$E|\xi|^r$，其中$r>0$为常数，若
    $$\lim_{n\to\infty}E|\xi_n-\xi|^r=0 \tag{217}$$
    则称$\{\xi_n(\omega)\}r$阶收敛于$\xi(\omega)$，记作$\xi_n(\omega)\xrightarrow{r}\xi(\omega)$。
    4. 以概率1收敛：对于随机变量$\xi_n(\omega)$和$\xi(\omega)$，若
    $$P\{\lim_{n\to\infty}\xi_n(\omega)=\xi(\omega)\}=1 \tag{218}$$
    则称$\{\xi_n(\omega)\}$以概率1收敛于$\xi(\omega)$，又称$\{\xi_n(\omega)\}$以几乎处处收敛于$\xi(\omega)$，记作$\xi_n(\omega)\xrightarrow{\mathrm{a.s.}}\xi(\omega)$。
    5. $\xi_n(\omega)\xrightarrow{r}\xi(\omega)\Rightarrow\xi_n(\omega)\xrightarrow{\mathrm{P}}\xi(\omega)\Rightarrow\xi_n(\omega)\xrightarrow{\mathrm{L}}\xi(\omega)$
    6. $\xi_n(\omega)\xrightarrow{\mathrm{a.s.}}\xi(\omega)\Rightarrow\xi_n(\omega)\xrightarrow{\mathrm{P}}\xi(\omega)$
    7. 若C为常数，则$\xi_n(\omega)\xrightarrow{\mathrm{P}}\xi(\omega)\Leftrightarrow\xi_n(\omega)\xrightarrow{\mathrm{L}}\xi(\omega)$。
4. Bochner—Хйнчин定理
    1. Bochner—Хйнчин定理：函数$f(t)$是特征函数的充分必要条件是：$f(t)$非负定，连续，且$f(0)=1$。
    2. Herglotz定理：复数列$C_n,n=0,\pm1,\cdots$可以表示为
    $$C_n=\int_{-\pi}^{\pi}\mathrm{e}^{inx}\mathrm{d}G(x) \tag{219}$$
    的充分必要条件是其为非负定的，即对于任意的正整数n及复数$\lambda_1,\lambda_2,\cdots$,均有
    $$\sum_{i,j=1}^{n}C_{i-j}\lambda_i\overline{\lambda_j}\geq0 \tag{220}$$
    其中$G(x)$是$[-\pi,\pi]$上有界、单调非减、左连续函数。

3.独立同分布场合的极限定理

1. Хйнчин大数定律：若$\xi_1,\xi_2,\cdots$是相互独立的随机变量序列，服从相同的分布，且具有有限的数学期望
$$a=E\xi_n \tag{221}$$

则对于任意的$\varepsilon > 0$，有

$$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{k=1}^{n}\xi_k - a| < \varepsilon\} = 1 \tag{222}$$

> 矩估计的相合性：若总体ξ的k阶原点矩$m_n = E\xi^k$存在，此时样本的k阶原点矩
>
> $$A_k = \frac{1}{n}\sum_{i=1}^{n}\xi_i^k \tag{223}$$
>
> 作为m_k的估计量成立
>
> $$A_k \xrightarrow{\mathrm{P}} m_k \tag{224}$$

2. 中心极限定理

    1. Lindeberg—Lévy中心极限定理：若$\xi_1, \xi_2, \cdots$是相互独立的服从相同分布的随机变量序列，且

    $$E\xi_n = \mu, D\xi_n = \sigma^2, n = 1, 2, \cdots \tag{225}$$

    其中$0 < \sigma^2 < \infty$，则其标准化随机变量

    $$\zeta_n = \frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{n}(\xi_k - \mu) \tag{226}$$

    的极限分布为

    $$\lim_{n\to\infty} P\{\zeta_n < x\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t \tag{227}$$

    > 若$m_{2k} = E\xi_n^{2k}$存在，则
    >
    > $$A_k = \frac{1}{n}\sum_{i=1}^{n}\xi_i^k \tag{228}$$
    >
    > 的分布渐进于
    >
    > $$N(m_k, \frac{m_{2k} - m_k^2}{n}) \tag{229}$$

    2. 若$\xi_1, \xi_2, \cdots$是相互独立的服从相同分布的随机变量序列，且

    $$E\xi_n = \mu, D\xi_n = \sigma^2, n = 1, 2, \cdots \tag{230}$$

    其中$0 < \sigma^2 < \infty$，则

    $$P\{a \le \sum_{k=1}^{n}\xi_k \le b\} \approx \frac{1}{\sqrt{2\pi}}\int_{\frac{a-n\mu}{\sigma\sqrt{n}}}^{\frac{b-n\mu}{\sigma\sqrt{n}}} \mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t \tag{231}$$

    3. 多元中心极限定理：若m维随机变量$\vec{\xi_1}, \vec{\xi_2}, \cdots$相互独立且具有相同的分布，其数学期望为$\vec{\mu}$，协方差矩阵为$\Sigma$，则其标准化随机变量

    $$\vec{\eta_n} = \frac{(\vec{\xi_1} - \vec{\mu}) + \cdots + (\vec{\xi_n} - \vec{\mu})}{\sqrt{n}} \tag{232}$$

    的极限分布为

    $$N(\vec{0}, \Sigma) \tag{233}$$

4.强大数定律

1. 以概率1收敛
    1. 极限事件
        1. 上限事件与下限事件：若$A_1, A_2, \cdots$为一列事件，记

$$\overline{\lim_{n\to\infty}} A_n = \bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty} A_n \tag{234}$$

$$\underline{\lim_{n\to\infty}} A_n = \bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty} A_n \tag{235}$$

        称$\overline{\lim}_{n\to\infty} A_n$为事件序列$\{A_n\}$的上限事件，$\underline{\lim}_{n\to\infty} A_n$为事件序列$\{A_n\}$的下限事件。

        2. 极限事件：若$A_1, A_2, \cdots$为一列事件，显然

$$\overline{\lim_{n\to\infty}} A_n \supset \underline{\lim_{n\to\infty}} A_n \tag{236}$$

        特别的，当$\overline{\lim}_{n\to\infty} A_n = \underline{\lim}_{n\to\infty} A_n$时，记$\lim_{n\to\infty} A_n \equiv \overline{\lim}_{n\to\infty} A_n = \underline{\lim}_{n\to\infty} A_n$，并称其为事件序列$\{A_n\}$的极限事件。

        3. > 由de Morgan定理
           >
           > $$\underline{\lim_{n\to\infty}} \overline{A_n} = \overline{\overline{\lim_{n\to\infty}} A_n} \tag{237}$$
           >
           > $$\overline{\overline{\lim_{n\to\infty}} A_n} = \overline{\underline{\lim_{n\to\infty}} A_n} \tag{238}$$

    2. Borel—Cantelli引理
        1. 若随机事件序列$\{A_n\}$满足

$$\sum_{n=1}^{\infty} P(A_n) < \infty \tag{239}$$

        则

$$P\{\overline{\lim_{n\to\infty}} A_n\} = 0, P\{\underline{\lim_{n\to\infty}} \overline{A_n}\} = 1 \tag{240}$$

        2. 若$\{A_n\}$是相互独立的随机事件序列，则

$$\sum_{n=1}^{\infty} P(A_n) = \infty \tag{241}$$

        成立的充分必要条件为

$$P\{\overline{\lim_{n\to\infty}} A_n\} = 0 \text{或} P\{\underline{\lim_{n\to\infty}} \overline{A_n}\} = 1 \tag{242}$$

2. 强大数定律：对于独立随机变量$\{\xi_n\}$，若

$$P\{\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n}(\xi_k - E\xi_k) = 0\} = 1 \tag{243}$$

    则称其满足强大数定律。

3. Borel强大数定律：若μ_n为某事件在n次独立试验中出现的次数，且在每次试验中该事件出现的概率均为p，则当$n \to \infty$时，成立

$$P\{\frac{\mu_n}{n}\to p\}=1 \tag{244}$$

4. Колмогóров强大数定律
    1. Колмогóров不等式：对于独立随机变量$\xi_1,\xi_2,\cdots$，若其方差有限，则对于任意的$\varepsilon>0$，成立

    $$P\{\max_{1\le k\le n}|\sum_{i=1}^{k}(\xi_i-E\xi_i)|\ge\varepsilon\}\le\frac{1}{\varepsilon^2}\sum_{k=1}^{n}D\xi_k \tag{245}$$

    2. Hājek—Rényi不等式：对于独立随机变量序列$\{\xi_n\}$，$D\xi_n=\sigma_n^2<\infty,n=1,2,\cdots$，而$\{C_n\}$是一列正的单调非增常数序列，则对于任意正整数$m<n$，及$\varepsilon>0$，均有

    $$P\{\max_{m\le k\le n}|\sum_{i=1}^{k}(\xi_i-E\xi_i)|\ge\varepsilon\}\le\frac{1}{\varepsilon^2}(C_m^2\sum_{i=1}^{m}\sigma_i^2+\sum_{i=m+1}^{n}C_i^2\sigma_i^2) \tag{246}$$

    3. Колмогóров强大数定律：对于独立随机变量序列$\{\xi_n\}$，若$\sum_{n=1}^{\infty}\frac{D\xi_n}{n^2}<\infty$，则成立

    $$P\{\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}(\xi_k-E\xi_k)=0\}=1 \tag{247}$$

5. 独立同分布场合的强大数定律：对于相互独立的服从相同分布的随机变量序列$\{\xi_n\}$，当$n\to\infty$时

    $$\frac{1}{n}\sum_{k=1}^{n}\xi_k\xrightarrow{\text{a.s.}}a \tag{248}$$

    成立的充分必要条件是$E\xi_n,n=1,2,\cdots$存在且为a。

5.中心极限定理

1. 对于相互独立的服从相同分布的随机变量序列$\{\xi_n\}$，其具有有限的数学期望和方差

    $$a_n=E\xi_n,b_n^2=D\xi_n,n=1,2,\cdots \tag{249}$$

    记

    $$B_n^2=\sum_{k=1}^{n}b_k^2 \tag{250}$$

    作标准化和数

    $$\zeta_n=\sum_{k=1}^{n}\frac{\xi_k-a_k}{B_n} \tag{251}$$

    本节即为寻找和数ζ_n的分布函数区域正态分布函数的充分必要条件。

2. Lindeberg条件：对于任意$\tau>0$，成立

    $$\lim_{n\to\infty}\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{|x-a_k|>\tau B_n}(x-a_k)^2\mathrm{d}F_k(x)=0 \tag{252}$$

    其中$F_n(x)$为随机变量ξ_n的分布函数。

3. Felelr条件：

    $$\lim_{n\to\infty}\max_{k\le n}\frac{b_k}{B_n}=0 \tag{253}$$

    $$\Leftrightarrow\lim_{n\to\infty}B_n=\infty\text{且}\lim_{n\to\infty}\frac{b_n}{B_n}=0 \tag{254}$$

4. Lindeberg—Felelr定理：对于和数ζ_n，成立

$$\lim_{n\to\infty} P\{\zeta_n < x\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-\frac{t^2}{2}} \mathrm{d}t \tag{255}$$

与Felelr条件的充分必要条件是成立Lindeberg条件。

5. 推论

   1. 对于相互独立随机变量序列$\{\xi_n\}$，若存在常数K_n，使得成立

      $$\max_{1\le k\le n} |\xi_k| \le K_n, n = 1, 2, \cdots \tag{256}$$

      且

      $$\lim_{n\to\infty} \frac{K_n}{B_n} = 0 \tag{257}$$

      则当$n \to \infty$时，成立

      $$P\{\sum_{k=1}^{n} \frac{\xi_k - a_k}{B_n} < x\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-\frac{t^2}{2}} \mathrm{d}t \tag{258}$$

   2. Lyapunov定理：若对于独立随机变量序列$\{\xi_n\}$，存在$\delta > 0$，使得当$n \to \infty$时，

      $$\frac{1}{B_n^{2+\delta}} \sum_{k=1}^{n} E|\xi_k - a_k|^{2+\delta} \to 0 \tag{259}$$

      则成立

      $$P\{\sum_{k=1}^{n} \frac{\xi_k - a_k}{B_n} < x\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-\frac{t^2}{2}} \mathrm{d}t \tag{260}$$

附录：概率模型

概率模型	概率分布$p(x)$	数学期望$E\xi$	方差$D\xi$	特征函数$f(t)$
退化分布$I_c(x)$	$p(x)=\begin{cases}1, & x=c\\ 0, & x\neq c\end{cases}$	c	0	e^{ict}
Bernoulli分布	$p(x)=\begin{cases}1-p, & x=0\\ p, & x=1\end{cases}$ $0<p<1$	p	$p(1-p)$	$p\mathrm{e}^{it}+1-p$
二项分布$B(n,p)$	$b(k;n,p)=\binom{n}{k}p^k(1-p)^{n-k}$ $k=0,\cdots,n;0<p<1$	np	$np(1-p)$	$(p\mathrm{e}^{it}+1-p)^n$
Poisson分布$P(\lambda)$	$p(k;\lambda)=\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}$ $k=0,1,\cdots;\lambda>0$	λ	λ	$\mathrm{e}^{\lambda(\mathrm{e}^{it}-1)}$
几何分布	$g(k;p)=p(1-p)^{k-1}$ $k=1,2,\cdots;0<p<1$	$\frac{1}{p}$	$\frac{q}{p^2}$	$\frac{p\mathrm{e}^{it}}{1-(1-p)\mathrm{e}^{it}}$
超几何分布	$p_k=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$ $M,n\leq N;M,N,n\in N^*$ $k=0,\cdots,\min(M,n)$	$\frac{nM}{N}$	$\frac{nM(N-M)(N-n)}{N^2(N-1)}$	$\sum_{k=0}^{n}\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}\mathrm{e}^{ikt}$
Pascal分布	$p_k=\binom{k-1}{r-1}p^r(1-p)^{k-r}$ $k=r,r+1,\cdots;0<p<1;r\in N^*$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$	$(\frac{(1-p)\mathrm{e}^{it}}{1-(1-p)\mathrm{e}^{it}})^r$
负二项分布	$p_k=\binom{-r}{k}p^r(p-1)^k$ $k=0,1,\cdots;0<p<1;r>0$	$\frac{r(1-p)}{p}$	$\frac{r(1-p)}{p^2}$	$(\frac{p}{1-(1-p)\mathrm{e}^{it}})^r$
正态分布$N(\mu,\sigma^2)$	$p(x)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $-\infty<x<\infty$	μ	σ^2	$\mathrm{e}^{i\mu t-\frac{1}{2}\sigma^2t^2}$
均匀分布$U[a,b]$	$p(x)=\begin{cases}\frac{1}{b-a}, & a\leq x\leq b\\ 0, & \text{其他}\end{cases}$ $a<b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{\mathrm{e}^{ibt}-\mathrm{e}^{iat}}{i(b-a)t}$
指数分布$\mathrm{Exp}(\lambda)$	$p(x)=\begin{cases}\lambda\mathrm{e}^{-\lambda x}, & x\geq 0\\ 0, & x<0\end{cases}$ $\lambda>0$	λ^{-1}	λ^{-2}	$(1-\frac{it}{\lambda})^{-1}$
χ^2分布	$p(x)=\begin{cases}\frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}x^{\frac{n}{2}-1}\mathrm{e}^{-\frac{x}{2}}, & x\geq 0\\ 0, & x<0\end{cases}$ $n\in N^*$	n	$2n$	$(1-2it)^{-\frac{n}{2}}$
Γ分布$\Gamma(r,\lambda)$（$r\in N^*$时为Erlang分布）	$p(x)=\begin{cases}\frac{\lambda^r}{\Gamma(r)}x^{r-1}\mathrm{e}^{-\lambda x}, & x\geq 0\\ 0, & x<0\end{cases}$ $r,\lambda>0$	$\frac{r}{\lambda}$	$\frac{r}{\lambda^2}$	$(1-\frac{it}{\lambda})^{-r}$
Cauchy分布	$p(x)=\frac{1}{\pi}\frac{\lambda}{\lambda^2+(x-\mu)^2}$ $-\infty<x,\mu<\infty;\lambda>0$	不存在	不存在	$\mathrm{e}^{i\mu t-\lambda\lvert t\rvert}$
t分布	$p(x)=\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(1+\frac{x^2}{n})^{-\frac{n+1}{2}}$ $-\infty<x<\infty;n\in N^*$	$0(n>1)$	$\frac{n}{n-2}(n>2)$	
Pareto分布	$p(x)=\begin{cases}rA^r\frac{1}{x^{r+1}}, & x\geq A\\ 0, & x<A\end{cases}$ $r,A>0$	($r>1$时存在)	($r>2$时存在)	
F分布	$p(x)=\begin{cases}\frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})}m^{\frac{m}{2}}n^{\frac{n}{2}}\frac{x^{\frac{m}{2}-1}}{(n+mx)^{\frac{m+n}{2}}}, & x\geq 0\\ 0, & x<0\end{cases}$ $m,n\in N^*$	$\frac{n}{n-2}(n>2)$	$\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}(n>4)$	
β分布	$p(x)=\begin{cases}\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}x^{p-1}(1-x)^{q-1}, & 0<x<1\\ 0, & \text{其他}\end{cases}$ $p,q>0$	$\frac{p}{p+q}$	$\frac{pq}{(p+q)^2(p+q+1)}$	$\frac{\Gamma(p+q)}{\Gamma(p)}\sum_{k=0}^{\infty}\frac{\Gamma(p+k)(it)^k}{\Gamma(p+q+k)\Gamma(k+1)}$
对数正态分布	$p(x)=\begin{cases}\frac{1}{\sqrt{2\pi}\sigma x}\mathrm{e}^{\frac{(\ln x-\alpha)^2}{2\sigma^2}}, & x>0\\ 0, & x\leq 0\end{cases}$ $\alpha,\sigma>0$	$\mathrm{e}^{\alpha+\frac{\sigma^2}{2}}$	$\mathrm{e}^{2\alpha+\sigma^2}(\mathrm{e}^{\sigma^2}-1)$	
Weibull分布	$p(x)=\begin{cases}\alpha\lambda x^{\alpha-1}\mathrm{e}^{-\lambda x^\alpha}, & x>0\\ 0, & x\leq 0\end{cases}$ $\lambda,\alpha>0$	$\Gamma(\frac{1}{\alpha}+1)\lambda^{-\frac{1}{\alpha}}$	$\lambda^{-\frac{2}{\alpha}}(\Gamma(\frac{2}{\alpha}+1)-(\Gamma(\frac{1}{\alpha}+1))^2)$	
Rayleigh分布	$p(x)=\begin{cases}x\mathrm{e}^{-\frac{x^2}{2}}, & x\geq 0\\ 0, & x<0\end{cases}$	$\sqrt{\frac{\pi}{2}}$	$2-\frac{\pi}{2}$	
Laplace分布	$p(x)=\frac{1}{2\alpha}\mathrm{e}^{-\frac{\lvert x\rvert}{\alpha}}$ $-\infty<x<\infty,\alpha>0$	0	$2\alpha^2$	$\frac{i\alpha t}{1+\alpha^2t^2}$