概率论

1 事件与概率

1.1 随机现象与统计规律性

- 不确定性
- 随机试验
- 事件
- 样本空间Ω：所有可能的实验结果组成的集合
- 概率 P(A)的含义：1°事件的频率(客观)；2°事件的置信度(主观)

1.2 样本空间与事件

- 样本/点：一个试验结果，ω
- 样本空间/全集：所有试验结果，Ω
- 事件/子集：部分试验结果，A,B,……,Ω,Ø
- 事件 A 发生：本次试验结果ω∈A
- 概率：可能性，P(A)
- “交”，A∩B，AB：事件 A 发生且事件 B 发生
- “并”，A∪B：事件 A 发生或事件 B 发生
- 不交，AB=Ø：不相容，互斥；此时 A∪B 也记为 A+B
- “补”，$\overline{A}=A^c=\{\omega:\omega\notin A\}$：逆事件，对立事件
- “差”，$A\backslash B=A\overline{B}$；当 $B\subseteq A$ 时，记为 A-B
- 有限交：$A_1,\cdots,A_n$ 全部发生
- 有限并：$A_1,\cdots,A_n$ 中至少有一个发生
- 可列交：所有事件 $A_i,i\geqslant 1$ 都发生
- 可列并：存在某个事件 A_i 发生
- 若 $B_1\subseteq B_2\subseteq\cdots$，则 $\lim\limits_{n\to\infty}B_n \underline{\underline{def}} \bigcup\limits_{n=1}^{+\infty}B_n$
- 若 $B_1\supseteq B_2\supseteq\cdots$，则 $\lim\limits_{n\to\infty}B_n \underline{\underline{def}} \bigcap\limits_{n=1}^{+\infty}B_n$

1.3 古典概型

- 有限样本空间：Ω={1,2,3,⋯,n}；基本事件：{i}；概率：$P(A)=\Sigma_{i\in A}p_i$；其中 $p_i\geqslant 0$，$\forall i$；$\Sigma_i p_i=1$；含义：权分配，p_i，P({i})
- 古典概率模型：$p_i=1/n$，事件 A 的概率定义为 P(A)=|A|/|Ω|
- Jordan 公式：$\left|\bigcup\limits_i A_i\right|=\sum\limits_i|A_i|-\sum\limits_{i<j}|A_iA_j|+\sum\limits_{i<j<k}|A_iA_jA_k|+\cdots$
- 基本性质：1. 非负性；2. 规范性；3. 可加性

1.4 几何概型

- 几何概率模型：$\Omega\subset R^d$，$|\Omega|<\infty$，事件 A 的概率定义为 $P(A)=\dfrac{|A|}{|\Omega|}$。

• 基本性质：1. 非负性；2. 规范性；3. 可加性；4. 连续性

1.5 概率空间

• 基本假设：非负性，$P(A)\geqslant 0$；归一化，$P(\Omega)=1$；

• 直观要求：可列可加性，$P(\sum_n A_n)=\sum_n P(A_n)$；

• 若 $F\subseteq 2^{\Omega}$满足下列三条，则称 F 为Ω的一个σ-代数：1° $\Omega\in F$；2° 若 $A\in F$，则 $A^c\in F$；3° 若 $A_1,A_2,\ldots\in F$，则$\cup_i A_i\in F$。

• σ-代数的含义：观测能力、信息量

• A 生成的σ-代数：包含 A 的最小σ-代数，$\sigma(A)=\bigcap_{F\text{是}\sigma-\text{代数},F\supseteq A} F$

• 离散σ-代数，基本事件：$A_i,i\in I$，其中 $I=\{1,2,3,\cdots,n\}$或$\{1,2,3,\cdots\}$是Ω的划分；$A=\{A_i:\forall i\in I\}$，$F=\sigma(A)=\{\bigcup_{i\in J}A_i : J\subseteq I\}$

• Borel σ-代数 B(·)：开集生成的σ-代数

• 假设 F 是Ω的一个σ-代数，若 $P:F\to R$ 满足下列三条，则称 P 为(Ω,F)上的一个概率：1° 非负性；2° 规范性；3° 可列可加性。其中Ω是样本空间，(Ω,F)是可测空间，(Ω,F,P)是概率空间。

• 概率的性质：

1° $P(A\cup B)=P(A)+P(B)-P(AB)$；

2° $P(A\cup B)\leqslant P(A)+P(B)$；

3° $P(\cup_n A_n)\leqslant \Sigma_n P(A_n)$（可列次可加性）；

4° $P(AB)\geqslant P(A)+P(B)-1$。

2 条件概率与统计独立性

2.1 条件概率，全概率公式，贝叶斯公式

• 假设(Ω,F,P)是一个概率空间，$B\in F$ 且 $P(B)>0$，$\forall A\in F$，称$\dfrac{P(AB)}{P(B)}$为在 B 发生的条件下，A 发生的条件概率。记为 $P(A|B)$或 $P_B(A)$。

• $P_B(A|C)=P(A|BC)$；$P(A_1A_2\cdots A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1})$。

• 全概率公式：假设 $A_i,i\in I$ 是Ω的一个可数划分，则 $P(B)=\Sigma_i P(A_i)P(B|A_i)$。

• 划分可改为 $P(A_iA_j)=0$；$P(\cup_i A_i)=1$；$P(A_i)\geqslant 0$。

• 贝叶斯公式：设 $A_i,i\in I$ 是Ω的一个可数划分，则 $P(A_i|B)=\dfrac{P(A_i)P(B\mid A_i)}{\sum_j P(A_j)P(B\mid A_j)}$。

2.2 事件的独立性

• 若 $P(AB)=P(A)P(B)$，则称 A,B 相互独立。

• 若 $P(A_iA_j)=P(A_i)P(A_j),\forall i\neq j$，则称 $A_i,i\in I$ 两两独立。

• 若 $P(A_{i_1}\cdots A_{i_k})=P(A_{i_1})\cdots P(A_{i_k}),\forall k\leq n,\forall 1\leq i_1<\cdots<i_k\leq n$，则称 $A_i,i\in I$ 相互独立。相互独立的事件集可以对每个事件做补集运算，依旧相互独立。

2.3 伯努利试验与直线上的随机游动

• （小）试验：(Ω_i,F_i,P_i)：第 i 个小试验，$i=1,2,\cdots,n$（或 $i\geqslant 1$）

• 大试验的样本空间$\Omega=\Omega_1\times\Omega_2\times\cdots\cdots\times\Omega_n$，$\omega=(\omega_1,\omega_2,\cdots,\omega_n)$

• 事件：$\overline{A_i}\to A_i=\{\omega:\omega_i\in A_i\}=\overline{\Omega}_1\times\overline{\Omega}_2\times\cdots\times\overline{\Omega}_{i-1}\times\overline{A_i}\times\overline{\Omega}_{i+1}\times\cdots\times\overline{\Omega}_n$

• σ代数 $F=\sigma(\{A_1A_2\cdots A_n\})$

• 概率 P：与小实验相容，$P(A_i)=P_i(\overline{A_i})$

• 相互独立的小试验：$P(A_1\cdots A_n)=\prod_i P(A_i)$，$P=P_1\times P_2\times\cdots\times P_n$

• 独立重复试验：$(\Omega_i,F_i,P_i)=(\Omega_1,F_1,P_1)$，$\forall$ i

3 随机变量与分布函数

3.1 随机变量及其分布

• 设 F 是Ω上的σ代数，若 $X:\Omega\to R$ 满足$\{X\leqslant x\}\in F$，$\forall x\in R$，则称 X 为一个随机变量。

• X 生成的σ代数，$\sigma(X):=\sigma(\{X\leqslant x\}|x\in R)=\{\{X\in B\},\ \forall B\in Borel(R)\}$。

• X 是随机变量当且仅当$\sigma(X)\subset F$。

• 分布μ：(R,B)上的概率，随机变量 X 的分布μ_X,$L(X):B\to P(X\in B)$。

• 分布列，$\mu(\{x_k\})=p_k$，其中 x_i 互不相等，$p_k\geqslant0$，$\Sigma_k p_k=1$。

• 离散型随机变量 X：$P(X=x_i)=p_i$。

• 伯努利分布 X~B(1,p)，P(X=1)=p，B(X=0)=q=1−p。

• X~B(1,p)，A={X=1}，则 $P(X=1_A)=1$ 记为 $X=1_A$ a.s.。

• 二项分布 X~B(n,p)，$P(x=k)=C_n^kp^kq^{n-k}$。（独立重复试验）

• 超几何分布，X~H(N,M,n)，$P(x=k)=C_M^kC_{N-M}^{n-k}/C_N^n$。（N 个产品 M 个正品抽 n 次）

• 给定 n，当 N→∞时，M/N→p 时，h(k;N,M,n)→b(k;n,p)。

• 几何分布：X~G(p)，$P(x=k)=q^{k-1}p$。（第 1 次打中是第 k 发的概率）

• 几何分布的无记忆性：P(x−k=m|x＞m)=P(x=k)。

•帕斯卡分布：X~P(r,p)，$P(x=k)=C_{k-1}^{k-r}q^{k-r}p^r$。（打中第 r 次是第 k 次打靶的概率）

• 负二项分布：X~NB(r,p)，$P(x=k)=C_{r+k-1}^kq^kp^r$。（打中第 r 次是第 k+r 次打靶的概率）

• 泊松分布：X~P(λ)，P(x=k)=$\frac{\lambda^k}{k!}e^{-\lambda}$。

• 概率密度函数 $p(x)$，$\mu((-\infty,x])=\int_{-\infty}^x p(y)dy$，其中 $p(x)$≥0，$\int p(x)dx=1$。

• 连续型随机变量 $P(X\le x)=\int_{-\infty}^x p(y)dy$。

• 单独谈论一个点 x 对应的 $p(x)$是没有意义的。

• 均匀分布：X~U(a,b)，$p(x)=1/b-a$，$a<x<b$。

• 指数分布：X~Exp(λ)，$p(x)=\lambda e^{-\lambda x}$，$x\geqslant0$。是几何分布的极限，其中λ=np，1/n 是时间间隔。

• 指数分布无记忆性，$P(X-t>s|X>t)=e^{-\lambda s}$。

• 正态分布：X~N(μ,σ^2)，$p(x)$=$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$。

• Gamma 分布：X~Γ(r,λ)，$p(x)=\dfrac{\lambda^r}{\Gamma(r)}x^{r-1}e^{-\lambda x}$。

• X 的分布函数：$F(x)=P(X\leqslant x)$。
• $F=F_X:X\to P(X\leqslant x)$满足：1°单调性；2°归一性；3°右连续性。

• 等价函数：$\hat{F}(x)=P(X<x)$。

• 尾分布函数：$G_X(x)=1-F(x)=P(X>x)$，G_X 表示由随机变量 X 诱导的函数。
• 连续型：F 是 R 上的连续函数，在一定条件下 $p_X(x)=-G'_X(x)$。
• 同分布：分布函数/分布列/密度相同。

3.2 随机向量,随机变量的独立性

• 随机向量：同一个(Ω,F)中的多个随机变量。
• $\{\boldsymbol{X}\leqslant x\}$：$\{X_1\leqslant x_1,\cdots,X_n\leqslant x_n\}=\{\boldsymbol{X}\in \mathrm{D}\}$，$\mathrm{D}=(-\infty,x_1)\times\cdots\times(-\infty,x_n)$。
• 联合分布：$\mathrm{B}\to\mu_\xi(\mathrm{B})$，B 是 $\mathrm{Borel}(\mathrm{R}^2)$。
• 联合分布函数 $F(x,y)=P(X\leqslant x,Y\leqslant y)$，满足：1°单调性；2°归一性；3°右连续性；4°对任意 $a_1<b_1$，$a_2<b_2$，有 $F(b_1,b_2)+F(a_1,a_2)-F(a_1,b_2)-F(a_2,b_1)\geqslant 0$。
• 边际分布列：$P(X=x_i),i\in I$。
• 条件分布列：固定 i，$P(Y=y_j|X=x_i), j\in J$。
• 联合分布列：$P(X=x_i,Y=y_j),i, j\in I,J$。

• 连续型：(X,Y)有联合概率密度函数 $p(x,y)$，$P(X\le x,Y\le y)=\int_{-\infty}^{x}\int_{-\infty}^{y}p(u,v)dudv$。

• 边缘密度：$p_X(x)=\int p(x,y)dy$。

• 条件密度：固定 x，$P_{Y|X}(y|x)=p(x,y)/p_X(x)$。

• 多元正态分布：$\boldsymbol{X}\sim N(\boldsymbol{\mu},\Sigma)$，$p(\boldsymbol{x})=\dfrac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\exp\{-\dfrac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\}$。

• 多元正态分布的边缘分布与条件分布都是正态分布。

• 二元正态分布，$\boldsymbol{\mu}=(\mu_1,\mu_2)$，$\Sigma=\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}$，则

$p(x,y)=\dfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\{-\dfrac{u^2-2\rho uv+v^2}{2(1-\rho^2)}\}$，其中$u=\dfrac{x-\mu_1}{\sigma_1},v=\dfrac{x-\mu_2}{\sigma_2}$。

• 随机变量的相互独立性：$P(X_1\le x_1,\cdots,X_n\le x_n)=P(X_1\le x_1)\cdots P(X_n\le x_n)$，则称 $X_1,\cdots,X_n$ 相互独立。
• 离散型：$X_1,\cdots,X_n$ 独立当且仅当 $P(X_1=x_1,\cdots,X_n=x_n)=P(X_1=x_1)\cdots P(X_n=x_n)$。
• 连续型：$X_1,\cdots,X_n$ 独立当且仅当 $p(\boldsymbol{X})=p(x_1)\cdots p(x_n)$。
• 独立充分条件：$p(x,y)=f(x)g(y)$。
• 独立充分条件：$p_{Y|X}(y|x)=g(y)$。
• 随机变量独立的性质：假设 $X_1,\cdots,X_n$ 相互独立，则：
1°假设 g_i 是一元可测函数，则 $g_i(X_i)$相互独立；
2°假设 $f(X_1,\cdots,X_k)$是 k 元可测函数，则 $f(X_1,\cdots,X_k),X_{k+1},\cdots,X_n$ 相互独立。

3.3 随机变量的函数及其分布

• $Y = f(X)$，f 是 Borel 函数。
• 目标：求 Y 的分布。
• 离散型：$P(Y = y_j) = \sum_{i:f(x_i)=y_j} p_i$。

• 分布函数的广义逆：$F^{-1}(u) := \inf\{x : F(x) \geq u\}$。

• $F^{-1}(u)$是 Borel 函数。
• $x_0=F^{-1}(u)\leqslant x$ iff $u\leqslant F(x)$。
• 取 $U\sim U(0,1)$，令 $X=F^{-1}(u)$，则 $F_X = F$。这表明任意分布函数都是某随机变量的分布函数。
• 连续型，f 严格单调，$y=f(x)$，$p_Y(y)|dy|=p_X(x)|dx|$。
• 若 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 独立，则 $f(\boldsymbol{X})$与 $g(\boldsymbol{Y})$独立。
• 连续型，$f:\mathrm{R}^n\to\mathrm{R}^n$，$x\to y$。
• 一对一：$p_X(x)|dx|=p_Y(y)|dy|$，$p_Y(y)=p_X(x)\cdot|dx/dy|$。
• 方法：分布函数法、补变量法。
• $\mu*\nu = L(x+y)$，其中 $X\sim\mu$，$Y\sim\nu$，且 X 与 Y 独立。
• 一族分布Π满足可加性/再生性是指$\mu*\nu\in\Pi$，$\forall\mu,\nu\in\Pi$。
• 若 $X_1,X_2,\cdots$ i.i.d.，$S_n=\Sigma_{i=1,2,\ldots}X_i$，则 $L(S_n)*L(S_m)=L(S_{n+m})$。
• $\chi^2(\mathrm{n})=\Gamma(n/2,1/2)$，具有自由度 n 的χ^2 分布。

4 数字特征与特征函数

4.1 数学期望

• 时间平均：大量观测值的算术平均。
• 空间平均：所有可能值的加权平均。
• 离散型：若$\Sigma_k p_k|x_k|<\infty$，则称$\Sigma_k p_k x_k$ 为 X 的数学期望，记为 EX。反之，则称 X 的期望不存在。
• 分布的数字特征：若 X=Y i.i.d，则 EX=EY。

• X 取非负整数，EX=$\sum_{n=1}^{+\infty} P(X \geq n) = \sum_{n=0}^{+\infty} P(X > n)$。

• 连续型：若$\int|x|p(x)dx<\infty$，则称$\int xp(x)dx$ 为 X 的数学期望，记为 EX。

• X 取非负，EX=$\int_0^\infty G(x)dx$。

• Lebesgue-Stieltjes 积分：$\int xdF(x) = \lim_{\Delta\to 0}\sum_i x_i(F(x_{i+1}) - F(x_i))$。

• 若$\int|x|dF(x)<\infty$，则称$\int xdF(x)$为 X 的期望，记为 EX。
• 若 X 有界：P(|X|≤M)=1，则期望存在。
• 数学期望的性质：1° x=c，则 EX=c；2° 单调性：若 X≥Y，则 EX≥EY；3° 线性：E(aX)=aE(X)，E(X+Y)=EX+EY。
• 若 X≥0 且 EX=0，则 X=0。
• 若 X≥0 且 EX<∞，则 $\lim_{x\to\infty} xG(x) = \lim_{x\to\infty} EX \cdot 1_{X>x} = 0$。

• 函数的期望：Y=f(X)，则 EY=$\int f(x)dF(x)$。

• 相互独立则：E(XY)=(EX)(EY)。
• Jensen 不等式：f(x)是凸函数，则 Ef(X)⩾f(EX)。
• Y 关于事件 A 的条件期望：$E(Y|A)=\int x dF_A(x)$，其中 $F_A(x)=P(Y\leqslant x|A)$。
• Y 关于 X 的条件变量：E(Y|X)=f(X)。
• X 是离散型：f(x)=E(Y|X=x)，$\Sigma_j y_j P(Y=y_j|X=x)$。
• X 是连续型：$f(x)=E(Y|X=x):=\lim_{\varepsilon\to 0}E(Y|x-\varepsilon<X<x+\varepsilon)$，$\int y p_{Y|X}(y|x)dy$。
• 重期望公式：EY=EE(Y|X)=Ef(X)。
• 离散型：$Ef(X)=\Sigma_i f(x_i)P(X=x_i)=\Sigma_{i,j}y_jP(X=x_i,Y=y_j)=EY$。

4.2 方差、相关系数、矩

• 方差：假设 E|X|<∞，若 $E(X-EX)^2=E(X^2)-(EX)^2$ 有限，则称它为 X 的方差，记为 var(X)或 DX。

• 标准差：$\sigma_X=\sqrt{\mathrm{var}(X)}$ 称为标准差。

• 矩：EX^k，$E(X-EX)^k$，Ee^{aX}。
• 线性变换：$\mathrm{var}(a+bX)=b^2\mathrm{var}(X)$。
• 标准化：$X^*=(X-\mu)/\sigma$，$EX^*=0$，$DX^*=1$。
• Chebyshev's 不等式：$P(|X-EX|\geqslant\varepsilon)\leqslant DX/\varepsilon^2$。
• 假设 $EX^2,EY^2<\infty$，称 E(X-EX)(Y-EY)为 X,Y 的协方差，记为 cov(X,Y)。
•和的方差：var(X+Y)=var(X)+var(Y)+2cov(X,Y)，独立 var(X+Y)=var(X)+var(Y)。
• 计算公式：cov(X,Y)=E(XY)-(EX)(EY)。
• 对称双线性函数：X'=aX+c，Y'=bX+d，则 cov(X',Y')=ab • cov(X,Y)。
• Cauchy 不等式：$(EXY)^2\leqslant EX^2\,EY^2$，取等号条件 Y=aX。
• $L^2=\{X|EX^2<\infty\}$，定义内积(X,Y)=EXY，夹角<X,Y>=‖X‖·‖Y‖·cosθ。
• $L_0^2=R^{\perp}=\{X\in L^2|EX=0\}$，有 $X=(X-EX)\oplus EX$。

• 相关系数：$\rho=\dfrac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)\,\mathrm{var}(Y)}}$。

• 若 X'=aX+c，Y'=bY+d，则$\rho_{X',Y'}=\rho_{X,Y}(ab>0)$或$-\rho_{X,Y}(ab<0)$。
• ρ=1 iff Y=aX+b，a>0；ρ=−1 iff Y=aX+b，a<0
• 不/正/负相关：ρ=0/⩾0/⩽0。
• 完全正/负相关：ρ=1/-1。
• 独立则线性不相关，但反之不然。
• 假设 A,B 是事件：A,B 不/正/负相关：P(AB)=/⩾/⩽P(A)P(B)。
• 二维正态分布时：不相关(ρ=0)当且仅当独立。
• $\boldsymbol{X}=(X_1,\cdots,X_n)^T$，期望：$E\boldsymbol{X}=(EX_1,\cdots,EX_n)^T$，协方差矩阵$\Sigma=(\sigma_{ij})_{n\times n}$，$\sigma_{ij}=\mathrm{cov}(X_i,X_j)$，∑是半正定对称矩阵，正定 iff $1,X_1,\cdots,X_n$ 线性无关。
• 最佳预测：目标：找 Y'∈**Y**，使得 $E(Y-Y')^2=\min_{V\in \mathbf{Y}}E(Y-V)^2$。结论：Y'是垂足。
• 若 **Y**=R，则 Y'=EY。
• 若 **Y**={a+bX|a,b∈R}，其中 EX=0，$EX^2=1$，则 b=cov(X,Y)。
• 若 $\mathbf{Y}=\{f(X):Ef(X)^2<\infty\}$，则 $f(X)=E(Y|X)$。
• 随机变量 X,Y 线性相关的强弱由$\rho_{X,Y}$刻画，$|\rho_{X,Y}|$越接近 1，越线性相关。

4.4 母函数

• 设 X 取非负整数，$P(X=k)=p_k$，称 $\sum_{k=0}^{+\infty}p_k s^k, s\in[-1,1]$ 为 X 的母函数，记为 $g_X(s)$。

• 分布的性质：X=Y 当且仅当 $g_X=g_Y$。
• $g(s)=\mathrm{E}s^X$。
• 矩：对 $s\in(-1,1)$，$g'(s)=p_1+2p_2s+\cdots+kp_ks^{k-2}+\cdots=\mathrm{E}Xs^{X-1}$。$g''(s)=\mathrm{E}X(X-1)s^{X-2}$。$g^{(l)}(s)=\mathrm{E}X(X-1)\cdots(X-l+1)s^{X-l}$。知 $\mathrm{E}X=g'(1),\mathrm{E}X^2=g''(1)+g'(1)$。
• 乘积：若 X 与 Y 独立，则 $g_{X+Y}(s)=g_X(s)g_Y(s)$。
• $X\sim G(p)$，则 $g(s)=\dfrac{ps}{1-qs}$；$X\sim B(n,p)$，则 $g(s)=(q+ps)^n$；$X\sim P(\lambda)$，则 $g(s)=e^{\lambda(s-1)}$。
• 复合：$\xi_1,\xi_2,\cdots$独立同分布，且它们与 W 独立。令 $X=\xi_1+\cdots+\xi_W$，则 $g_X=g_W(g_\xi)$。
• 凸组合：X,Y,ξ相互独立，$\xi\sim B(1,p)$。令 $W=X\cdot 1_{\{\xi=1\}}+Y\cdot 1_{\{\xi=0\}}$，则 $g_W=pg_X+qg_Y$。

4.5 特征函数

• 称 $\mathrm{E}e^{itX}=\mathrm{E}\cos(tX)+i\mathrm{E}\sin(tX)$为 X 的特征函数，记为$f_X(t)$。
• 基本性质：$f(0)=1$；$\|f(t)\|\leqslant 1$；f 一致连续；f 半正定（$t_1,t_2,\cdots,t_n\in\mathbb{R}$，$a_{ij}=f(t_i-t_j)$，则$(a_{ij})$半正定）。
• B-K 定理：$f:\mathbb{R}\to\mathbb{C}$ 满足 $f(0)=1$，连续，半正定，则存在 X 使得 $f=f_X$。
• 逆转公式&唯一性定理：$F(x)-F(y)=\dfrac{1}{2\pi}\lim\limits_{T\to\infty}\int_{-T}^{T}\dfrac{e^{-ity}-e^{-itx}}{it}f(t)dt$，$x,y\in C(F)$（连续点）。若$\int|f(x)|dx<\infty$，则分布函数连续可导，且 $p(x)=\dfrac{1}{2\pi}\int e^{-itx}f(t)dt$。
• 性质：乘积：X 与 Y 独立，则 $f_{X+Y}(t)=f_X(t)f_Y(t)$。
• $X\sim B(n,p)$，$f_n(t)=(q+pe^{it})^n$；$X\sim P(\lambda)$，$f_\lambda(t)=e^{\lambda(e^{it}-1)}$。
• 凸组合：X,Y,ξ相互独立，$\xi\sim B(1,p)$。令 $W=X\cdot 1_{\{\xi=1\}}+Y\cdot 1_{\{\xi=0\}}$，则 $f_W=pf_X+qf_Y$。
• 若 $\mathrm{E}X^k$ 存在，则 $f^{(k)}(0)=i^k\mathrm{E}X^k$，且成立 Peano 余项的 Taylor 公式。
• X 与 Y 独立 iff $f_{X,Y}(t,s)=f_X(t)f_Y(s)$。
• X 与 Y 独立$\Rightarrow f_{X+Y}(t)=f_X(t)f_Y(t)$，反之不然（cauchy distribution）。

4.6 多元正态分布

• 密度函数：$p_X(x)=\dfrac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\exp\{-\dfrac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\}$
• 非退化线性变换：$Y=v+BX\sim N(v+B\mu,B\Sigma B^T)$。取 $B=\sqrt{\Sigma}^{-1}$，$v=\mu$，则 $Y\sim N(0,I)$。
• 特征函数：$Z\sim N(0,I)$，$f_Z(t)=Ee^{itZ}=\prod\limits_{i=1}^{n}Ee^{it_iZ_i}=e^{-\frac{1}{2}\|t\|^2}$。
$X\sim N(\mu,\Sigma)$，$f_X(t)=Ee^{itX}=Ee^{it(\mu+\sqrt{\Sigma}Z)}=\exp\{it\cdot\mu-\dfrac{1}{2}t^T\Sigma t\}$。
• 高斯分布：Σ半正定，若 X 的联合特征函数为$f(t)=\exp\{it\cdot\mu-\dfrac{1}{2}t^T\Sigma t\}$，则称 X 服从高斯分布 $N(\mu,\Sigma)$，也称 X 为一个高斯向量。
• 数字特征：期望：$\mathrm{E}X=\mu$，协方差矩阵：Σ。
• $X\sim N(\mu,\Sigma)$ iff $\forall a_1,\cdots,a_n$，$Y:=\Sigma_i a_iX_i\sim N(\mu,\sigma^2)$。
• 高斯分布：不相关$\Leftrightarrow$独立。

• 一般情形：rank(Σ)=r，则存在 $1\leqslant i_1<\cdots<i_r\leqslant n$ 使得 Y~正态分布。
• 存在 $Z\sim N(0,I_{r\times r})$以及列满秩矩阵 $A_{n\times r}$ 使得 X=μ+AZ。
• 存在 $Z\sim N(0,I_{n\times n})$以及秩为 r 的方阵 $A_{n\times n}$ 使得 X=μ+AZ。
• $X=(Y_1,\cdots,Y_r,W_{r+1},\cdots,W_n)\sim N(\mu,\Sigma)$，假设 Y 服从正态分布，求 $L(W|Y=y)$。
• 正交分解：找 $B_{(n-r)\times r}$ 使得 V=(W-BY)且 V⊥Y。
• 协方差：$\mathrm{cov}(V_k,Y_i)=(\Sigma_{21}-B\Sigma_{11})_{ki}$。取 $B=\Sigma_{21}\Sigma_{11}^{-1}$，则 $L(By+V)=N(w+B(y-v),\Sigma_{22})$。

5 极限定理

5.1 伯努利试验场合的极限定理

• 假设 $X=X_1,X_2,\cdots$独立同分布，P(X=1)=p，P(X=0)=q。
• 伯努利大数定律：$P(|S_n/n-p|\geqslant\varepsilon)\to 0$。
• 如果任意ε＞0，都有 $\lim P(|\xi-\xi_n|\geqslant\varepsilon)\to 0$，则称$\xi_n$依概率收敛到ξ，$\xi_n \xrightarrow{P} \xi$。
• $S_n^*=(S_n-np)/\sqrt{npq}$，则 $P(a<S_n^*\leqslant b)\to \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$。

5.2 依概率收敛&几乎必然收敛

• 如果 $\lim E|\xi_n-\xi|^r=0$，称ξ_n r 阶收敛到ξ，$\xi_n \xrightarrow{r} \xi$。
• r 阶收敛⇒依概率收敛，反之不然。
• ξ_n依概率收敛到ξ，$E|\xi_n|^r\to E|\xi|^r$，则ξ_n r 阶收敛到ξ。
• 弱大数定律：若 $X_1,X_2,\cdots$满足***，则 S_n/b_n-a_n 依概率收敛到 0。
• 切比雪夫弱大数定律：***：两两不相关，$\mathrm{var}(X_i)\leqslant M$，则$(S_n-ES_n)/n$ 依概率收敛到 0。
• 马尔可夫弱大数定律：***：$\mathrm{var}(S_n)=o(n^2)$，则$(S_n-ES_n)/n$ 依概率收敛到 0。
• 有界收敛定理：ξ_n依概率收敛到ξ，且 $P(|\xi_n|\leqslant M)=1$，则 $\lim E\xi_n=E\xi$。
• 定理：设 $X=X_1,X_2,\cdots\cdots$独立同分布，E|X|＜∞，则 S_n/n 依概率收敛于 EX。

5.4 几乎必然收敛&强大数定律

• 如果 $P(\lim\xi_n=\xi)=1$，那么称ξ_n几乎必然收敛到ξ，记为$\xi_n\to\xi$ a.s.。
• 令 $A_{n,\varepsilon}=\{|\xi_n-\xi|>\varepsilon\}$，则 $\{\lim_{n\to\infty}\xi_n=\xi\}^c=\bigcup_{k=1}\bigcap_{N\geq 1}\bigcup_{n\geq N} A_{n,\frac{1}{k}}$。
• 对任意事件 $A_1,A_2,\cdots$，令 $\{A_n\ i.o.\}=\overline{\lim_{n\to\infty}} A_n := \bigcap_{N\geq 1}\bigcup_{n\geq N} A_n$。
• Borel-Cantelli 引理：$s=\sum_{n=1}^{+\infty} P(A_n)$。若 s＜∞，则 $P(A_n$ i.o.)=0；若 s=∞，且 $A_1,\cdots$相互独立，则 $P(A_n$ i.o.)=1。
• 几乎必然收敛⇒依概率收敛，反之不然；几乎必然收敛和 r 阶矩收敛没有互推关系。
• 若ξ_n依概率收敛到ξ，则存在子列$\{n_k\}$使得ξ_{n_k}几乎必然到ξ。
• ξ_n依概率收敛到ξ iff 任意子列都有子列几乎必然收敛到ξ。
• 强大数定律：若 X_1,X_2 满足***，则 S_n/b_n-a_n 几乎必然收敛到 0。
• Borel-Cantelli 强大数律：$X_1,X_2\cdots$相互独立，$E(X_i-EX_i)^4\leqslant M$，则$(Sn-ES_n)/n$ 几乎必然收敛到 0。

• 定理：$X=X_1,X_2,\cdots$独立同分布，$EX^2<\infty$，则S_n/n几乎必然收敛到EX。
• Kolmogorov 强大数律：$X=X_1,X_2,\cdots$独立同分布，$E|X|<\infty$，则S_n/n几乎必然收敛到EX。

• Kolmogorov 强大数律：$X_1,\cdots$独立，$\sum_{n=1}^{+\infty}\frac{DX_n}{n^2}<\infty$，则$P(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(X_i-EX_i)=0)$
=1。
• 若$X_1,X_2,\cdots$独立，且S_n/n几乎必然收敛到Y，则Y退化。
• 若i.i.d.，且S_n/n几乎必然收敛到a，则a=EX。
• 若i.i.d，则X_n/n几乎必然收敛到0 iff $E|X|<\infty$。
• 定理：$X=X_1,X_2,\cdots$独立同分布，$EX=\infty$，S_n/n几乎必然收敛到∞。

5.3 依分布收敛&中心极限定理

• 如果$\forall x\in C(F_\xi)$，都有$\lim F_{\xi n}(x)=F_\xi(x)$，则称$\xi_n$依分布收敛于$\xi$（没有考察$\xi_n$的取值，只考察了$\xi_n$的分布；原像ω测度一样，但原像ω取值可以不一样）。
• 依概率收敛⇒依分布收敛，反之不然。
• 依分布收敛于常数⇒依概率收敛于常数。
• 有界收敛定理：ξ_n有界，依分布收敛于ξ，则$E\xi_n\to E\xi$。
• ξ_n依分布收敛到ξ iff $Ef(\xi_n)=Ef(\xi)$，$\forall f\in F$。（F 可取：$1_{(-\infty,b]}$，其中$b\in C(F_\xi)$(的稠密子集)；$1_{(a,b]}$，其中$a,b\in C(F_\xi)$；阶梯函数；有界连续函数；三角函数）
• 定理：ξ_n依分布收敛到ξ iff 特征函数$f_{\xi n}(t)\to f_\xi(t)$收敛。
• 定理：若$f_{\xi n}(t)\to f(t)$且f在t=0连续，则f是特征函数，ξ_n依分布收敛。
• 中心极限定理：$X=X_1,X_2,\cdots$i.i.d，$0<DX<\infty$，则S_n^*依分布收敛到N(0,1)。

• $\frac{1}{\sqrt{n}}(S_n-ES_n)\xrightarrow{d}\sigma Z\sim N(0,\sigma^2)$。

• 研究对象：$S_n^*=\sum_{k=1}^{n}\frac{X_k-\mu}{\sqrt{\mathrm{var}(S_n)}}$。假设$X_1,X_2,\cdots$独立同分布，$EX_k=\mu_k$，$DX_k=\sigma_k^2$，$B_n^2=\sum_{k=1,\dots,n}\sigma_k^2$。

• Feller 条件：$\frac{\max_{1\le k\le n}\sigma_k}{B_n}\to 0\ iff\ B_n\to\infty\ and\ \frac{\sigma_n}{B_n}\to 0$。

• Lindeberg 条件：$\frac{1}{B_n^2}\sum_{k=1}^{n}E|X_k-\mu_k|^2 1_{\{|X_k-\mu_k|>\varepsilon B_n\}}\to 0$。

•Lindeberg-Feller CLT：Lindeberg 成立 iff S_n^*依分布收敛到N(0,1)且 Feller 成立。