概率论 课堂笔记

原生生物

斗满人概，人满天概。——管子

目录

一 概率空间与独立性

§1.1 事件与概率

样本点-具体结果 **样本空间** Ω-样本点的全体 **事件** A-样本空间的子集

例 1.1 掷硬币 $\Omega=\{H,T\},A=\{H\}$
电子自旋 $\Omega=\{\uparrow,\downarrow\},A=\{\uparrow\}$
掷骰子 $\Omega=\{1,2,3,4,5,6\},A=\{2,4,6\}$

* 上方的例子中，样本点均只有有限多个。

例 1.2 道琼斯指数 样本点-连续曲线 $x(t),t\in[0,T]$ 样本空间 $\Omega=\{x(t)\in C[0,T]\}$

事件运算 $\longleftrightarrow$ 集合运算
事件 A 发生 $\longleftrightarrow$ 试验结果 $\omega\in A$
$A=\varnothing$ **不可能事件** $A=\Omega$ **必然事件**
事件交、并、余 $\longleftrightarrow$ $A\cap B,A\cup B,A^{\mathrm{c}}$
事件的交-两事件均发生 事件的并-两事件至少发生一个 事件的余-事件未发生
* 可记 $A\cap B$ 为 AB
A 与 A^{c} 称**对立事件**。
A 发生则 B 亦发生 $\longleftrightarrow$ $A\subset B$
$A\cap B=\varnothing$ 时称 A,B **不相容**。
* 由此亦可定义 $A_1,\dots,A_n,\dots$ 互不相容

问题：是否要将 Ω 的所有子集定义为随机事件？
* 良好定义的随机事件为 Ω 的一个子集族，且至少要求对交、并、余三种运算**封闭**。

例 1.3 掷硬币第一次正面向上的时刻 $\Omega=\{1,2,\dots\}$ 此时样本空间为无限集合，有限交并性质不足！

定义 1.1 $\mathcal{F}$ 为 Ω 某些子集构成的子集族，称其为事件域 (σ 域)，若：
1. $\Omega\in\mathcal{F}$
2. $A\in\mathcal{F}\implies A^{\mathrm{c}}\in\mathcal{F}$
3. $A_n\in\mathcal{F},n\in\mathbb{N}^*\implies\bigcup\limits_{n=1}^{\infty}A_n\in\mathcal{F}$

并称 $(\Omega,\mathcal{F})$ 为一个可测空间。

例 1.4 最大 σ 域，Ω 的幂集
最小 $\mathcal{F}=\{\Omega,\varnothing\}$
中间域，如 $A\neq\varnothing,\Omega$ 时 $\mathcal{F}=\{\Omega,\varnothing,A,A^{\mathrm{c}}\}$

定义概率：直观想法-**频率稳定性** (重复试验后计数发生次数)
重复 N 次，A 发生 N_A 次，经验表明 $\lim\limits_{N\to\infty}\dfrac{N_A}{N}=c$，记 c 为 $P(A)$。明显性质：$P(\varnothing)=0,P(\Omega)=1$。
若 $A\cap B=\varnothing$，$N_{A\cup B}=N_A+N_B\implies P(A\cup B)=P(A)+P(B)$

定义 1.2 $P:\mathcal{F}\to\mathbb{R}$ 为 $(\Omega,\mathcal{F})$ 上的概率测度，若：
1. 非负性 $P(A)\geq 0$
2. 规范性 $P(\Omega)=1$
3. 可列可加性 若 $\{A_n\}$ 互不相容，则 $P\left(\bigcup\limits_n A_n\right)=\sum\limits_n P(A_n)$

并称 $(\Omega,\mathcal{F},P)$ 为一个概率空间。

例 1.5 掷硬币 $\Omega=\{H,T\},\mathcal{F}=2^\Omega,P(\{H\})=p,P(\{T\})=q,p+q=1$

例 1.6 $\Omega=\{1,2,3,4,5,6\},\mathcal{F}=2^\Omega$ 假定公平，则 $P(A)=\dfrac{|A|}{6}$

样本点有限，且每个样本点等概率发生，则称**古典概型**。

定理 1.1 概率测度基本性质

1. $P(A^c)=1-P(A)$
2. $A\subset B$ 时 $P(B)=P(B\backslash A)+P(A)\geq P(A)$
3. $P(A\cup B)=P(B)+P(A)-P(AB)$
4. *Jordan* 公式 $P\left(\bigcup\limits_{i=1}^n A_i\right)=\sum\limits_{k=1}^n\sum\limits_{i_1<\cdots<i_k}(-1)^{k-1}P(A_{i1}\ldots A_{ik})$

定理 1.2 概率测度连续性

1. 单调增事件列 $A_1\subset\cdots\subset A_n\subset\cdots$，记 $A=\bigcup\limits_{i=1}^\infty A_i=\lim\limits_{n\to\infty}A_n$，则 $P(A)=\lim\limits_{n\to\infty}P(A_n)$。
2. 单调减事件列 $B_1\supset\cdots\supset B_n\supset\cdots$，记 $B=\bigcap\limits_{i=1}^\infty B_i=\lim\limits_{n\to\infty}B_n$，则 $P(B)=\lim\limits_{n\to\infty}P(B_n)$。

证明：$A=A_1\cup(A_2\backslash A_1)\cup\cdots\cup(A_n\backslash A_{n-1})\cup\cdots$。此处的并均为无交并，因此可由定义与引理计算即得结果。对于 B，取余即可化为 A 的情况。

§1.2 条件概率与独立性

直观想法-重复试验，B 发生 N_B 次，B 发生条件下 A 发生次数 N_{AB}，次数足够多时**条件概率**可看作 $\dfrac{N_{AB}}{N_B}$。

定义 1.3 条件概率

设 $P(B)>0$，B 发生条件下 A 的条件概率 $P(A|B)=\dfrac{P(AB)}{P(B)}$

变形有乘法公式 $P(AB)=P(B)P(A|B)$

*$B_1,\ldots,B_n$ 为 Ω 的互不相容子集，且 $\bigcup\limits_i B_i=\Omega$，则称其为 Ω 的一个**划分**。

定理 1.3 全概率公式

$B_1,\ldots,B_n$ 为 Ω 的一个划分，$P(B_i)>0$，则 $A=A\cap\Omega=\bigcup\limits_i(B_i\cap A)\Rightarrow P(A)=\bigcup\limits_i P(B_i)P(A|B_i)$。

例 1.7 坛子里有 3 白 2 红共 5 个球，每次无放回摸出一个球，$A=${第二次摸到白球}，按第一次抽到白或红划分可知 $P(A)=\dfrac{3}{5}\cdot\dfrac{1}{2}+\dfrac{2}{5}\cdot\dfrac{3}{4}=\dfrac{3}{5}$(注意到，在每个轮次抽出白球的概率一致)。

定理 1.4 贝叶斯公式

$A_1,\ldots,A_n$ 为 Ω 的一个划分，$P(A_i)>0$，则 $P(B)>0$ 时 $P(A_i|B)=\dfrac{P(A_i)P(B|A_i)}{\sum_j P(A_j)P(B|A_j)}$

例 1.8 发出 s 时，收到 s 概率为 0.8，收到 t 概率为 0.2；发出 t 时，收到 s 概率为 0.1，收到 t 概率为 0.9。且发出 s 概率为 0.6，发出 t 概率为 0.4。

收到 s 的情况下，发出 s 的概率为：$\dfrac{0.6\cdot0.8}{0.6\cdot0.8+0.4\cdot0.1}=0.923$

掷硬币，B 代表第一次正面，A 代表第二次正面，则 $P(A|B)=P(A)$，即 $P(AB)=P(A)P(B)$，由此引出定义：

定义 1.4 独立性
称 A,B 独立，若 $P(AB)=P(A)P(B)$
更一般，称 $A_1,\ldots,A_n$ 相互独立是指 $\forall 2\leq k\leq n, i_1<\cdots<i_k, P(\prod_{j=1}^{k} A_{i_k})=\prod_{j=1}^{k} P(A_{i_k})$；
两两独立是指只需 $k=2$ 时满足。

两两独立与相互独立不同，举例如下：

例 1.9 古典概型中，$\Omega=\{1,2,3,4\}, A=\{1,3\}, B=\{1,2\}, C=\{1,4\}$。
计算可知 A,B,C 两两独立，但不相互独立。

定理 1.5 若 A,B 独立，则 A 与 B^{c}，A^{c} 与 B，A^{c} 与 B^{c} 独立。更一般地，若一些事件相互独立，将其中部分改为其对立事件后仍然相互独立。

证明：两事件时，由对称，只需证明 A 与 B^{c} 独立。由 $P(AB^{\mathrm{c}})+P(AB)=P(A)$ 可算出结果。多事件时类似两事件一个个调整即可。

例 1.10“重复独立试验，小概率事件必发生”
记事件为 A，$A_k=$\{第 k 次试验中 A 发生\}，则 $\bigcup_{k=1}^{n} A_k$ 表示 \{前 k 次试验中 A 发生\}。
$P(\bigcup_{k=1}^{n} A_k)=1-P(\bigcap_{k=1}^{n} A_k^{\mathrm{c}})$，由独立性，此式为 $1-\prod_{k=1}^{n}(1-P(A_k))$，由此知结果。

§1.3　概率模型

例 1.11 对称随机游走
赌徒财富 k，庄家 $N-k$，掷均匀硬币，正面则赌徒赢庄家 *1*，否则庄家赢赌徒 *1*，赌至一方输光，问赌徒输光概率。

记赌徒初始为 k 且输光的事件为 A_k，B 为首局正面，由此 $P(A_k)=P(B)P(A_k|B)+P(B^{\mathrm{c}})P(A_k|B^{\mathrm{c}})=\dfrac{P(A_{k-1})+P(A_{k+1})}{2}$，且 $P(A_0)=1, P(A_N)=0$，由等差数列知 $P(A_k)=\dfrac{N-k}{N}$。

计数问题

例 1.12 坛子里有 *4* 白 *6* 红共 *10* 个球，随机取 *4* 个，求 *2* 白 *2* 红概率。

样本点数目 C_{10}^4，事件发生的样本点数目 $\mathrm{C}_4^2\mathrm{C}_6^2$，结果为 $\dfrac{3}{7}$。
* **古典概型**重要应用：**排列组合**计算样本点数目
经典问题：n 个对象中选 m 个，问选法种数 (是否可重复？是否考虑顺序？)。
有序不重复：A_n^m 无序不重复：C_n^m 有序可重复：n^m
无序可重复：插板法，看作 m 个小球 n 个盒子 ($n-1$ 个挡板)，知结果为 C_{n-1+m}^m

例 1.13 将 n 个小球投入到 $N\geq n$ 个盒子中，投法等可能，求前 n 个盒子中各一个球的概率 *(*球是否可分辨？盒子是否有容量限制？*)*。

(1) 球可区分，盒子无限制 (麦克斯韦-玻尔兹曼统计)：样本点个数 N^n，合要求个数 $n!$，概率 $\dfrac{n!}{N^n}$
(2) 球不可区分，盒子无限制 (玻色-爱因斯坦统计)：化为无序可重复，样本点个数 C_{n+N-1}^n，合要求个数 1，概率 $\dfrac{1}{\mathrm{C}_{n+N-1}^n}$
(3) 球不可区分，盒子容量一 (费米-狄拉克统计)：样本点个数 C_N^n，合要求个数 1，概率为 $\dfrac{1}{\mathrm{C}_N^n}$

例 1.14 *Polya* 模型
坛子里有一些球，b 黑 r 红，先摸出一个记下颜色后放回，并且再放入 c 个同色球。记 B_n 表示第 n 次抽到黑球，求概率。

观察可发现，n 次抽取抽中 k 次黑球，任意给定次序概率相同，为 $D_k(b)=\dfrac{\prod_{i=0}^{k-1}(b+ic)\prod_{i=0}^{n-k-1}(r+ic)}{\prod_{i=0}^{n-1}b+r+ic}$，有 $B_{n+1}=\sum\limits_{k=0}^{n}\mathrm{C}_n^kD_k(b)\dfrac{b+kc}{b+r+nc}=\dfrac{b}{b+r}\sum\limits_{k=0}^{n}\mathrm{C}_n^kD_k(b+r)$，而由概率含义 $\sum\limits_{k=0}^{n}\mathrm{C}_n^kD_k(b+r)$ 构成整个样本空间，因此为 1，因此 $B_{n+1}=\dfrac{b}{b+r}$。
*$c=-1$ 即为无放回，$c=0$ 即为有放回。

二 随机变量与分布函数

§2.1 随机变量

$\mathcal{F}_1,\mathcal{F}_2$ 为 Ω 上的 σ 域，可验证其交亦为 σ 域。更一般地，给定某指标集 I，$\mathcal{F}_i,i\in I$ 的交集亦为 σ 域。
$\mathbb{R}$ 上 Borel 域定义为包含所有 $(a,b],a,b\in\mathbb{R}$ 的**最小** σ 域 (最小定义：所有包含的取交)，记为 $B(\mathbb{R})$。
$\{b\}=\bigcap\limits_n\left(b-\dfrac{1}{n},b\right]\in B(\mathbb{R})$，类似知 $(a,b),[a,b],[a,b)\in B(\mathbb{R})$。
$\mathcal{B}(\mathbb{R}^n)$ 为包含所有左开右闭区间笛卡尔积形成的矩形的最小 σ 域，Borel 域中的集合称 Borel 集。

定义 2.1 随机变量、概率分布函数
$(\Omega,\mathcal{F},P)$ 中，称 $X:\Omega\to\mathbb{R}$ 为一个随机变量，若 $\forall x\in\mathbb{R}$，有 $\{\omega\in\Omega:X(\omega)\le x\}\in\mathcal{F}$。
此时记后方集合为 $\{X\le x\}$，称 $F(x)=P(\{X\le x\})$ 为随机变量 X 的 *(*概率*)* 分布函数。

例 2.1 掷均匀硬币
$\Omega=\{H,T\},X:\Omega\to\mathbb{R},X(H)=1,X(T)=-1$

$$P(\{X\le x\})=\begin{cases}1 & x\ge 1\\ 0.5 & -1\le x<1\\ 0 & x<-1\end{cases}$$

定理 2.1 分布函数 $F(x)$ 性质
1. 单调增
2. 负无穷极限 *0*，正无穷极限 *1*
3. 右连续 $\lim\limits_{\sigma\to 0^+}F(x+\sigma)=F(x)$

证明：
1. 利用包含关系说明。
2. 取一列数趋向正/负无穷，利用概率的极限等于极限的概率知结论。
3. 类似 2，取子列说明。
注：
(1) 若某函数这三条性质，一定为某随机变量的概率分布函数，因此，一般将满足三条性质的函数称为分布函数。
(2) 另一种定义分布函数的方式：$G(x)=P(\{X<x\})$，此时其具有**左**连续性，一二两条不变。
(3) 分布函数**丢失**了关于样本空间的信息，与样本空间无关。

例 2.2 若 $X=c$ 概率为 *1*，称 X 几乎处处常值，则 $F(x)=\begin{cases}1 & x\ge c\\ 0 & x<c\end{cases}$

例 2.3 *Bernoulli* 两点分布

若 $P(X=1)=p, P(X=0)=q, p+q=1$，则 $F(x)=\begin{cases}1 & x\geq 1\\ q & 0\leq x<1\\ 0 & x<0\end{cases}$

特别地，$A\in\mathcal{F}$，示性函数 $I_A(\omega)=\begin{cases}1 & \omega\in A\\ 0 & \omega\notin A\end{cases}$ 亦为随机变量，满足 *Bernoulli* 两点分布，且有 $P(I_A=1)=P(A)$。

定理 2.2 随机变量性质

1. $P(X>x)=1-F(x)$
2. $P(x<X\leq y)=F(y)-F(x)$
3. $P(X=x)=F(x)-F(x^-)$

定理 2.3 设 X 为 $(\Omega,\mathcal{F},P)$ 上随机变量，则任意 *Borel* 集的原象为事件域中元素。

证明：记 $\mathcal{A}=\{A\subset\mathbb{R}:X^{-1}(A)\in\mathcal{F}\}$，由于 $X^{-1}(A^{\mathrm{c}})=X^{-1}(A)^{\mathrm{c}}, X^{-1}\left(\bigcup\limits_n A_n\right)=\bigcup\limits_n X^{-1}(A_n)$ 分别验证三条性质可知 $\mathcal{A}$ 为 $\mathbb{R}$ 上 σ 域，。因此，$(a,b]=(-\infty,b]-(-\infty,a]\in\mathcal{A}$，进一步可知 $B(\mathbb{R})\subset A$，即可得证。
* 随机变量相加后仍为随机变量
证明：令 r_n 为一切有理数的一个排列，证出 $\{X+Y\leq x\}=\bigcap\limits_{n=1}^{\infty}(\{X\leq r_n\}\cup\{Y\leq x-r_n\})$ 即可说明结论。
右包含左易证，故只需证明 ω 不属于左侧时亦不属于右侧。若其不属于左侧，取 m 使 $X(\omega)>r_m>Y(\omega)$ 即发现其不属于右侧。

§2.2　随机向量

定义 2.2 离散型随机变量
随机变量 X 取值至多可列个 $x_1,x_2,\ldots$，则称 X 为离散型随机变量。
记 $p_k=P(X=x_k)$，则 $\{p_k\}$ 为 X 的分布列，此时分布函数 $F(x)=\sum\limits_{k:x_k\leq x}p_k$ 在 x_k 处跳跃，又称原子分布。

定义 2.3 连续型随机变量
若随机变量 X 的分布函数 $F(x)=\displaystyle\int_{-\infty}^{x}f(u)\mathrm{d}u$，其中 f 非负可积，则称 X 为连续型随机变量，称 f 为 X 的密度函数。

注：
(1) 密度函数含义：当 $x=x_0$ 为 f 连续点时，$\Delta x\to 0$，则 $P(x_0<X\leq x_0+\Delta x)=f(x_0)\Delta x$。
(2) 密度函数改变有限多个值仍为密度函数。
(3) $P(X=a)\leq\displaystyle\int_{a-1/n}^{a}f(u)\mathrm{d}u\to 0$，因此 X 在任意有限多个点取值概率为 0。
(4) 若 F 连续且除去有限多个点外 $F'(x)$ 存在且连续，则 X 为连续型随机变量，且 F' 可作为一个密度函数。
(5) X 为连续型随机变量，则 F 绝对连续。

例 2.4 钟表指针

$\Omega=[0,2\pi),\mathcal{F}=B(\mathbb{R})\cap\Omega,P(A)=\dfrac{|A|}{2\pi}$，$|A|$ 指勒贝格测度。
令 $X(\omega)=\omega,Y(\omega)=\omega^2$，则

$$F_X(x)=\begin{cases}x<0 & 0\\ \dfrac{x}{2\pi} & x\in(0,2\pi]\\ x\geq 2\pi & 1\end{cases},F_Y(x)=\begin{cases}x<0 & 0\\ \dfrac{\sqrt{y}}{2\pi} & x\in(0,4\pi^2]\\ x\geq 4\pi^2 & 1\end{cases}$$
，求导知 f_X,f_Y。

分布函数 F 性质：
(1) 单调 $\to$ 不连续点至多可数
(2) **勒贝格分解** $F=c_1F_d+c_2F_c+c_3F_s$ 其中 $c_i\geq 0,\sum\limits_i c_i=1$，$F_d$ 为离散型随机变量的分布函数，F_c 为连续型随机变量的分布函数，F_s 为奇异的。

定义 2.4 随机向量
$X_1,\dots,X_n$ 为 $(\Omega,\mathcal{F},P)$ 上的随机变量，称 $\vec{X}=(X_1,\dots,X_n)$ 为 n 维随机向量，$F(x_1,\dots,x_n)=P(X_1\leq x_1,\dots,X_n\leq x_n)$ 为 $\vec{X}$ 的联合分布函数。
离散型：$\vec{X}$ 取值至多可列多个，联合分布列$f(x_1,\dots,x_n)=P(X_1=x_1,\dots,X_n=x_n)$。
连续型：$F(x_1,\dots,x_n)=\displaystyle\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n}f(u_1,\dots,u_n)\mathrm{d}u_1\dots\mathrm{d}u_n$，$f$ 非负可积，称 f 为 X 的联合密度函数。

定理 2.4 考虑二维随机向量的联合分布函数 $F(x,y)$：
1. $F(x,y)$ 关于 x,y 均单调增。
2. $F(x,y)$ 关于 x,y 均右连续。
3. $F(x,y)$ 在 x,y 趋近负无穷时极限均为 *0*，x,y 均趋近正无穷时极限为 *1*。
4. $x_1\leq x_2,y_1\leq y_2$ 时 $F(x_2,y_2)-F(x_1,y_2)-F(x_2,y_1)+F(x_1,y_1)=P(X\in(x_1,x_2],Y\in(y_1,y_2])\geq 0$

注：
(1) 取极限可发现 4 **蕴含** 1，反之不然 (举例：$F(x)=\begin{cases}1 & x+y\geq 0\\ 0 & x+y<0\end{cases}$ 满足 1,2,3 但不满足 4)。
(2) 若某二元函数满足 2,3,4 三条性质，一定为某随机向量的联合分布函数。
$F_X(x)=P(X\leq x)=\lim\limits_{y\to+\infty}F(x,y)$ 称为**边际分布**。
连续型随机变量 $F_X(x)=\displaystyle\int_{-\infty}^{x}\int_{-\infty}^{+\infty}f(u,v)\mathrm{d}v\mathrm{d}u$，$f_X(x)=\displaystyle\int_{-\infty}^{+\infty}f(x,v)\mathrm{d}v$ 称为**边际密度**。

例 2.5 三项分布
$\Omega=\{H,T,E\}$，均匀“三面硬币”，设扔 n 次后三种次数分别为 H_n,T_n,E_n，有 $H_n+E_n+T_n=n$，则 $P\big((H_n,T_n,E_n)=(h,t,e)\big)=\dfrac{n!}{h!t!e!}\dfrac{1}{3^n}$

例 2.6 $G\subset\mathbb{R}^n$ 为有限区域，则联合密度函数 $f(x_1,\dots,x_n)=\dfrac{1}{|G|},\vec{x}\in G$
特别地，$G=[0,1]^2$ 时，$f(x,y)=\begin{cases}1 & (x,y)\in G\\ 0 & (x,y)\notin G\end{cases}$

三　离散型随机变量

§3.1　分布列与独立性

回顾：离散型随机变量 X 取值至多可列个 $x_1,x_2,\dots$，记 $p_k=P(X=x_k)$，则 $\{p_k\}$ 为 X 的分布列。

例 3.1 二项分布
$P(x=k)=\mathrm{C}_n^k p^k q^{n-k},p+q=1$ 时称 X 符合二项分布，记为 $X\sim B(n,p)$。
背景：抛 n 次硬币，X 为正面向上次数，

例 3.2 几何分布
$P(X=k)=q^{k-1}p,p+q=1$ 时称 X 符合几何分布，此时 $P(X>k)=q^k$。
背景：抛 n 次硬币，X 为第一次正面向上时抛的次数。
几何分布具有无记忆性：$P(X-m=k|X>m)=P(X=k)$。反之，若取值为 $\mathbb{N}^*$ 的某随机变量满足无记忆性，即对任意 m,k 符合上式，则必须服从几何分布。

例 3.3 泊松分布
$P(X=k)=\dfrac{\lambda^k}{k!}e^{-\lambda},\lambda>0$ 时称 X 符合泊松分布，记为 $X\sim P(\lambda)$。
背景：网站访问量、百科新词条

* 放射性粒子数：体积为 V 的小物块分为 n 等份，每一小块 $\Delta v=\dfrac{V}{n}$，假设每一小块在 7.5s 内放出 1 个 α 粒子的概率为 $p=\mu\cdot\Delta v$，放出更多概率为 0，且各小块放出与否相互独立。
分析：n 块共放出 k 个概率符合二项分布，令 $\lambda=\mu V$，则
$P(X=k)=\mathrm{C}_n^k p^k(1-p)^{n-k}=\dfrac{n\dots(n-k+1)}{k!}\left(\dfrac{\lambda}{n}\right)^k\left(1-\dfrac{\lambda}{n}\right)^{n-k}$，固定 k，令 n 趋向无穷，此式极限即为 $\dfrac{\lambda^k}{k!}e^{-\lambda}$。由此可知，二项分布可以**逼近**泊松分布。

定义 3.1 独立性
若 $\forall x,y\in\mathbb{R},P(X=x,Y=y)=P(X=x)P(Y=y)$，则称离散型随机变量 X,Y 独立。
更一般，称 $X_1,\dots,X_n$ 相互独立，若 $\forall x_i\in\mathbb{R},P(X_1=x_1,\dots,X_n=x_n)=P(X_1=x_1)\cdots P(X_n=x_n)$。

定理 3.1 离散型随机变量 X,Y 独立，当且仅当 $P(X\le x,Y\le y)=P(X\le x)P(Y\le y)\Leftrightarrow F(x,y)=F_X(x)F_Y(y)$。

证明：利用分布列 $f(x,y)$ 与分布函数 $F(x,y)$ 关系，利用求和可证明仅当，利用左极限可证明当。

例 3.4 泊松翻转
抛均匀硬币 *1* 次，记 X,Y 为正反出现的次数，计算容易发现不独立。
抛 N 枚均匀硬币，$N\sim P(\lambda)$，计算 $f(x,y)=P(X=x,Y=y,N=x+y)=\dfrac{\lambda^{x+y}}{(x+y)!}e^{-\lambda}\dfrac{\mathrm{C}_{x+y}^x}{2^{x+y}}$
$=\left(\dfrac{\lambda}{2}\right)^x\dfrac{e^{-\lambda/2}}{x!}\left(\dfrac{\lambda}{2}\right)^y\dfrac{e^{-\lambda/2}}{y!}$，注意到 $f_X(x)=\sum\limits_y f(x,y)=\left(\dfrac{\lambda}{2}\right)^x\dfrac{e^{-\lambda/2}}{x!}$，由此知 X,Y 独立。

定理 3.2 离散型随机变量 X,Y 独立，g,h 是 $\mathbb{R}$ 上的 *Borel* 可测函数，则 $g(X),h(Y)$ 独立。

证明：$P(g(X)=a,h(Y)=b)=P\left(\bigcup\limits_{g(x)=a}\{X=x\},\bigcup\limits_{h(y)=a}\{Y=y\}\right)$ 分解求和。

§3.2　数学期望

定义 3.2 数学期望
离散型随机变量 X 对应分布列 f，$\sum\limits_{x:f(x)>0}xf(x)$ 若绝对收敛，则称为 X 的数学期望，记为 $E[X]$。

*$E[x]=\sum\limits_k x_k p_k$(原则上 x_i 互不相同，事实上相同不会影响计算)

定理 3.3 佚名统计学家公式
g 为 $\mathbb{R}$ 上函数，$Y=g(X)$，X 分布列为 f，则 $E[Y]=\sum_x g(x)f(x)$(假定右侧绝对收敛)。

证明：考虑 Y 分布列即可。

定义 3.3 数字特征
离散型随机变量 X 的 k 阶矩为 $m_k=E[X^k]$，k 阶中心矩 $\sigma_k=E[(X-m_1)^k]$。
方差 $\mathrm{Var}(X)$ 为二阶中心矩，$\mathrm{Var}(X)=\sum_x (x-m_1)^2 f(x)=E[X^2]-2m_1E[X]+m_1^2=E[X^2]-E^2[X]$。
标准差定义为 $\sqrt{\mathrm{Var}(X)}$

例 3.5 *Bernoulli* 分布
$P(X=1)=p,P(X=0)=q,p+q=1 \Rightarrow E[X]=p,\mathrm{Var}(X)=p-p^2=pq$

例 3.6 二项分 $X\sim B(n,p)$
$E[X]=\sum_{k=0}^{n} k\mathrm{C}_n^k p^k q^{n-k}=np\sum_{k=0}^{n-1}\mathrm{C}_{n-1}^k p^k q^{n-1-k}=np$
$E[X(X-1)]=\sum_{k=0}^{n} k(k-1)\mathrm{C}_n^k p^k q^{n-k}=n(n-1)p^2$
$E[X^2]=np(np+q),\mathrm{Var}(X)=npq$

定理 3.4 数学期望性质
1. 非负性：$X\geq 0\Rightarrow E[X]\geq 0$
2. 归一性：$E[1]=1$
3. 线性性：$E[aX+bY]=aE[X]+bE[Y]$
由此，E 可以看作一个期望算子。

线性性证明：令**示性函数** $A_x=\{X=x\}$，则 $X=\sum_x xI_{A_x},E[X]=\sum_x xP(A_x)$，对 Y 用 B_y 类似处理，则 $aX+bY=\sum_x xI_{A_x}+\sum_y yI_{B_y}=\sum_{Ax+By} I_{A_x}I_{B_y}$。
* 观点：扩展至量子物理、**非线性**期望

定理 3.5 X,Y 独立且期望存在，$E[XY]=E[X]E[Y]$。

证明：$E[X]=\sum_x xI_{A_x},E[Y]=\sum_y yI_{B_y}\Rightarrow E[X]E[Y]=\sum_{x,y} xyI_{A_xB_y}$。

定理 3.6 方差性质
1. $\mathrm{Var}(aX+b)=a^2\,\mathrm{Var}(X)$
2. $\mathrm{Var}(X+Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)+2(E(XY)-E(X)E(Y))$，$X,Y$ 独立时即可加。

例 3.7 期望不存在的例子
$P(X=x_k)=\dfrac{1}{2^k}$，$x_k=(-1)^k\dfrac{2^k}{k}$，不绝对收敛，$\sum_k x_kp_k=-\ln 2$，而期望不存在。
若 $P(X=x_k)=2^{k-1}$，可发现期望趋向于无穷。

§3.3 协方差

例 3.8 考虑一个随机的 n 元置换 π，记不动点 $\pi(x)=x$ 的个数为 N，求分布列 $P(N=r), r=0,1,\dots,n$。
记 $A_k=\{k$ 为不动点 $\}$，对应的示性函数为 I_k，令 $X=\sum\limits_{i_1<\cdots<i_r}^{i_{r+1}<\cdots<i_n} I_{i_1}I_{i_2}\dots I_{i_r}(1-I_{i_{r+1}})\dots(1-I_{i_n})$，其中 i_k 为 1 到 n 的排列，则 $X=I_{\{N=r\}}$。

$$E[X]=\mathrm{C}_n^r E[I_1\dots I_r(1-I_{r+1})\dots(1-I_n)]=\mathrm{C}_n^r\sum_{s=0}^{n-r}(-1)^s\mathrm{C}_{n-r}^s\frac{(n-r-s)!}{n!}=\frac{1}{r!}\sum_{s=0}^{n-r}\frac{(-1)^s}{s!}$$

例 3.9 不计算分布列，求上例中的期望与方差。
$N=I_1+I_2+\cdots+I_n$。则 $E[N]=nE[I_1]=n\dfrac{(n-1)!}{n!}=1, E[N^2]=\sum\limits_{i,j=1}^{n}E[I_iI_j]=nE[I_1^2]+n(n-1)E[I_1I_2]=1+1=2, \mathrm{Var}(N)=1$。

例 3.10 *Erdos* 概率方法：正 *17* 边形染红 *5* 个顶点，证明存在 *7* 个相邻顶点中至少有 *3* 个红点。
建立模型，*17* 个点中等概率随机取一个，令 I_k 为 k 为红色的示性函数，再令 $X(k)=I_{k+1}+\cdots+I_{k+7}$，则 $E[X]=\dfrac{35}{17}>2$，由此 $P(X>2)>0$，得证。

类似随机变量的情况，对随机向量有结论：$f(x,y)$ 为 (X,Y) 的联合分布列，$g:\mathbb{R}^2\to\mathbb{R}$ 的 Borel 可测函数，则 $E[g(X,Y)]=\sum\limits_{x,y}g(x,y)f(x,y)$(假定右侧绝对收敛)。

定义 3.4 协方差与相关系数
协方差 $\mathrm{Cov}(X,Y)=E[(X-E[X])(Y-E[Y])]=E[XY]-E[X]E[Y]$
相关系数 $\rho(X,Y)=\dfrac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}$，为 *0* 时称两变量不相关。

* 对 n 维随机向量 $\vec{X}=(X_1,\dots,X_n)$，协方差矩阵 $\Sigma=(\sigma_{ij}),\sigma_{ij}=\mathrm{Cov}(X_i,X_j)$
性质：协方差矩阵对称且非负定。
非负定性证明：$\sum\limits_{i,j}t_it_j\sigma_{ij}=\sum\limits_{i,j}t_it_jE[(X_i-E[X_i])(X_j-E[X_j])]=E\left[\left(\sum\limits_i t_i(X_i-E[X_i])\right)^2\right]\geq 0$

定理 3.7 相关系数性质
1. $|\rho|\leq 1$
2. X,Y 独立或不相关 $\Leftrightarrow \rho=0$
3. $\rho=\pm1\Leftrightarrow P(Y=aX+b)=1$

定理 3.8 *Cauchy* 不等式
$(E[XY])^2\leq E[X^2]E[Y^2]$，等号成立当且仅当存在不全为 *0* 实数 a,b 使得 $P(aX=bY)=1$。

证明：先利用佚名统计学家公式说明 $E[X^2]=0$ 时可推出 $P(X=0)=1$，从而计算得 $E[XY]=0$；若 $E[X^2]>0$，$E[(Y-tX)^2]=t^2E[X^2]-2tE[XY]+E[Y^2]\geq 0$，利用判别式得证，取等时再利用 $E[X^2]=0$ 的条件即可。

例 3.11 多项分布
$\vec{X}=(X_1,\dots,X_r), P(X_1=k_1,\dots,X_r=k_r)=\dfrac{n!}{k_1!\dots k_r!}p_1^{k_1}\dots p_r^{k_r}$，其中 $\sum\limits_i p_i=1,\sum\limits_i k_i=n$。
计算 $\mathrm{Cov}(X_i,X_j),\rho(X_i,X_j)$。
$i\neq j$ 时 $X_i\sim B(n,p_i), X_i+X_j\sim B(n,p_i+p_j)$，由此 $E[X_i]=np_i, \mathrm{Var}(X_i)=np_i(1-p_i), \mathrm{Var}(X_i+X_j)=n(p_i+p_j)(1-p_i-p_j)$，利用定理 *3.6* 可知 $\mathrm{Cov}(X,Y)=\dfrac{1}{2}(\mathrm{Var}(X+Y)-\mathrm{Var}(X)-\mathrm{Var}(Y))=-np_ip_j$，相关系数为 $-\sqrt{\dfrac{p_ip_j}{(1-p_i)(1-p_j)}}$。

§3.4　条件分布与条件期望

定义 3.5 条件分布

(X,Y) 为离散型随机向量，给定 $X=x$，且 $P(X=x)>0$，Y 的条件分布列 $f_{Y|X}(y|x)=P(Y=y|X=x)$，对应条件分布函数 $F_{Y|X}(y|x)=P(Y\le y|X=x)$

注：$\sum\limits_y f_{Y|X}(y|x)=1$，其亦为一个分布函数。

$f_{Y|X}(y|x)=\dfrac{f(x,y)}{f_X(x)}, f_X(x)>0$

定义 3.6 条件期望

给定 $X=x$ 下，Y 的条件期望 $\psi(x)=E[Y|X=x]=\sum\limits_y yf_{Y|X}(y|x)$(假设其绝对收敛)，并称 $\psi(X)$ 为 Y 关于 X 的条件期望，记为 $E[Y|X]$。

定理 3.9 全期望公式

$E[E[Y|X]]=E[Y]$ * 另一种形式：$E[Y]=\sum\limits_x f_X(x)E[Y|X=x]$

证明：$E[\psi(X)]=\sum\limits_x \psi(x)f_X(x)=\sum\limits_x f_X(x)\sum\limits_y yf_{Y|X}(y|x)=\sum\limits_{x,y} yf(x,y)=\sum\limits_y yf_Y(y)$，注意步骤中交换次序要求绝对收敛。

* 类似可定义关于事件的条件期望 $E[X|A]=E[X|I_A=1]$。

例 3.12 多项分布的条件期望 $i\ne j, E[X_j|X_i>0]$

利用全期望公式 $E[X_j]=P(X_i=0)E[X_j|X_i=0]+P(X_i>0)E[X_j|X_i>0]$。又由于 $P(X_j=k|X_i=0)=\dfrac{P(X_j=k,X_i=0)}{P(X_i=0)}=\mathrm{C}_n^k\left(\dfrac{p_j}{1-p_1}\right)^k\left(1-\dfrac{p_j}{1-p_1}\right)^{n-k}$，因此 $E[X_j|X_i=0]=\dfrac{np_j}{1-p_i}$，进而算出 $E[X_j|X_i>0]=\dfrac{np_j(1-(1-p_i)^{n-1})}{(1-(1-p_i)^n)}$

例 3.13 鸟下 N 枚蛋，$N\sim P(\lambda)$，每颗蛋独立以概率 p 变为小鸟，记 K 为小鸟数，计算 $E[K|N]$、$E[K]$、$E[N|K]$。

记 $q=1-p$，由于 $f_{K|N}(k|n)=\mathrm{C}_n^kp^kq^{n-k}$，因此 $E[K|N=n]=np$，即 $E[K|N]=Np$，进而 $E[K]=E[E[K|N]]=pE[N]=p\lambda$。

而 $f_{N|K}(n|k)=\dfrac{P(N=n,K=k)}{P(K=k)}=\dfrac{P(K=k|N=n)P(N=n)}{\sum_m P(K=k|N=m)P(N=m)}=\dfrac{(\lambda q)^{n-k}e^{-\lambda q}}{(n-k)!}$，由此 $E[N|K=k]=\sum\limits_{n\ge k} n\dfrac{(\lambda q)^{n-k}e^{-\lambda q}}{(n-k)!}=\sum\limits_{n\ge 0}(n+k)\dfrac{(\lambda q)^n e^{-\lambda q}}{n!}=\lambda q+k$，即 $E[N|K]=\lambda q+K$。

定理 3.10 记 $\psi(X)=E[Y|X]$，g 保证所述期望均存在，则 $E[\psi(X)g(X)]=E[Yg(X)]$。

§3.5　随机游走

$S_0=a, S_n=S_{n-1}+X_n=S_0+\sum\limits_{i=1}^n X_i$，$\{X_i\}$ 相互独立且同分布，取值为 ±1，$P(X_i=1)=p, P(X_i=-1)=q, p+q=1$。

* 直线上简单随机游走，当 $p=\dfrac{1}{2}$ 时称为**对称的**。

例 3.14 自由随机游走，$S_0=a$，求 $P(S_n=b)$，已知 $2|a+b+n$。

设向右游走次数为 r，向左游走次数为 l，则 $\begin{cases} r+l=n \\ r-l=b-a \end{cases} \Rightarrow r=\dfrac{n+b-a}{2}$，因此 $P(S_n=b)=\mathrm{C}_n^{(n+b-a)/2}p^{(n+b-a)/2}q^{(n-b+a)/2}$

定理 3.11 简单随机游走性质

1. 空间齐性 $P(S_n = j + b|S_0 = a + b) = P(S_n = j|S_0 = a)$

2. 时间齐性 $P(S_{n+m} = j|S_m = a) = P(S_n = j|S_0 = a)$

3. 马氏性 $P(S_{m+n} = j|S_0 = j_0, \ldots, S_m = j_m) = P(S_{m+n} = j|S_m = j_m)$

*** 要求等式两边有意义

证明：利用 $P(S_n = j|S_0 = a) = P(\sum_{i=1}^{n} X_i = j - a|S_0 = a)$ 转化即可计算得出结果。

* **轨道计数**

平面表示 $\{(n, S_n) : n = 0, 1, \ldots\}$

记 $N_n(a, b)$ 为 $(0, a) \to (n, b)$ 轨道个数，$N_n^{\circ}(a, b)$ 为 $(0, a) \to (n, b)$ 且访问 x 轴至少一次的轨道个数。

定理 3.12 反射原理

若 $a, b > 0$，则 $N_n^{\circ}(a, b) = N_n(-a, b)$。

证明：寻找第一个交点，利用一一对应。

定理 3.13 $N_n(a, b) = \mathrm{C}_n^{(n+b-a)/2}$

* 关心问题：**返回出发点、游走最远距离、首次击中某点**

定理 3.14 投票定理

$b > 0$，则 $(0, 0) \to (n, b)$ 不再过 x 轴的轨道个数为 $(1, 1) \to (n, b)$ 不过 x 轴的轨道个数，即 $N_{n-1}(1, b) - N_{n-1}^{\circ}(1, b) = \dfrac{b}{n} N_n(0, b)$。

例 3.15 A 得票 a，B 得票 $b < a$，求 A 得票始终大于 B 的概率。

问题可化为 $(0, 0)$ 到 $(a + b, a - b)$ 轨道中不再过 x 轴轨道数与总数之比，为 $\dfrac{a - b}{a + b}$。

定理 3.15 不返回出发点

$S_0 = 0, n \neq 1$，则 $P(S_1 \ldots S_n \neq 0, S_n = b) = \dfrac{|b|}{n} P(S_n = b)$。

由此可推得 $P(S_1 \ldots S_n \neq 0) = \dfrac{E[|S_n|]}{n}$

证明：利用投票定理计算即可。

* 记最到达的最右端 $M_n = \max\{S_i : 0 \le i \le n\}$，$S_0 = 0 \Rightarrow M_n \ge 0$

定理 3.16 最右端

$$P(M_n \ge r, S_n = b) = \begin{cases} P(S_n = b) & b \ge r \\ \left(\dfrac{q}{p}\right)^{r-b} P(S_n = 2r - b) & b < r \end{cases}$$

证明：考虑第一次到达 r 的点，反射其后的部分，即可与 $S_n = 2r - b$ 的轨道数量一一对应，而反射产生的概率差别为 $\left(\dfrac{q}{p}\right)^{r-b}$。

* 推论：$p = q$ 时可以计算出 $P(M_n \ge r) = P(S_n = r) + 2P(S_n \ge r + 1)$。

定理 3.17 首中时定理

$S_0 = 0$，时刻 n 首次击中 b 概率为 $f_b(n) = \dfrac{|b|}{n} P(S_n = b)$。

证明：不妨设 $b>0$，注意条件等价于 $M_{n-1}=S_{n-1}=b-1, S_n=b$ 即可算出结果。

定理 3.18 反正弦律
对称随机游走，$S_0=0$，记 $T_{2n}=\max\{0\le i\le 2n: S_i=0\}$，则 $P(T_{2n}=2k)=P(S_{2k}=0)P(S_{2n-2k}=0)$

证明：利用 $P(S_{2k+1}\dots S_{2n}\neq 0|S_{2k}=0)=P(S_1\dots S_{2n-2k}\neq 0|S_0=0)$。此外 $P(S_1\dots S_{2m}\neq 0)=P(S_2\dots S_{2m}\neq 0|S_1=1)=P($ 第 $2m$ 次后首次击中 $-1)$，由此可计算出结论。
* 此分布称为**反正弦律**
利用 Stirling 公式可计算 $P(T_n\le 2xn)\sim \dfrac{1}{n}\sum\limits_{\frac{k}{n}<x}\dfrac{1}{\pi}\dfrac{1}{\sqrt{\frac{k}{n}(1-\frac{k}{n})}}\sim\int_0^x\dfrac{1}{\pi\sqrt{u(1-u)}}\mathrm{d}u=\dfrac{2}{\pi}\arcsin\sqrt{x}$，$\dfrac{T_{2n}}{2n}$ 渐近分布 $\dfrac{2}{\pi}\arcsin\sqrt{x}$。

$\mathbb{Z}^d$ 上的随机游走：$S_n=S_{n-1}+X_n=S_0+\sum\limits_{i=1}^{n}X_i$，向量 $\{X_i\}$ 相互独立且同分布，仅有一个不为 0 且取值为 ±1 的分量。

例 3.16 平面上对称随机游走，求 $P(S_{2n}=\mathbf{0})$。
考虑上下左右次数知结果为 $\sum\limits_{k=0}^{n}\dfrac{(2n)!}{k!k!(n-k)!(n-k)!}\dfrac{1}{4^{2n}}=C_{2n}^{n}\dfrac{1}{4^{2n}}\sum\limits_{k=0}^{n}(\mathrm{C}_n^k)^2=\dfrac{1}{4^{2n}}(\mathrm{C}_{2n}^{n})^2$。

§3.6 母函数

* 数列的母函数：$\{a_i, i\in\mathbb{N}\}$，母函数 $G_a(s)=\sum\limits_{i=0}^{\infty}a_is^i$，如 C_n^i 对应 $(1+s)^n$

卷积：$c_n=\sum\limits_{k=0}^{n}a_kb_{n-k}$，记为 $c=a*b$，验证有 $G_c(s)=G_a(s)G_b(s)$。

例 3.17 随机游走 $S_0=0$，求 $S_{2n}=0, S_i\ge 0$ 的轨道个数 C_n。
设首次返回原点为 $2k$ 时，则考虑第一步后可知从 *0* 到 $2k$ 的轨道个数为 C_{k-1}，因此有 $C_n=\sum\limits_{k=0}^{n-1}C_{k-1}C_{n-k}$，于是 $G(s)-1=sG^2(s)$，考虑合理解知 $G(s)=\dfrac{1-\sqrt{1-4s}}{2s}$，展开可得 $C_n=\dfrac{(2n)!}{n!(n+1)!}$*(*卡特兰数*)*。

定义 3.7 非负整值随机变量的母函数
$G_X(s)=E[S^X]=\sum\limits_{i=1}^{\infty}s^iP(X=i)=\sum\limits_{i=1}^{\infty}f(i)s^i$
* 其收敛半径 $R\ge 1$，在收敛域内可微
*$G(0)=P(X=0), G(1)=1, f(i)=\dfrac{G^{(i)}(0)}{i!}$
* 只要系数非负，$G(1)=1$，即可看作概率母函数

1. 二项分布 $B(n,p)$ 母函数 $(ps+q)^n$
2. 几何分布 $P(X=i)=pq^{i-1}$ 母函数 $\dfrac{ps}{1-qs}$
3. 泊松分布 $P(\lambda)$ 母函数 $e^{\lambda(s-1)}$

定理 3.19 母函数与数字特征矩
1. $E[X]=G'(1)$
2. $E[x(x-1)\dots(x-k+1)]=G^{(k)}(1)$
3. $\mathrm{Var}(X)=G''(1)-G'(1)(1-G'(1))$
此处 $G^{(k)}(1)$ 指 $\lim\limits_{x\to 1^-}G^{(k)}(x)$，由阿贝尔定理此定义合理。

证明：直接求导即可说明。

定理 3.20 随机变量卷积
设非负整值随机变量 $X_1,\dots,X_n$ 互相独立，$Y=\sum_{k=1}^{n}X_k$，则 $G_Y=\prod_{k=1}^{n}G_{X_k}$。

证明：利用卷积归纳即可。

定理 3.21 复合分布
设 X_i 独立同分布，母函数 G_X，N 与 X_i 独立，母函数 G_N，$Y=\sum_{k=1}^{N}X_k$ 母函数为 $G_N(G_X)$。

证明：利用全概率公式分解计算。

定义 3.8 联合母函数
X,Y 联合母函数 $G(s,t)=E[s^xt^y]=\sum_{i,j}s^it^jP(X=i,Y=j)$。

定理 3.22 X,Y 独立等价于 $G(s,t)=G_X(s)G_Y(t)$

证明：比较系数。

例 3.18 掷三颗均匀骰子，求点数和分布。
设每个骰子点数 X_i，和为 Y，则 $G_Y(s)=G_X^3(s)=\dfrac{s^3(1-s^6)^3}{6^3(1-s)^3}$，考虑每项系数即为分布。

* 二项分布**再生性**：独立变量 $X_1\sim B(n_1,p),X_2\sim B(n_2,p)$，考虑母函数可发现 $X_1+X_2\sim B(n_1+n_2,p)$。
泊松分布也有再生性：X_1,X_2 独立泊松分布 $P(\lambda_1),P(\lambda_2)$，其和分布为 $P(\lambda_1+\lambda_2)$。
* 离散随机变量的母函数是否可以定义为 $G(s)=\sum_i s^{x_i}P(X=x_i)$？其他类型随机变量呢？(第五章内容)

四 连续型随机变量

§4.1 独立性

X 连续型随机变量，有**密度函数** f，非负且在实轴上积分为 1，分布函数 $F(x)=\int_{-\infty}^{x}f(u)du$。

*$P(X=x)=0,P(a<X<b)=\int_a^b f(u)du$
* **均匀分布** $X\sim U[a,b]$
$f(x)=\dfrac{1}{b-a},x\in[a,b]$
* **指数分布** $X\sim \mathrm{Exp}(\lambda)$
$f(x)=\lambda\mathrm{e}^{-\lambda x},F(x)=1-\mathrm{e}^{-\lambda x}$
背景：$P(t\le X\le t+\Delta t|X>t)=\lambda\Delta t$ 微分方程 (如半衰期等)
也具有**无记忆性**
* **正态分布** $X\sim N(\mu,\sigma^2)$
$f(x)=\dfrac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\dfrac{1}{2\sigma^2}(x-\mu)^2\right)$
μ 为对称轴且为最大值点
$\mu\pm\sigma$ 为拐点
背景：**随机误差**分布
Wigner **半圆律**
$f(x)=\dfrac{1}{2\pi\sigma^2}\sqrt{4\sigma^2-x^2}$
背景：随机矩阵、自由概率论 (和正态分布同地位)

例 4.1 半圆律中 $P(X \in (0,\sigma))$
积分可得结果为 $\dfrac{1}{6}+\dfrac{\sqrt{3}}{4\pi}$

例 4.2 $X \sim N(0,1)$，求 $Y=X^2$ 密度函数
$P(Y \le y)=P(-\sqrt{y}<X<\sqrt{y})=\displaystyle\int_{-\sqrt{y}}^{\sqrt{y}}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\mathrm{d}x$，求导得密度函数为 $\dfrac{\mathrm{e}^{-y/2}}{\sqrt{2y\pi}}$。

* 正态分布常用 $\phi(x)$ 表示密度函数，$\Phi(x)=\displaystyle\int_{-\infty}^{x}\phi(u)\mathrm{d}u$，$N(0,1)$ 称标准正态分布。
*$X \sim U(0,1)$，计算可得 $Y=\Phi^{-1}(X)\sim N(0,1)$

定义 4.1 一般随机变量的独立性
$X_1,\dots,X_n$ 满足 $P(X_1 \le x_1,\dots,X_n \le x_n)=\displaystyle\prod_{k=1}^{n}P(X_k \le x_k)$，则称随机变量相互独立。

定理 4.1 等价刻画 $X_1,\dots,X_n$ 相互独立等价于 $\forall B_1,\dots,B_n \in B(\mathbb{R}), P(\forall i, X_i \in B_i)=\displaystyle\prod_{k=1}^{n}P(X_k \in B_k)$

(证明须更多测度论知识)

定理 4.2 设 $g_1,\dots,g_n$ 均 *Borel* 可测，$X_1,\dots,X_n$ 独立，则 $g_1(X_1),\dots,g_n(X_n)$ 独立。

证明：利用 Borel 集原像为 Borel 集可由上一定理说明。

定理 4.3 连续型随机变量的独立
$X_1,\dots,X_n$ 密度函数 $f_1,\dots,f_n$，独立等价于 $f_1\dots f_n$ 为联合密度函数。

证明：利用多重积分性质计算。
* **卷积**：X,Y 独立时，称 $Z=X+Y$ 为其卷积，$f_Z=f_X * f_Y$

定理 4.4 (X,Y) 密度函数为 $f(x,y)$，则 $f_{X+Y}(x)=\displaystyle\int_{\mathbb{R}}f(x,z-x)\mathrm{d}x=\int_{\mathbb{R}}f(z-y,y)\mathrm{d}y$

证明：利用二重积分变量代换公式计算。
*X,Y 独立时，$f(x,y)=f_X(x)f_Y(y)$，由此 $X+Y$ 密度函数成为两函数卷积

例 4.3 X,Y 独立标准正态分布，求 $X+Y$ 分布函数。
利用定理计算积分得 $X+Y \sim N(0,2)$。

§4.2 期望

定义 4.2 连续型随机变量的期望
当 $\displaystyle\int_{\mathbb{R}}xf(x)\mathrm{d}x$ 绝对收敛，定义其为 X 的数学期望 $E[X]$。

定理 4.5 X 有密度函数 f，期望存在，则 $E[X]=\displaystyle\int_0^{\infty}(1-F_X(x))\mathrm{d}x-\int_{-\infty}^{0}F_X(x)\mathrm{d}x$。

证明：利用分部积分公式计算。

定理 4.6 复合的期望
g 为 *Borel* 可测函数，X 与 $g(X)$ 均为连续型随机变量，则 $E[g(X)]=\displaystyle\int_{\mathbb{R}}g(x)f_X(x)\mathrm{d}x$

证明：利用上一定理由 F 推导计算。

定理 4.7 连续型随机向量情形

g 是二元 *Borel* 可测函数, X,Y 联合分布 $f(x,y)$, $g(X,Y)$ 为连续型随机向量且期望存在, 则 $E[g(X,Y)] = \int_{\mathbb{R}^2} g(x,y)f(x,y)\mathrm{d}x\mathrm{d}y$。特别地，当 $g=ax+by$ 时代入可发现 $E[aX+bY]=aE[X]+bE[Y]$。

* 可用于计算**协方差**

矩 $m_k = E[X^k]$，方差 $\mathrm{Var}(X)=E[X^2]-E^2[X]$

协方差 $\mathrm{Cov}(X,Y)=E[XY]-E[X]E[Y]$，相关系数 $\rho(X,Y)=\dfrac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}$

定理 4.8 柯西不等式

(X,Y) 连续型随机向量，且 $E[X^2],E[Y^2]$ 存在，则 $(E[XY])^2 \le E[X^2]E[Y^2]$

证明：与离散情形相同构造。

* 由此知 $|\rho|\le 1$，当 $P(Y=aX)=1$ 时可取等。

例 4.4 正态分布

积分计算期望：$\int_{\mathbb{R}} x\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)\mathrm{d}x = \int_{\mathbb{R}}(x-\mu)\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)\mathrm{d}x+\mu$
$= \int_{\mathbb{R}} x\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}x^2\right)\mathrm{d}x+\mu=\mu$。

同样直接代入计算可知方差为 σ^2。

例 4.5 柯西分布

$f(x)=\dfrac{1}{\pi(1+x^2)}$，积分可知期望不存在。

例 4.6 二元正态分布

$f(x,y)=\dfrac{1}{\sqrt{2\pi(1-\rho^2)}}\exp\left(-\dfrac{x^2-2\rho xy+y^2}{2(1-\rho^2)}\right),\rho\in(-1,1)$

计算可知 $f_X(x)=\dfrac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}$，因此 $X\sim N(0,1)$，由对称性知 Y 亦如此，$E[X]=E[Y]=0$。

$\mathrm{Cov}(X,Y)=E[XY]=\int_{\mathbb{R}^2}xyf(x,y)\mathrm{d}x\mathrm{d}y$ 将 xy 写为 $x(y-\rho x)+\rho x^2$，可发现左侧项积分为 *0*，考虑右侧按先 y 后 x 积分得 ρ。

此时 ρ 即为相关系数，$\rho=0 \Leftrightarrow X,Y$ 独立。

条件期望：直观上考虑 $P(Y\le y|x<X\le x+\Delta x),\Delta x\to 0$，可得 $\dfrac{\int_{-\infty}^{y}f(x,u)\mathrm{d}u}{f_X(x)}$

定义 4.3 设 (X,Y) 联合密度函数 $f(x,y)$，X 分布为 $f_X(x)$

条件密度 $f_{Y|X}(y|x)=\dfrac{f(x,y)}{f_X(x)}$ 关于 y 构成密度函数。

$F_{Y|X}(y|x)=\dfrac{\int_{-\infty}^{y}f(x,u)}{f_X(x)}\mathrm{d}u$

条件期望 $E[Y|X=x]=\int_{\mathbb{R}}yf_{Y|X}(y|x)\mathrm{d}y$，由此有随机变量 $E[Y|X]$。

定理 4.9 期望形式的全概率公式

$E[E[Y|X]]=E[Y]$

例 4.7 二元正态分布的条件期望

关于 y 的分布为 $N(\rho x,1-\rho^2)$，由此 $E[Y|X]=\rho X$。

§4.3 多元正态分布

一般分析结论：随机向量 (X_1,X_2) 密度函数 $f(x_1,x_2)$，映射 $T:(X_1,X_2)\to(Y_1,Y_2)$，T 为将 $D\subset\mathbb{R}^2$ 映射到 $R\subset\mathbb{R}^2$ 的一一映射，$T:(x_1,x_2)\to(y_1(x_1,x_2),y_2(x_1,x_2))$，设其逆映射 $(x_1(y_1,y_2),x_2(y_1,y_2))$ 有连续偏导数。则 $f_{Y_1,Y_2}(y_1,y_2)=f(x_1,x_2)|J|I_{T(D)}$，其中 J 是 (x_1,x_2) 的 Jacobi 行列式。
* 证明：利用变量替换，考虑 (Y_1,Y_2) 取在 Borel 集 $(-\infty,y_1]\times(-\infty,y_2]$ 中的概率。
(若在 $P(x_1,x_2\in D_0)=1,D_0\subset D$ 上一一映射，结论仍成立)

例 4.8 X,Y 独立 $\sim N(0,1)$，$X=R\cos\Theta,Y=R\sin\Theta$，$R\ge 0,\Theta\in[0,2\pi]$，求 R,Θ 分布。
$f_{X,Y}(x,y)=\dfrac{1}{2\pi}\mathrm{e}^{-x^2/2-y^2/2}$，由此 $f_{R,\Theta}(r,\theta)=\dfrac{1}{2\pi}r\mathrm{e}^{-r^2/2},r\ge 0,\theta\in(0,2\pi]$，有 $f_\Theta(\theta)=\dfrac{1}{2\pi},f_R(r)=r\mathrm{e}^{-r^2/2}$。

*U_1,U_2 独立 $\sim U(0,1)$，则 $X=\sqrt{-2\ln U_1}\cos(2\pi U_2),Y=\sqrt{-2\ln U_1}\sin(2\pi U_2)$ 为独立标准正态分布。
一元正态分布 $N(\mu,\sigma^2)$，密度函数 $f(x)=\dfrac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\dfrac{1}{2\sigma^2}(x-\mu)^2\right)$
二元正态分布 $f(x,y)=\dfrac{1}{\sqrt{2\pi(1-\rho^2)}}\exp\left(-\dfrac{x^2-2\rho xy+y^2}{2(1-\rho^2)}\right)$
一般情况：$C_n\mathrm{e}^{-Q(x_1,\dots,x_n)}$ 二次型 $Q=\sum_{i,j}x_ix_jA_{ij},A^T=A$，需**加条件**

定义 4.4 多元正态分布
$\vec{X}=(X_1,\dots,X_n)$ 密度函数 $f(\vec{x})=\dfrac{1}{\sqrt{(2\pi)^n|\Sigma|}}\exp(-\dfrac{1}{2}(\vec{x}-\vec{\mu})\Sigma^{-1}(\vec{x}-\vec{\mu})^T)$，$\Sigma=(\sigma_{ij})$ 正定，则记 $\vec{X}\sim N(\vec{\mu},\Sigma)$
*$n=2$ 时可简化表达

定理 4.10 参数性质
设 $\vec{X}\sim N(\vec{\mu},\Sigma)$，则
1. $E[\vec{X}]=\vec{\mu}$
2. $E[(\vec{X}-\vec{\mu})^T(\vec{X}-\vec{\mu})]=\Sigma$，即 $\sigma_{ij}=\mathrm{Cov}(X_i,X_j)$

证明：设 $\Sigma=B^T\Lambda B$ 为正交相似对角化，分析计算。

定理 4.11 线性变换下不变性
$\vec{X}\sim N(\vec{\mu},\Sigma)$，$D$ 为 n 阶可逆方阵，则 $\vec{X}D\sim N(\vec{\mu}D,D^T\Sigma D)$。

证明：对分量分别计算可知结论。

定理 4.12 $\vec{X}\sim N(\vec{\mu},\Sigma)$, $\Sigma=\mathrm{diag}(\Sigma_1,\Sigma_2)$, 两分块均为方阵, 将向量与期望对应拆分为 $\vec{X}_1,\vec{X}_2$ 与 $\vec{\mu}_1,\vec{\mu}_2$，则 $\vec{X}_1\sim N(\vec{\mu}_1,\Sigma_1),\vec{X}_1\sim N(\vec{\mu}_1,\Sigma_2)$。

证明：分离变量得结果。

定理 4.13 $\vec{X}\sim N(\vec{\mu},\Sigma)$，$\Sigma=\begin{pmatrix}\Sigma_{11}&\Sigma_{12}\\\Sigma_{21}&\Sigma_{22}\end{pmatrix}$，将随机向量与期望对应拆分为 $\vec{X}_1,\vec{X}_2$ 与 $\vec{\mu}_1,\vec{\mu}_2$，则 $\vec{X}_1\sim N(\vec{\mu}_1,\Sigma_{11}),\vec{X}_1\sim N(\vec{\mu}_1,\Sigma_{22})$。

证明：对角化后利用分部积分计算。

定理 4.14 更广泛的线性变换下不变性
$\vec{X}\sim N(\vec{\mu},\Sigma)$，$D$ 为 $n\times m$ 阶列满秩方阵，则 $\vec{X}D\sim N(\vec{\mu}D,D^T\Sigma D)$。

证明：与之前类似，拆分计算。

定理 4.15 独立性
$\Sigma = \mathrm{diag}(\Sigma_1, \Sigma_2)$，各分量相互独立当且仅当协方差矩阵对角。

证明：分离变量得结果。

* 定义**均值** $\bar{X} = \dfrac{X_1 + \cdots + X_n}{n}$，**方差** $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

定义 4.5 卡方分布
当密度函数 $f(x) = \dfrac{1}{2^{d/2}\Gamma(d/2)} x^{d/2-1}e^{-x/2}, x > 0$，称 X 服从 d 个自由度卡方分布，记为 $X \sim \chi^2(d)$。

定理 4.16 卡方分布性质
$Y_1, \ldots, Y_d$ 独立同分布 $N(0,1)$，则 $\sum_{i=1}^{d} Y_i^2 \sim \chi^2(d)$。

证明：利用极坐标换元计算。

定理 4.17 均值、方差性质
设 X_i 独立同分布 $N(\mu, \sigma^2)$，则：
1. $\bar{X} \sim N(\mu, \sigma^2/n)$
2. $(n-1)/\sigma^2 S^2 \sim \chi^2(n-1)$
3. $\bar{X}, S^2$ 独立

复平面二维随机向量：$Z = X + Y\mathrm{i}$
$E[Z] = E[X] + \mathrm{i}E[Y]$
复高斯分布：$N_{\mathbb{C}}(\mu, \sigma^2), \mu \in \mathbb{C}, \sigma > 0$，联合密度 $f(z) = \dfrac{1}{\pi\sigma^2}\exp\left(-\dfrac{1}{\sigma^2}|z-\mu|^2\right)$。
结论：$Z \sim N_{\mathbb{C}}(0,1) \Rightarrow E[Z^k\overline{Z^l}] = \begin{cases} k! & k = l \\ 0 & k \neq l \end{cases}$

五　中心极限定理

§5.1　一般随机变量的期望

1. 记号准备
对随机变量 X 与分布函数 F，离散型/连续型具有分布列/分布函数 f，从而对 $xf(x)$ 求和/积分可得期望。而由于 $\mathrm{d}F$ 分别为 $F(x) - F(x^-)/f(x)\mathrm{d}x$，期望可统一写为 $E[X] = \int_{\mathbb{R}} x\mathrm{d}F$，佚名统计学家公式也可写为 $E[g(X)] = \int_{\mathbb{R}} g(x)\mathrm{d}F$。

2. 抽象积分
对一般 $(\Omega, \mathcal{F}, P)$，随机变量 X 与分布函数 F，如何定义 $E[X]$？
STEP 1 对**简单** (取有限个值) 随机变量
可记为 $X = \sum_{i=1}^{n} x_i I_{A_i}$，$A_1, \ldots, A_n$ 为 Ω 的划分，则 $E[X] = \sum_{i=1}^{n} x_i P(A_i)$。
STEP 2 对**非负**随机变量
存在单调增的简单随机变量列，收敛到此随机变量：记 $X_n = nI_{A_n} + \sum_{j=1}^{n2^n} \dfrac{j-1}{2^n} I_{A_{nj}}$，其中 $A_{nj} = \{\dfrac{j-1}{2^n} \leq X < \dfrac{j}{2^n}\}, A_n = \{X \geq n\}$。

再说明若简单随机变量序列 X_n, Y_n 均单调增收敛至 X，则期望的极限相同 (此处相同包含正无穷，由此可由期望的极限定义 X 的期望)。
STEP 3 对**一般**随机变量
将一般随机变量 X 写为 $X^+ - X^-, X^+ = \max\{X, 0\}, X^- = -\max\{-X, 0\}$，当 $E[X^+], E[X^-]$ 至少一个有限时，可定义 $E[X] = E[X^+] - E[X^-]$。特别地，若两者都有限，则称 X 的**数学期望**存在。
* 统一记号为 $E[X] = \displaystyle\int_\Omega X \mathrm{d}P$ 或 $\displaystyle\int_\Omega X(\omega) P(\mathrm{d}\omega)$

3 期望性质
非负性：$X \geq 0 \Rightarrow E[X] \geq 0$ (由定义过程知成立)
规范性：$c \in \mathbb{R} \Rightarrow E[c] = c$
线性性：$E[aX + bY] = aE[X] + bE[Y]$
“**连续**性”：X_n 趋向 X(或偏移的概率为 0)，只要满足以下条件之一即有期望亦有极限：

1. 单调收敛：X_n 单调
2. 控制收敛：$|X_n| \leq Y, E[Y] < \infty$
3. **有界收敛**：$|X_n| \leq c, c \in \mathbb{R}$

4 Lebesgue-Stieltjes 积分
对随机变量 X 与分布函数 F，引入 $(\mathbb{R}, R(\mathbb{R}))$ 上概率测度 $\mu_F((a, b]) = F(b) - F(a)$，则 $(\mathbb{R}, R(\mathbb{R}), \mu_F)$ 构成概率空间。对 Borel 可测函数 g，有抽象积分 $\displaystyle\int g \mathrm{d}F$，称为 Lebesgue-Stieltjes 积分。
有结论：$E[g(X)] = \displaystyle\int g \mathrm{d}F$。
证明：回到三步的定义方式**逐步证明**。
* 此结论对多元也成立
***Fatou 引理**：X_i 非负随机变量，则 $E[\liminf\limits_{n\to\infty} X_n] \leq \liminf\limits_{n\to\infty} E[X_n]$

5 独立随机变量乘积期望与期望乘积相同

定理 5.1 X, Y 为独立随机变量，且期望与乘积的均存在，则 $E[XY] = E[X]E[Y]$

证明：
对简单随机变量：直接拆分计算即可。
对非负随机变量：注意到可取出两列递增的**独立**简单随机变量趋向 X_n 与 Y_n，由此得证。
对一般随机变量：利用线性性拆分计算即可。

§5.2 特征函数

* 非负整值母函数定义：$G_X(s) = \displaystyle\sum_{k=0}^{\infty} P(X = k) s^k = E[s^X]$
当 $s = \mathrm{e}^t$，有**矩母函数** $M_X(t) = E[\mathrm{e}^{tX}]$
* “**好**”的情形：存在 0 的邻域使矩母函数存在
* 不好的例子：$f(x) = \dfrac{1}{\pi(1 + x^2)}$

定义 5.1 X, Y 为 $(\Omega, \mathcal{F}, P)$ 上随机变量，称 $Z = X + Yi$ 为复随机变量。

(1) 实质为二维随机向量
(2) Z_1, Z_2 独立，指 (X_1, Y_1) 与 (X_2, Y_2) 独立，即 $P(X_1 \le x_1, Y_1 \le y_1, X_2 \le x_2, Y_2 \le y_2) = P(X_1 \le x_1, Y_1 \le y_1)P(X_2 \le x_2, Y_2 \le y_2)$。
(3) $E[Z] = E[X] + \mathrm{i}E[Y]$
(4) Z_i 相互独立时，$E[Z_1 \dots Z_n] = E[Z_1] \dots E[Z_n]$。

定义 5.2 特征函数
$\phi(t) = E[\mathrm{e}^{\mathrm{i}tX}]$，有时用 $\phi_X(t)$ 表示。

(1) $\phi(t) = E[\cos(tX)] + \mathrm{i}E[\sin(tX)]$
(2) 由于 $|\mathrm{e}^{\mathrm{i}tx}| = 1$，$\phi(t)$ 总存在。
(3) $\phi(t) = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}tx}\mathrm{d}F(x)$，有分布函数时可写为 $\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}tx} f(x)\mathrm{d}x$。

定理 5.2 基本性质
1. $\phi(0) = 1, |\phi(t)| \le 1, \overline{\phi(t)} = \phi(-t)$
2. ϕ 在 $\mathbb{R}$ 上一致连续。
3. 非负定性：$t_i \in \mathbb{R}, z_i \in \mathbb{C}$，有 $\sum_{j,k=1}^{n} \phi(t_j - t_k) z_j \overline{z_k} \ge 0$。

证明：
2. $|\phi(t+h) - \phi(t)| \le \int_{\mathbb{R}} |\mathrm{e}^{\mathrm{i}tx}(\mathrm{e}^{\mathrm{i}hx} - 1)|\mathrm{d}F \le \int_{\mathbb{R}} |\mathrm{e}^{\mathrm{i}hx} - 1|\mathrm{d}F$，有界收敛定理可知一致趋于 0。
3. 原式 $= E\Big[\big|\sum_j z_j \mathrm{e}^{\mathrm{i}t_j x}\big|^2\Big] \ge 0$。
* 满足定理 5.2 三条性质的函数必为某随机变量的特征函数

定理 5.3 $E[|X|^k] < \infty$，则 $\forall j < k, \phi^{(j)}(0) = \mathrm{i}^j E[X^j]$，进而 $\phi(t) = \sum_{j=0}^{k} \frac{(\mathrm{i}t)^j}{j!} E[X^j] + o(t^k)$。

证明：由题 6.5.4 知 $E[|X|^j] < \infty$，由此积分与求导可交换，再由求导结果可知成立。

定理 5.4 两变量特征函数关系
1. $Y = aX + b$ 时 $\phi_Y(t) = \mathrm{e}^{\mathrm{i}bt}\phi_X(at)$。
2. X, Y 独立时 $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$。

证明：直接计算即可。

定义 5.3 多元特征函数
$\vec{X} = (X_1, \dots, X_n)$ 的特征函数为 $\phi_{\vec{X}}(\vec{t}) = E[\mathrm{e}^{\mathrm{i}\sum_j t_j X_j}]$。

定理 5.5 独立性
X, Y 独立当且仅当 $\phi_{X,Y}(s,t) = \phi_X(s)\phi_Y(t)$。

证明：左推右可直接计算，右推左需要**反转公式** (见下节)。

例 5.1 *Bernoulli* 分布
$\phi(t) = p\mathrm{e}^{\mathrm{i}t} + q$
由此亦可知二项分布母函数为 $(p\mathrm{e}^{\mathrm{i}t} + q)^n$

* 对非负整值随机变量，母函数 G，则 $G(\mathrm{e}^{\mathrm{i}t})$ 即为特征函数。

例 5.2 指数分布
$f(x)=\lambda\mathrm{e}^{-\lambda x}, x>0$
$\phi(t)=\dfrac{\lambda}{\lambda-\mathrm{i}t}$

证明：可分别计算实部虚部积分，亦可通过复分析直接计算。

例 5.3 标准正态分布 $N(0,1)$
$\phi(t)=\dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{\mathbb{R}}\mathrm{e}^{itx-x^2/2}\mathrm{d}x=\mathrm{e}^{-t^2/2}$

证明：“物理方法”假设 $\mathrm{i}t$ 为实数，再解析延拓；“数学方法”计算导数说明。
*$N(\mu,\sigma^2)$ 的分布函数为 $\mathrm{e}^{\mathrm{i}\mu t-\sigma^2t^2/2}$

例 5.4 多元正态分布 $N(\vec{\mu},\Sigma)$
$\phi(\vec{t}\,)=\exp\left(\dfrac{\mathrm{i}\vec{\mu}\vec{t}^{\,T}-\vec{t}\,\Sigma\vec{t}^{\,T}}{2}\right)$

证明：设 $Y=\vec{X}\cdot\vec{t}^{\,T}$，将其化为一元正态分布的情况。

例 5.5 均匀分布 $U(-1,1)$
$\phi(t)=\dfrac{\sin t}{t}$

§5.3 反转与连续性定理

定理 5.6 反转公式
$-\infty<a<b<\infty, \dfrac{F(b)+F(b^-)}{2}-\dfrac{F(a)+F(a^-)}{2}=\displaystyle\lim_{T\to\infty}\int_{-T}^{T}\dfrac{\mathrm{e}^{-\mathrm{i}at}-\mathrm{e}^{-\mathrm{i}bt}}{2\pi\mathrm{i}t}\phi(t)\mathrm{d}t$

理解：类似**傅里叶反变换**
证明：记极限中的积分为 I_T，$I_T=\displaystyle\int_{-T}^{T}\dfrac{\mathrm{e}^{-\mathrm{i}at}-\mathrm{e}^{-\mathrm{i}bt}}{2\pi\mathrm{i}t}\int_{\mathbb{R}}\mathrm{e}^{\mathrm{i}tx}\mathrm{d}F\mathrm{d}t$，由 **Fubini 定理**积分符号可交换，由积分区域对称可转化为 $\displaystyle\int_{\mathbb{R}}g_T(x)\mathrm{d}F, g_T(x)=\dfrac{1}{\pi}\int_0^T\left(\dfrac{\sin t(x-a)}{t}-\dfrac{\sin t(x-b)}{t}\right)\mathrm{d}t$。由于 $\displaystyle\int_0^{+\infty}\dfrac{\sin tx}{\pi t}=\begin{cases}1/2 & x>0\\0 & x=0\\-1/2 & x<0\end{cases}$，$g_T(x)$ 有界，由控制收敛定理其极限为 $\begin{cases}1 & x\in(a,b)\\1/2 & x=a,b\\0 & \text{其他}\end{cases}$，因此结果为 $P(X\in(a,b))+\dfrac{P(X=a)+P(X=b)}{2}$，即为左侧。
* 推论：**特征函数相同即可知同分布**
证明：记 C_F 为 F 连续点，$\mathbb{R}\backslash C_F$ 至多可数。让 $a,b\in C_F, a\to-\infty$，可唯一确定 $F(b)$，由此连续点已唯一确定。再用连续点从右侧逼近可唯一确定不连续点处。

定理 5.7 假设随机向量对任何长方体，落入其表面概率为 0，则有：
$P(a_j<X_j\le b_j, j=1,\ldots,n)=\displaystyle\lim_{T_i\to\infty}\int_{-T_n}^{T_n}\cdots\int_{-T_1}^{T_1}\prod_{j=1}^{n}\dfrac{\mathrm{e}^{-\mathrm{i}a_jt}-\mathrm{e}^{-\mathrm{i}b_jt}}{2\pi\mathrm{i}t_j}\phi(t_1,\ldots,t_n)\mathrm{d}t_1\ldots\mathrm{d}t_n$

* 由此即可说明特征函数可分离变量时随机变量独立

例 5.6 求 $\cos t$ 对应的分布函数。
可构造出 $P(X=1)=P(X=-1)=\dfrac{1}{2}$，由此分布函数即为结果。

* 对随机变量 X_n，分布函数 F_n，特征函数 ϕ_n，F_n,ϕ_n 收敛性关系？

例 5.7 $X_n = \dfrac{1}{n}$，分布函数的极限为 $\begin{cases} 1 & x > 0 \\ 0 & x \le 0 \end{cases}$，不满足右连续性。

定义 5.4 分布函数的收敛
F, F_n 为分布函数，称 F_n 弱收敛至 F，若对 F 的任意连续点 x，$\lim\limits_{n\to\infty} F_n(x) = F(x)$，记作 $F_n \xrightarrow{W} F$。

定理 5.8 *Lévy-Cramér* 连续性定理
假设 F_n 为分布函数，对应特征函数 ϕ_n
1. 若 $F_n \xrightarrow{W} F$，F 为分布函数，对应特征函数 ϕ，则 ϕ_n 内闭一致收敛到 ϕ。
2. 若 ϕ_n 逐点收敛到 ϕ，ϕ 在 $t = 0$ 处连续，则 $\phi(t)$ 为特征函数，其对应分布函数 F，且 $F_n \xrightarrow{W} F$。

例 5.8 $X \sim U(-n, n)$，则 $\phi_n(t) = \dfrac{\sin nt}{nt}$，特征函数极限为 $\begin{cases} 0 & t \neq 0 \\ 1 & t = 0 \end{cases}$，不满足连续性。

§5.4 极限定理

1. 问题：研究 $T_n = \dfrac{1}{B_n}\sum\limits_{i=1}^{n}(X_i - a_i)$ 的极限性质
*X_k 性质-独立同分布
*$a_k = E[X_k], B_n = c\sqrt{n}$
* 研究不同收敛意义下极限

2. 大数定律 (LLN)、中心极限定理 (CLT)

定义 5.5 弱收敛
若 $F_{X_n} \xrightarrow{W} F_X$，则称 X_n 依分布收敛 *(*弱收敛*)* 到 X，记为 $X_n \xrightarrow{D} X$。

定理 5.9 大数定律
X_i 独立同分布，期望 $\mu = E[X_i]$ 存在，令 $S_n = \sum\limits_{k=1}^{n} X_k$，则 $\dfrac{S_n}{n} \xrightarrow{D} \mu$，即 $\dfrac{S_n}{n} - \mu \xrightarrow{D} 0$。

证明：运用连续性定理，由 $\phi_X'(0) = \mu\mathrm{i}$ 将 X_i 的分布函数在 0 处展开一次项即可计算得结果。
*$\dfrac{S_n}{n} - \mu$ 的无穷小阶数可推测为 $\dfrac{1}{\sqrt{n}}$，这是由于 $\operatorname{Var}(S_n) = n\operatorname{Var}(X_1)$，具体定理为：

定理 5.10 中心极限定理
X_i 独立同分布，期望 $\mu = E[X_i]$，方差 $\sigma^2 = \operatorname{Var}(X_i), \sigma > 0$ 存在，则 $\dfrac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{D} N(0,1)$。

证明：通过平移放缩可不妨设 $\mu = 0, \sigma = 1$，再对 $\dfrac{S_n}{\sqrt{n}}$ 展开估算。

例 5.9 各次测量值独立同分布，方差为 4，欲以 95% 把握保证测量精度达 ± 0.5，求最低测量次数。
$Z_n = \dfrac{S_n - n\mu}{2\sqrt{n}} \xrightarrow{D} N(0,1)$，$P\left(\dfrac{S_n}{n} - \mu\right) \le 0.5 = P\left(|Z_n| \le \dfrac{\sqrt{n}}{4}\right)$，将 Z_n 视为正态分布，查表得至少需要 $n = 62$。

3. Lindeberg 条件 (处理独立但**未必同分布**)
设 $a_k = E[X_k], b_k^2 = \operatorname{Var}(X_k), B_n^2 = \sum\limits_{k=1}^{n} b_k^2$，$F_k$ 为 X_k 分布函数。

L 条件：$\forall \varepsilon > 0, \lim\limits_{n\to\infty} \dfrac{1}{B_n^2}\sum\limits_{k=1}^{n}\int_{|x-a_k|>\varepsilon B_n}(x - a_k)^2 \mathrm{d}F_k = 0$

定理 5.11 *Lindeberg - Feller CLT*
X_i 相互独立，满足 L 条件，则 $\dfrac{\sum_{k=1}^n (X_k - a_k)}{B_n} \xrightarrow{D} N(0,1)$，且 $\dfrac{\max_k b_k^2}{B_n^2} \longrightarrow 0$。
特别地，X_i 相互独立，$a_i = 0$，$E[|X_i|^3] < \infty$，且 $\dfrac{1}{B_n^3}\displaystyle\sum_{k=1}^n E[|X_k|^3] \longrightarrow 0$，则 $\dfrac{\sum_{k=1}^n X_k}{B_n} \xrightarrow{D} N(0,1)$。

* 由 $\displaystyle\int_{|x|>\varepsilon B_n} x^2 \mathrm{d}F_k \le \frac{1}{\varepsilon B_n}\int_{|x|>\varepsilon B_n} |x|^3 \mathrm{d}F_k$ 即可验证特殊情况。
注：
(1) L 条件的概率意义：
$$\sum_{k=1}^n \int_{|x-a_k|>\varepsilon B_n} \frac{(x-a_k)^2}{B_n^2}\mathrm{d}F_k \ge \varepsilon^2 \sum_{k=1}^n P\left(\frac{|X_k - a_k|}{B_n} > \varepsilon\right) \ge \varepsilon^2 P\left(\bigcup_{k=1}^n \left\{\frac{|X_k - a_k|}{B_n} > \varepsilon\right\}\right)$$
$= \varepsilon^2 P\left(\dfrac{\max_k |X_k - a_k|}{B_n} > \varepsilon\right)$，由此 L 条件可推出 $P\left(\dfrac{\max_k |X_k - a_k|}{B_n} > \varepsilon\right) \longrightarrow 0$。
(2) X_i 相互独立时，L 条件增加 $B_n \longrightarrow \infty, \dfrac{\max_k b_k^2}{B_n^2} \longrightarrow 0$(Feller 条件) 即为 CLT 的**必要条件**。
(3) Lyapunov 条件：$\exists \delta > 0, E[|X_i - a_i|^{2+\delta}] < \infty$，且 $\dfrac{1}{B_n^{2+\delta}}\displaystyle\sum_{k=1}^n E[|X_k - a_k|^{2+\delta}] \longrightarrow 0$
类似特殊情况的推导知 Lyapunov 条件可推出 L 条件，从而推出 CLT。
(4) 利用 $b_k^2 = \displaystyle\int_{\mathbb{R}} (x-a_k)^2 \mathrm{d}F_k$ 可拆分为两段，放缩知 $\dfrac{b_k^2}{B_n^2}$ 在极限时一致 $\le \varepsilon^2$，从而且 $\dfrac{\max_k b_k^2}{B_n^2} \longrightarrow 0$。

4. 局部极限定理 (LLT)
* **二项分布**的正态逼近
取 X_i 独立同分布 $B(1,p)$，由再生性知 $S_n \sim B(n,p)$，由此可进行估计：

定理 5.12 二项分布的局部极限定理
$p \in (0,1), q = 1-p, x_k = \dfrac{k-np}{\sqrt{npq}}$，对 $|x_k| \le A$，$n \to \infty$ 时一致地有 $\mathrm{C}_n^k p^k q^{n-k} \sim \dfrac{1}{\sqrt{2\pi npq}}\mathrm{e}^{-x_k^2/2}$。

证明：利用 Stirling 公式估算系数。

定理 5.13 积分形式 *LLT*
$S_n \sim B(n,p)$，则 $P\left(a < \dfrac{S_n - np}{\sqrt{npq}} \le b\right) \longrightarrow \dfrac{1}{\sqrt{2\pi}}\displaystyle\int_a^b \mathrm{e}^{-x^2/2}\mathrm{d}x$

证明：取上一定理中 x_k，有：
左 $= \displaystyle\sum_{k:x_k\in(a,b]} \mathrm{C}_n^k p^k q^{n-k} \sim \sum_{k:x_k\in(a,b]} \frac{\mathrm{e}^{-x_k^2/2}}{\sqrt{2\pi npq}} = \sum_{k:x_k\in(a,b]} \frac{\mathrm{e}^{-x_k^2/2}}{\sqrt{2\pi}}(x_{k+1} - x_k) \longrightarrow$ 右，最后一步是由于黎曼和极限为积分。
*n 固定时，x_k 与 k 一一对应，形成等距分划

5. 矩方法
$\gamma_k = \displaystyle\int_{\mathbb{R}} x^k \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}\mathrm{d}x = \begin{cases} 0 & 2 \nmid k \\ (k-1)!! & 2 \mid k \end{cases}$，可看作 $1,2,\ldots,k$ 两两配对的方式数。

定理 5.14 X_i 独立，且期望均为 0，方差均为 1，$\forall m \ge 3, C_m = \sup\limits_k E[|X_k|^m] < \infty$(一致有界高阶矩)，则 $S_n = \displaystyle\sum_{i=1}^n X_i$ 有 $E\left[\left(\dfrac{S_n}{\sqrt{n}}\right)^k\right] \longrightarrow \gamma_k$，进而 $\dfrac{S_n}{\sqrt{n}} \xrightarrow{D} N(0,1)$。

证明：$E\left[\left(\dfrac{S_n}{\sqrt{n}}\right)^k\right]=n^{-k/2}\sum\limits_{i_1,\ldots,i_k}E[X_{i_1}\ldots X_{i_k}]$，其中非零项每个随机变量次数必至少为 2。由于高阶矩一致有界，若选出的项小于 $k/2$ 个，会在极限中趋于 0，由此只有 $k=2m$ 时可两两配对 $i_1,\ldots,i_{2m}$ 得出项，再由每种配对方式对应极限为 1 可算出结果。
* 由矩的极限推出依分布收敛：

定理 5.15 矩收敛定理
条件：
1. $k\in\mathbb{N},\gamma_{k,n}=\int x^k\mathrm{d}F_n$ 存在
2. $\forall k,\lim\limits_{n\to\infty}\gamma_{k,n}=\gamma_k$
3. $\gamma_k=\int x^k\mathrm{d}F$，且满足 Carleman 条件$\sum\limits_{k=1}^{\infty}(\gamma_{2k})^{-1/(2k)}=\infty$
则 $F_n\xrightarrow{W}F$。

六　几种收敛

§6.1　四种收敛方式

定义 6.1 假设 X,X_n 为 $\{\Omega,\mathcal{F},P\}$ 上的随机变量，则：

1. 几乎处处收敛 (以概率 1 收敛)：

   $P(\{\omega\in\Omega:X_n(\omega)-X(\omega)\longrightarrow 0\})=1$，

   记作 $X_n\xrightarrow{a.s.}X$。

2. **r** 阶收敛：

   $r\in\mathbb{N}^*,\forall n,E[|X_n|^r]<\infty$，且 $E[|X_n-X|^r]\longrightarrow 0$，

   记作 $X_n\xrightarrow{r}X$。$r=1$ 时称平均收敛，$r=2$ 时称均方收敛。

3. 依概率收敛：

   $\forall\varepsilon>0,P(|X_n-X|>\epsilon)\longrightarrow 0$，

   记作 $X_n\xrightarrow{P}X$。

4. 依分布收敛：

   F_X 的连续点处 $P(X_n\le x)\longrightarrow P(X\le x)$，

   记作 $X_n\xrightarrow{D}X$。

* 依分布收敛与样本空间选择无关，具有特殊性

定理 6.1 四种收敛的关系
$X_n\xrightarrow{a.s.}X\Rightarrow X_n\xrightarrow{P}X\Rightarrow X_n\xrightarrow{D}X$
$r>s,X_n\xrightarrow{r}X\Rightarrow X_n\xrightarrow{s}X\Rightarrow X_n\xrightarrow{P}X$
且反方向均无法推出，由此强弱有严格性。

定理证明拆分为以下：

1. $X_n\xrightarrow{P}X\Rightarrow X_n\xrightarrow{D}X$

   $F_n(x)=P(X_n\le x,X\le x+\varepsilon)+P(X_n\le x,X>x+\varepsilon)\le F(x+\varepsilon)+P(|X-X_n|>\varepsilon)$，同理 $F(x-\varepsilon)\le F_n(x)+P(|X-X_n|>\varepsilon)$，连续点处考虑 $F_n(x)$ 上下极限可知结果。

例 6.1 $X_n \xrightarrow{D} X \nRightarrow X_n \xrightarrow{P} X$

$P(X=0)=P(X=1)=\dfrac{1}{2}, Y=1-X, X_n=X$，则 $X_n \xrightarrow{D} Y$，但其他三种收敛均不成立。

2. $r>s, X_n \xrightarrow{r} X \Rightarrow X_n \xrightarrow{s} X$

   利用问题 14.4.28(可由 **Hölder 不等式**证明) 有 $E[|X_n-X|^s]^{1/s} \le E[|X_n-X|^r]^{1/r}$，由此得结论。

3. $X_n \xrightarrow{1} X \Rightarrow X_n \xrightarrow{P} X$

   考虑概率测度在 $|X| \ge a$ 部分的积分得 **Markov 不等式**：$a>0, P(|X|>a) \le \dfrac{E[|X|]}{a}$，由此取 $a=\varepsilon$ 知结论。

   ***Chebyshev 不等式**：$a>0, P(|X-E[X]|>a) \le \dfrac{\mathrm{Var}(X)}{a^2}$

   例 6.2 $X_n \xrightarrow{P} X \nRightarrow X_n \xrightarrow{r} X$

   $\Omega=(0,1]$，P 为 *Lebesgue* 测度，$X_n(\omega)=\begin{cases} n^{1/r} & \omega \in \left(0, \dfrac{1}{n}\right] \\ 0 & \omega \in \left(\dfrac{1}{n}, 1\right] \end{cases}$，$X=0$，计算可验证依概率收敛但不 r 阶收敛。

4. $X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{P} X$

   分析得 $\{\omega \in \Omega : X_n(\omega)-X(\omega) \longrightarrow 0\} = \bigcap_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} \{\omega \in \Omega : |X_n(\omega)-X(\omega)| \le \dfrac{1}{k}\}$，由此知 $X_n \xrightarrow{a.s.} X \Leftrightarrow P\left(\bigcup_{k=1}^{\infty} \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \{\omega \in \Omega : |X_n(\omega)-X(\omega)| > \dfrac{1}{k}\}\right) = 0$，类似分析得其等价于 $\forall \varepsilon>0, \lim_{m\to\infty} P\left(\bigcup_{n=m}^{\infty} \{\omega \in \Omega : |X_n(\omega)-X(\omega)| > \varepsilon\}\right) = 0$，由此得结论。

   例 6.3 $X_n \xrightarrow{r} X \nRightarrow X_n \xrightarrow{a.s.} X$，由此 $X_n \xrightarrow{P} X \nRightarrow X_n \xrightarrow{a.s.} X$

   X_i 相互独立, $X_n \sim B(1, \dfrac{1}{n})$, 可验证任意 r 阶均收敛, 但 $P\left(\bigcup_{n=m}^{\infty} \{\omega \in \Omega : |X_n(\omega)-X(\omega)| > \varepsilon\}\right) = 1$，故不几乎处处收敛。

定理 6.2 反向的成立条件

1. 若 $X_n \xrightarrow{D} c \in \mathbb{R}$，则 $X_n \xrightarrow{P} c$。

2. 若 $\exists k, P(|X_n| \le k)=1, X_n \xrightarrow{P} X$，则 $X_n \xrightarrow{r} X$。

3. 若 $\forall \varepsilon>0, \sum_{n=1}^{\infty} P(|X_n-X|>\varepsilon) < \infty$，则 $X_n \xrightarrow{a.s.} X$。

证明：

1. 利用 $P(|X_n-c|>\varepsilon) = P(X_n>c+\varepsilon)+P(X_n<c-\varepsilon)$，利用分布函数估计知结果。
2. 先拆分 X 为 $|X-X_n|+|X_n|$ 估计出 X 有同样的界，再将差的期望拆分为 $|X_n-X| \le \varepsilon$ 与 $|X_n-X|>\varepsilon$ 的部分可知期望的极限。
3. 利用并的概率小于等于概率的和直接估算。

定理 6.3 弱大数律

X_i 独立同分布，期望 $\mu=E[X_i]$ 存在，令 $S_n=\sum_{k=1}^{n} X_k$，则 $\dfrac{S_n}{n} \xrightarrow{P} \mu$。

证明：由大数定律与收敛于常随机变量得出。

定理 6.4 Skorokhodem 表示定理
设 $X_n \xrightarrow{D} X$，则存在 $(\Omega, \mathcal{F}, P)$，其上的 Y_n, Y 满足 Y_n 与 X_n，Y 与 X 同分布，$Y_n \xrightarrow{a.s.} Y$。

定理 6.5 弱收敛性质
$X_n \xrightarrow{D} X \Leftrightarrow \forall g \in C_b(\mathbb{R})$(有界连续函数)，$E[g(X_n)] \longrightarrow E[g(X)]$。

证明：左推右由表示定理将 X_n, X 替换为 Y_n, Y 利用控制收敛定理可说明 $E[g(Y_n)] \longrightarrow E[g(Y)]$；右推左利用 $P(X_n \leq x) = E[I_{(-\infty,x]}(X_n)]$，以有界连续函数逼近知结论。

§6.2　重要结论

1. **不等式**
* 记 $||X||_p = E[|X|^p]^{1/p}, p \geq 1$
Hölder 不等式：$\dfrac{1}{p} + \dfrac{1}{q} = 1, E[|XY|] \leq ||X||_p ||Y||_q$
Minkowski 不等式：$||X+Y||_p \leq ||X||_p + ||Y||_p$
Markov 不等式：$a > 0, aP(|X| > a) \leq E[|X|]$
Chebyshev 不等式：$a > 0, a^2 P(|X - E[X]| > a) \leq \mathrm{Var}(X)$

例 6.4 $\exists r > 0, E[|X|^r] = 0 \Rightarrow P(X = 0) = 1$
由 *Markov* 不等式，$\forall \varepsilon, P(|X| \geq \varepsilon) \leq \dfrac{E[|X|^r]}{\varepsilon^r} = 0$，由此 $P(|X| < \varepsilon) = 1 \Rightarrow P(|X| \leq 2\varepsilon) = 1$，利用右连续性有结论。

2. 收敛

定理 6.6 记 $\square$ 为 $a.s.$ 或 r 或 P，有：
1. $X_n \xrightarrow{\square} X, X_n \xrightarrow{\square} Y \Rightarrow P(X = Y) = 1$
2. $X_n \xrightarrow{\square} X, Y_n \xrightarrow{\square} Y \Rightarrow X_n + Y_n \xrightarrow{\square} X + Y$
3. 前两条对依分布收敛一般不成立，但 *1.* 可以改为 X, Y 同分布

证明：利用拆分与不等式放缩可验证成立，对依分布收敛，取 $P(X = 1) = P(X = -1) = \dfrac{1}{2}, Y_n = X_n = Y = -X$ 即为前两条的反例，利用连续点处相等与单调右连续可知同分布。

3. **Borel - Cantelli 引理**
事件列的**上下极限**：
A_n 的上限事件：$\limsup\limits_{n\to\infty} A_n = \bigcap\limits_{n=1}^{\infty} \bigcup\limits_{m=n}^{\infty} A_m$ (A_n 中发生无穷多次的样本点)，
A_n 的下限事件：$\liminf\limits_{n\to\infty} A_n = \bigcup\limits_{n=1}^{\infty} \bigcap\limits_{m=n}^{\infty} A_m$ (A_n 中至多有限多次不发生的样本点)
*A_n 的上限事件记为 $\{A_n\ i.o.\}$。

定理 6.7 *Borel - Cantelli* 引理
1. $\sum\limits_{n=1}^{\infty} P(A_n) < \infty \Rightarrow P(A_n\ i.o.) = 0$
2. $\sum\limits_{n=1}^{\infty} P(A_n) = \infty$，$A_i$ 独立 $\Rightarrow P(A_n\ i.o.) = 1$

证明：利用交、并、独立与概率的连续性放缩。
* 由此可发现，若 A_i 独立，$P(A_n\ i.o.)$ 只能为 0 或 1(**零一律**)。

例 6.5 $X_i \sim \mathrm{Exp}(1)$ 独立同分布，则 $P\left(\limsup\limits_{n\to\infty}\dfrac{X_n}{\ln n}=1\right)=1$。
令 $A_n=\left\{\dfrac{X_n}{\ln n}\geq 1+a\right\}$，则 $P(A_n)=\dfrac{1}{n^{1+a}}$ 且相互独立。
可发现 $a\in(-1,0]$ 时 $P(A_n\ i.o.)=1$(A_n 几乎处处发生无穷多次)，$a>0$ 时 $P(A_n\ i.o.)=0$(A_n 几乎处处只发生有限多次)，由此知结论。

§6.3　强大数律

弱大数律：X_i 独立同分布，期望 $\mu=E[X_i]$ 存在，令 $S_n=\sum\limits_{k=1}^{n}X_k$，则 $\dfrac{S_n}{n}\xrightarrow{P}\mu$。

定理 6.8 强大数律
X_i 独立同分布，期望 $\mu=E[X_i]$ 存在，$E[X_i^2]$ 存在，令 $S_n=\sum\limits_{k=1}^{n}X_k$，则 $\dfrac{S_n}{n}\xrightarrow{2}\mu,\dfrac{S_n}{n}\xrightarrow{a.s.}\mu$。

证明：对均方收敛，可由独立性直接计算 $E\left[\left(\dfrac{S_n}{n}-\mu\right)^2\right]=\dfrac{\mathrm{Var}(X_1)}{n}\longrightarrow 0$。
对几乎处处收敛，先找几乎处处收敛的子列，再证明对非负的 X_k 成立，最后推至一般。

定理 6.9 柯尔莫哥洛夫强大数律
X_i 独立同分布，则 $\dfrac{X_1+\cdots+X_n}{n}\xrightarrow{a.s.}\mu\Leftrightarrow E[|X_1|]$ 存在且 $E[X_1]=\mu$。

左推右证明：由于可拆分为两部分，不妨设随机变量均非负。

STEP 1 截尾：取 $Y_n=X_nI_{X_n<n}$，拆分估计可知 $\sum\limits_{n=1}^{\infty}P(X_n\neq Y_n)=\sum\limits_{n=1}^{\infty}P(X_n\geq n)<\infty$，因此 $P(X_n\neq Y_n\ i.o.)=0$，因此 $\dfrac{\sum_{k=1}^{n}(X_k-Y_k)}{n}\xrightarrow{a.s.}0$，由此只需证结论对各阶矩存在的 Y_n 成立。

STEP 2 几乎处处收敛子列：对 $\alpha>1$，令 $\beta_k=\lceil\alpha^k\rceil$，则 $\dfrac{\beta_{k+1}}{\beta_k}\longrightarrow\alpha$，且存在 A 使 $\forall m,\sum\limits_{k=m}^{\infty}\dfrac{1}{\beta_k^2}\leq\dfrac{A}{\beta_m^2}$。记 $S'_n=\sum\limits_{i=1}^{n}Y_i$，利用二阶矩估计知 $\sum\limits_{n=1}^{\infty}P\left(\left|\dfrac{S'_{\beta_n}-E[S'_{\beta_n}]}{\beta_n}\right|>\varepsilon\right)<\infty$，由此 $\left|\dfrac{S'_{\beta_n}-E[S'_{\beta_n}]}{\beta_n}\right|>\varepsilon\xrightarrow{a.s.}0$，进而 $\dfrac{S'_{\beta_n}}{\beta_n}\xrightarrow{a.s.}\mu$。

STEP 3 收敛：由于 S'_n 单调增，$\beta_m\leq n<\beta_{m+1}$ 有 $\dfrac{\beta_m}{n}\cdot\dfrac{S'_{\beta_m}}{\beta_m}<\dfrac{S'_n}{n}<\dfrac{\beta_{m+1}}{n}\cdot\dfrac{S'_{\beta_{m+1}}}{\beta_{m+1}}$，取上下极限可估计得结论成立

* 推论 (**Borel 强大数律**)：试验中事件 A 发生概率 p，S_n 为 n 次独立重复试验中 A 发生次数，则 $\dfrac{S_n}{n}\xrightarrow{a.s.}p$。

定理 6.10 重对数律
X_i 独立同分布，期望 0，方差 1，则：
$P\left(\limsup\limits_{n\to\infty}\dfrac{S_n}{\sqrt{2n\ln\ln n}}=1\right)=1$
$P\left(\liminf\limits_{n\to\infty}\dfrac{S_n}{\sqrt{2n\ln\ln n}}=-1\right)=1$

* 意味着 CLK 不能加强到更强的收敛

七　概率论外篇

§7.1　信息熵

记事件 E 发生概率 $p=P(E)$，定义“惊奇程度”$S(p)$，基本要求：$S(1)=0$、$S(p)$ 严格单调减、关于 p 连续、$S(pq)=S(p)+S(q)$ (直观理解：独立事件引起惊奇程度为分别发生之和)。这些要求可确定：

定理 7.1 $S(p) = -c\ln p, c > 0$。

证明：先考虑 $p_0^{m/n}$，再由连续推到一切 p。

定义 7.1 *Shannon* 熵
离散型随机变量 X，取不同点概率为 $p_1, \dots, p_n, \dots$，定义 $H(X) = -\sum_k p_k \ln p_k$。
联合熵：$H(x,y) = -\sum_{i,j} P(x_i, y_j) \ln P(x_i, y_i)$。
相对熵：$H_Y(X) = \sum_j H_{Y=y_j}(X) P_Y(y_j)$，其中 $H_{Y=y_j}(X) = -\sum_i P(x_i|y_j) \ln P(x_i|y_j)$。

注：$H_Y(X) = -\sum_{i,j} P(x_i, y_j) \ln \frac{P(x_i, y_j)}{P_Y(y_j)}$

定理 7.2 $H(X,Y) = H(Y) + H_Y(X)$

证明：直接计算知结论。

定理 7.3 $H_Y(X) - H(X) \le 0$

证明：由 $\ln x \le x - 1$ 估计知结论。
* 由**凸性**可知 $P(X = x_i) = p_i, i = 1, 2, \dots, n$，$p_1 = \cdots = p_n = \frac{1}{n}$ 时 $H(X) = \ln n$ 最大，即熵可代表**不确定程度**的大小。

定义 7.2 连续型随机变量的熵 X 连续，密度函数 $f(x)$，则 $H(X) = -\int_{\mathbb{R}} f(x) \ln f(x) \mathrm{d}x$。
联合熵：$H(X,Y) = -\int_{\mathbb{R}^2} f(x,y) \ln f(x,y) \mathrm{d}x\mathrm{d}y$。

***Gibbs 不等式** (利用分析知识可证明)：$u - u\ln u \le v - u\ln v$，积分得 $\int_{\mathbb{R}} f(1-\ln f)\mathrm{d}x \le \int_{\mathbb{R}} g(1-\ln g)\mathrm{d}x$，再利用密度函数在实轴积分为 1 知 $\int_{\mathbb{R}} -f\ln f \mathrm{d}x \le \int_{\mathbb{R}} -f \ln g \mathrm{d}x$。

定理 7.4 熵最大的条件
令 $D = \{x \in \mathbb{R} : f(x) \neq 0\}$，则：
1. $D = \mathbb{R}, E[X] = 0, \mathrm{Var}(X) = 1$ 时，正态分布 $N(0,1)$ 熵最大，为 $\ln \sqrt{2\pi \mathrm{e}}$。
2. $D = (0, \infty), E[X] = \frac{1}{\lambda}$ 时，指数分布 $\mathrm{Exp}(\lambda)$ 熵最大，为 $\ln \frac{\mathrm{e}}{\lambda}$。
3. $D = [0, a]$ 时，均匀分布 $U(0,a)$ 熵最大，为 $\ln a$。

证明：分别取 g 为三种分布的密度函数，利用 Gibbs 不等式估算即可。

***Boltzmann 熵**：$S = k\ln\Omega$，$k = k_B$ 为玻尔兹曼常数，Ω 为微观状态数 (类似离散型均匀分布时的情况)。

§7.2 Linderberg 替换理论

*L 条件形式 CLK：
设 $a_k = E[X_k], b_k^2 = \mathrm{Var}(X_k), B_n^2 = \sum_{k=1}^n b_k^2$，$F_k$ 为 X_k 分布函数。
$\forall \varepsilon > 0, \lim_{n\to\infty} \frac{1}{B_n^2} \sum_{k=1}^n \int_{|x-a_k|>\varepsilon B_n} (x-a_k)^2 \mathrm{d}F_k = 0$，则 $\frac{\sum_{k=1}^n (X_k - a_k)}{B_n} \xrightarrow{D} N(0,1)$，且 $\frac{\max_k b_k^2}{B_n^2} \longrightarrow 0$。

定理 7.5 $X_n \xrightarrow{D} X$ 等价于下列条件之一：
1. $\forall g \in C_b(\mathbb{R})$*(有界连续函数)*，$E[g(X_n)] \longrightarrow E[g(X)]$。
2. *任意有界一致连续* g，$E[g(X_n)] \longrightarrow E[g(X)]$。
3. *给定* $m \in \mathbb{N}$，$\forall g \in C_b(\mathbb{R}), g', g'', \ldots, g^{(m)} \in C_b(\mathbb{R})$，$E[g(X_n)] \longrightarrow E[g(X)]$。
4. $\varphi_{X_n}(t) \longrightarrow \varphi_X(t)$*(逐点收敛)*。

Linderberg 思想：取至三阶导均有界连续的 g，独立随机变量列 $Y_n \sim N(0, b_n^2)$，与 X_n 亦独立。
定义 $\zeta_{nk} = \sum\limits_{1\le i<k} X_k + \sum\limits_{k<i\le n} Y_i$，则 $\zeta_{nn} + X_n = S_n, \zeta_{n1} + Y_1 = B_n \cdot N(0,1)$，$\zeta_{nk} + X_k = \zeta_{n,k+1} + Y_{k+1}$，利用逐项相消，可将 X_i **替换**为 Y_i：
$E\left[g\left(\dfrac{S_n}{B_n}\right)\right] - E[g(Y)] = \sum\limits_{k=1}^{n}\left(E\left[g\left(\dfrac{\zeta_{nk}+X_k}{B_n}\right)\right] - E\left[g\left(\dfrac{\zeta_{nk}+Y_k}{B_n}\right)\right]\right)$
再利用 ζ_{nk}, X_k, Y_k 独立泰勒展开估算 $h(t) = \sup\limits_x\{g(x+t) - g(x) - g'(x)t - \dfrac{1}{2}g''(x)t\}$，拆分证明。

§7.3 随机矩阵

1. 起源
统计 (样本协方差阵)：
$X_k = \begin{pmatrix} X_{1k} & \cdots & X_{pk}\end{pmatrix}^T, X = \begin{pmatrix} X_1 & \cdots & X_n\end{pmatrix}$ 为 $p \times n$ 矩阵。
当 X_{ij} 独立同 $N(0,1)$ 时，X 的联合密度 $f(X) = \dfrac{1}{\sqrt{2\pi}^{pn}} \exp\left(-\dfrac{1}{2}\operatorname{tr}(XX^T)\right)$
物理：波函数、随机矩阵模拟

2. 高斯正交系综 (Gaussian Orthogonal Ensemble)
$X_n : \Omega \longrightarrow M_{n\times n}(\mathbb{R}), X_n(\omega) = (x_{ij}(\omega))_{i,j=1}^n$
x_{ij} 独立同 $N(0,\sigma^2)$，记 $A_n = \dfrac{X_n + X_n^T}{2}$，计算可发现 $a_{ii} \sim N(0,\sigma^2), a_{ij} \sim N\left(0, \dfrac{\sigma^2}{2}\right)(i \ne j)$，且 A_n 的上半三角部分独立，由此直接计算乘积可知 A_n(上半三角) 的分布：
$f(A_n) = 2^{-n/2}(\pi\sigma^2)^{-n(n+1)/4}\exp\left(-\dfrac{1}{2\sigma^2}\operatorname{tr} A_n^2\right)$，记为 $A_n \sim \mathrm{GOE}_n(\sigma)$
* 正交变换下不变性：Q 为正交阵，$B_n = Q^T A_n Q$，则 $B_n \sim \mathrm{GOE}_n(\sigma)$

3. 半圆律
X 分布函数 $\omega(x) = \dfrac{1}{2\pi}\sqrt{4-x^2}$，记 $\gamma_k = E[X^k]$，计算知其为 $\begin{cases} 0 & k = 2m+1 \\ \dfrac{1}{m+1}\mathrm{C}_{2m}^m & k = 2m \end{cases}$。
实 Wigner 矩阵：$A_n = (a_{ij})$ 为实对称阵，a_{ii} 独立与 Y 同分布，$a_{ij}(i > j)$ 独立与 z 同分布，$E[Y] = E[Z] = 1, \operatorname{Var}(Z) = 1, \operatorname{Var}(Y) < \infty$，$|Y|, |Z|$ 各高阶矩存在。

定理 7.6 $k \in \mathbb{N}, \dfrac{1}{n}E\left[\operatorname{tr}\left(\dfrac{A_n}{\sqrt{n}}\right)\right] \longrightarrow \gamma_k$

证明：左 $= n^{-1-k/2}\sum\limits_{i_1,\ldots,i_k} E[a_{i_1i_2}a_{i_2i_3}\ldots a_{i_ki_1}]$，类似定理 5.14 可证明不消失的项中必然每个不同的 a_{ij} 出现两次，进而说明 $i_1, \ldots, i_k$ 选取方法与 $1, 2, \ldots, k$ **不相交** (对任何 $a < b < c < d$，不存在配对 $(a,c),(b,d)$) 的两两配对数 ($k = 2m$ 时即为**卡特兰数** C_m) 一一对应，再利用组合计算知结论。

4. Wishart 矩阵模型
$X = (x_{ij})_{p\times n}$，矩阵元独立同 $N(0,1)$，设 $n - p = \alpha$ 固定，则 $\dfrac{1}{p}E\left[\operatorname{tr}\left(\dfrac{1}{p}XX^T\right)^m\right] \longrightarrow C_m$。