01 随机事件与概率基础

纸面目录

页	内容
1	随机试验、事件、概率公理
2	排列组合、抽样与几何概型
3	条件概率、全概率与贝叶斯
4	独立性、可靠性与共同条件
5	随机变量与常用离散分布
6	CDF、密度、端点与混合型
7	均匀、指数及计数—等待关系
8	正态分布与标准化
9	随机变量函数的分布
10	联合、边际、条件与联合 CDF
11	二维密度、区域积分与二维均匀
12	随机变量独立性与二维正态

范围：第一章至第三章 §3.4。专业术语附英文；本册只讲知识与方法，题目见另册《重点题目与标准解答》。

随机试验与样本空间

随机试验（random experiment）可以在相同条件下重复，但单次结果事先不确定；全部可能结果构成**样本空间（sample space）** S 或 Ω，其中一个结果 ω 称样本点（outcome）。

样本空间可有限、可列无限或不可列；取决于记录哪些结果。选定样本空间后，事件应覆盖所需的观察信息。

事件及其发生

事件（event）是样本空间的可测子集。试验结果 $\omega \in A$ 就是 A 发生。S 必然发生，$\varnothing$ 不可能发生；单元素事件称基本事件。

概率是模型对事件发生可能性的量化；**频率（relative frequency）** n_A/n 是有限次记录。相同条件下独立大量重复时，频率趋于稳定，但不必等于某次有限试验的概率值。

事件运算 Event algebra

关系／文字	数学表达
A 发生必有 B 发生	$A \subseteq B$
两者同时发生	$AB = A \cap B$
至少一个发生	$A \cup B$
A 不发生	A^c 或 $\overline{A}$
A 发生且 B 不发生	$A - B = AB^c$

互斥（mutually exclusive）：$AB = \varnothing$；互逆／互补还要求 $A \cup B = S$。并交满足交换、结合、分配律。德摩根律：

$$(A \cup B)^c = A^c B^c, \qquad (AB)^c = A^c \cup B^c.$$

无限并表示至少一个，无限交表示全部。三个事件：至少两个为 $AB \cup AC \cup BC$；恰两个为 $ABC^c \cup AB^c C \cup A^c BC$；最多一个为 $(AB \cup AC \cup BC)^c$；至少一个不发生为 $(ABC)^c$。

概率公理 Axioms

1. 非负：$P(A) \geq 0$。
2. 规范：$P(S) = 1$。
3. 可列可加：两两互斥的 A_i 满足 $P(\bigcup_i A_i) = \sum_i P(A_i)$。

由此得到 $P(\varnothing) = 0$、$P(A^c) = 1 - P(A)$、$0 \leq P(A) \leq 1$。若 $A \subseteq B$，则 $P(A) \leq P(B)$。

一般差事件：$P(A - B) = P(A) - P(AB)$。

加法与容斥 Inclusion–exclusion

$$P(A \cup B) = P(A) + P(B) - P(AB).$$

$$\begin{aligned} P(A \cup B \cup C) = {} & P(A) + P(B) + P(C) \\ & - P(AB) - P(AC) - P(BC) \\ & + P(ABC). \end{aligned}$$

一般为单项加、两两交减、三重交加，依交集项数交替。不能对有重叠的事件直接加概率。

零概率不等于不可能。非空集合也可能概率为 0；概率 1 也不必等于整个样本空间。

02 计数、抽样与几何概型

古典概型 Classical probability

若样本空间有限且每个结果等可能，则 $P(A)=|A|/|S|$。关键是等可能性；不能只按结果的文字类别数数。

n 个不同对象取 k 个：

条件	结果数
不放回、有序	$A_n^k=n!/(n-k)!$
不放回、无序	$\binom{n}{k}=n!/[k!(n-k)!]$
独立均匀放回、有序	n^k

分子、分母须使用相同的顺序与可区分性。放回的无序组合通常不等可能。

无放回抽样

总体 a 个成功对象、b 个失败对象，无放回取 n 个，恰 k 成功：

$$P(X=k)=\frac{\binom{a}{k}\binom{b}{n-k}}{\binom{a+b}{n}},$$

$$\max(0,n-b)\le k\le \min(a,n).$$

上界受成功数与样本量限制，下界受失败数限制。无放回各位置边际概率相同，但各次通常不独立；观察前次结果后要更新剩余数量。

球盒与生日模型

n 个不同球独立均匀投入 m 个不同盒，总结果 m^n。某盒恰 k 球的概率：

$$\binom{n}{k}(1/m)^k(1-1/m)^{n-k}.$$

n 人生日独立、365 天等可能，$n\le 365$ 时至少两人同生日：

$$1-\frac{365\cdot 364\cdots(365-n+1)}{365^n}.$$

先算“全不相同”，再取补；$n>365$ 时概率 1。同月模型将 365 换成 12。

配对与错排 Matching / Derangement

n 个对象与位置随机一一配对。至少一个配对用容斥；指定 k 个配对的概率是 $(n-k)!/n!$。

$$P(\text{至少一配对})=\sum_{k=1}^{n}\frac{(-1)^{k+1}}{k!}.$$

无人配对概率 $\sum_{k=0}^{n}(-1)^k/k!$，错排数为此概率乘 $n!$。若只限制指定的少数对象，直接用条件概率，未必需要全体错排。

几何概型 Geometric probability

在有限正度量区域 S 内均匀取点：

$$P(A)=\frac{m(A\cap S)}{m(S)},\quad 0<m(S)<\infty.$$

m 可为长度、面积或体积。先与总区域求交；没有均匀性就不能直接取度量比。均匀依赖于随机参数的选择，不能在无限平面直接取面积比。

线段中的单点、平面中的通常曲线等可非空但度量为 0，因而均匀模型中概率为 0。

蒲丰投针 Buffon's needle

线距 a，短针长 $b<a$；中点到最近线距离 x 在 $[0,a/2]$ 均匀，夹角 φ 在 $[0,\pi)$ 均匀且独立。针相交当且仅当

$$x\le (b/2)\sin\varphi.$$

参数空间为矩形，面积 $a\pi/2$；相交区域面积 $\int_0^{\pi}(b/2)\sin\varphi\,d\varphi=b$。故

$$P(\text{相交})=\frac{2b}{a\pi}.$$

该式要求短针；均匀选两端点会改变模型。

随机模拟 Monte Carlo

投 n 次，相交 m 次，以频率近似概率，可得 $\pi\approx 2bn/(am)$（$m>0$）。有限试验仅为近似；模拟要按原模型在参数矩形中均匀生成点。

03 条件概率、全概率与贝叶斯

条件概率 Conditional probability

已知 B 发生，所求概率要在 B 内重新归一化：

$$P(A \mid B) = \frac{P(AB)}{P(B)}, \qquad P(B) > 0.$$

$P(AB)$ 是共同发生概率；$P(A \mid B)$ 是已知条件后的比例；$P(B \mid A)$ 交换了条件，一般不同。

给定同一条件后，补事件、加法、容斥仍成立：

$$P(A^c \mid B) = 1 - P(A \mid B).$$

乘法公式 Multiplication rule

$$P(AB) = P(A)P(B \mid A) = P(B)P(A \mid B).$$

$$P(ABC) = P(A)P(B \mid A)P(C \mid AB).$$

多次依次发生时，后面的条件须保留前面全部历史。若独立才可将条件概率换成边际概率。

概率树沿同一路径乘，不同互斥路径相加；每个分叉处的条件概率和为 1。

来源划分 Partition

H_i 两两互斥且并为 S，称完备事件组／划分。选盒、厂别、天气、转球颜色均可作划分；分类应不重不漏。

全概率（law of total probability）：

$$P(E) = \sum_i P(H_i)P(E \mid H_i).$$

每项是“来源与结果共同发生”的概率；求总结果时按各来源加权。

贝叶斯（Bayes' theorem）：

$$P(H_i \mid E) = \frac{P(H_i)P(E \mid H_i)}{\sum_j P(H_j)P(E \mid H_j)}.$$

观测后反推来源，分母为观测的总概率，须正。忽略先验权重只比较条件概率，通常会得错结论。

条件体系内仍可分类

已知另一个条件 C 后，H_i 仍作来源划分：

$$P(E \mid C) = \sum_i P(H_i \mid C)P(E \mid H_iC).$$

因此“已知第一件优质，再问第二件”有两处变化：来源权重更新；若无放回，来源内部的剩余组成也更新。

若条件交集概率为零，对应项略去；不要保留没有定义的条件概率。

根据所求选择方法

所求	方法
几个事件依次共同发生	链式乘法
不同来源合并后出现某结果	全概率
已看到结果，反推来源	贝叶斯
已知信息后再问后续结果	条件体系内的全概率

条件信息如何进入计算

- **选盒：**来源权重来自选盒机制，等可能选盒不等于从全部球中均匀选球。
- **转球：**按转出的颜色分类，更新接收盒的总数和颜色组成。
- **搜索失败：**观测不只是成功，失败也改变来源权重；尚未搜索的区域不能乘已搜索的失败率。
- **共同环境：**同一天、同厂别、同电压要先固定共同条件，不能先平均成功率再直接相乘。

三门模型的信息条件

主持人知道奖品位置，总开未选的空门，并总提供换门：换门获胜等价于初选错误，因此获胜概率为 $2/3$，坚持为 $1/3$。

这个结论依赖主持人的规则。若条件还指定其具体开哪一门，需要把选择规则写入条件概率；不能将“剩两门”直接当作等可能。

条件题的四步

定义来源与观测 → 写各来源下的观测概率 → 加权求分母 → 计算后验或后续条件概率。后验概率之和应为 1。

04 事件独立性与系统可靠性

独立 Independence

A, B 独立当且仅当

$$P(AB) = P(A)P(B).$$

$P(B) > 0$ 时，等价于 $P(A \mid B) = P(A)$，即知道 B 不改变 A 的概率。

独立与互斥不同：互斥是不能同时发生。若两者均正概率，互斥就不独立。概率为 0 或 1 的事件与任意事件独立；A 与自身独立当且仅当 $P(A) \in \{0, 1\}$。

取补保持独立：A^c, B、A, B^c、A^c, B^c 也独立。

两两与相互独立

两两独立只验证每一对。相互独立（mutual independence）要求每个含至少两个事件的子组都满足

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P(A_{i_j}).$$

三个事件需三条两两乘法式及一条三重乘法式。仅两两独立不保证三重交，单独三重乘法也不够。

相互独立事件中任取若干换成补事件仍独立；对不重叠事件组分别作并交补，所得事件也独立。两两独立不足以保证这些性质。

独立重复试验

各次结果独立，试验条件与成功率相同。单次成功率 p，n 次至少一次成功：

$$1 - (1 - p)^n.$$

要求大于 q，转成 $(1 - p)^n < 1 - q$，取满足原不等式的最小整数。严格“大于”与“不小于”取整可能不同。

系列赛可预先设想打满全部局数，再计算达到获胜场数的概率；提前停止不改变最终胜负。

系统可靠性 Reliability

元件正常事件相互独立，正常率为 p_i：

串联系统须全部正常，可靠性 $\prod_i p_i$；并联系统至少一个正常，可靠性 $1 - \prod_i (1 - p_i)$。

n 个同可靠性 p 的独立元件，至少 r 个正常时工作：

$$\sum_{k=r}^{n} \binom{n}{k} p^k (1 - p)^{n-k}.$$

不同 p_i 时，按成功元件的具体子集求和；不能套普通二项。复杂结构可按关键元件状态分类。

条件独立 Conditional independence

给定 H 后独立是

$$P(AB \mid H) = P(A \mid H)P(B \mid H).$$

它一般既不推出无条件独立，也不由无条件独立推出。

若两次事件给定来源 H_i 后独立，来源权重 w_i，各次条件成功率均 p_i，则

$$P(E_1 E_2) = \sum_i w_i p_i^2,$$

而 $P(E_1)P(E_2) = (\sum_i w_i p_i)^2$，一般不同。

先固定共同来源计算，再平均。共用天气、厂别、电压时通常这样处理；若已观察一次结果，下一步使用来源的后验权重。

无条件独立与条件独立如何选

题目明确“各元件独立”，可按给定元件概率相乘；若只说“给定电压后各元件独立”，则先在每个电压条件下算系统可靠性，再全概率加权。

仅给单个元件的故障率不足以确定系统可靠性，还需元件之间的依赖条件。

概率参数与补事件

同一模型内正常率与故障率互补。建立系统事件时先统一“成功”的定义，避免把元件正常率代入按故障数构建的公式。

05 随机变量与常用离散分布

随机变量 Random variable

随机变量是定义在样本空间上的实值函数 $X : S \to \mathbb{R}$ 。映射规则确定，输入的试验结果随机。大写 X 表示变量，小写 x 表示取值。

$\{X = x\}$、$\{X \in A\}$ 是事件；分布描述变量的取值及对应概率。若全部取值有限或可列无限，称离散型（discrete）。

分布律／概率质量函数 PMF：

$$p_k = P(X = x_k) \geq 0, \qquad \sum_k p_k = 1.$$

$$P(X \in A) = \sum_{x_k \in A} p_k.$$

列分布律须列全取值，再求各取值对应事件的概率，最后检查和为 1。

伯努利 Bernoulli

一次试验只有成功／失败两类，成功记 1，失败记 0 。 $P(X=1)=p, P(X=0)=1-p$ ， 记 $\text{Bernoulli}(p) = B(1,p)$。

二项 Binomial $B(n,p)$

n 次独立且每次成功率相同，X 记录成功次数：

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, \ldots, n.$$

$p^k(1-p)^{n-k}$ 是一种指定成功顺序的概率，$\binom{n}{k}$ 是成功位置的选择数。归一性由二项式定理得到。

n 固定、数成功次数；各次不独立或成功率不同，通常不能用普通二项。

超几何 Hypergeometric

有限总体无放回计数，分布律与范围见第 2 页。它与二项的分界是抽样机制，不能仅凭公式相似判断。

停止与截断

几何和负二项数“达到成功条件需几次”。若另设最大试验数，最后一项往往合并所有尚未结束的路径，须按事件重算，不能照搬无限等待公式。

泊松 Poisson $\text{Pois}(\lambda)$

描述固定时段或空间范围内的事件次数，计数没有预设上限，允许 0。$\lambda > 0$ 是整个观察窗口的平均次数：

$$P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, \ldots.$$

归一性由 $\sum_{k\geq 0} \lambda^k/k! = e^{\lambda}$ 得到。

常用建模条件是稳定发生速率、不相交范围计数独立、小范围内多次发生的概率可忽略。速率 ν、观察时长 t 时，$\lambda = \nu t$；先统一单位。

“至少 r 次”常用 $1 - \sum_{k=0}^{r-1} P(X=k)$。

泊松近似 Poisson approximation

n 大、p 小且 np 适中时，$B(n,p) \approx \text{Pois}(np)$。极限条件为 $n \to \infty, p \to 0, np \to \lambda$；经验阈值不能保证尾部概率精度。

几何 Geometric

各次独立、成功率固定 $0 < p < 1$，G 是首次成功的总试验次数：

$$P(G=k) = (1-p)^{k-1}p, \quad k = 1, 2, \ldots.$$

前 $k-1$ 次必须全失败，第 k 次成功。

若记录的是首次成功前的失败数 $J = G - 1$，则 $P(J=j) = (1-p)^j p$，$j \geq 0$。先确认从 0 还是从 1 计数。

巴斯卡／负二项 Pascal / Negative binomial

K 记录第 r 次成功所需总次数，r 正整数：

$$P(K=k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \quad k \geq r.$$

第 k 次必须成功，前 $k-1$ 次恰有 $r-1$ 次成功。$r=1$ 即几何。

若记录失败数 $J = K - r$，则

$$P(J=j) = \binom{j+r-1}{r-1} p^r (1-p)^j, \quad j \geq 0.$$

06 分布函数、密度与端点处理

分布函数 CDF：统一描述各种分布

$$F_X(x) = P(X \le x), \quad x \in \mathbb{R}.$$

它记录阈值左侧的累计概率，适用于离散、连续、混合型。即使变量范围有限，CDF 仍须定义在整个实轴。

基本性质：单调不减、右连续；$F(-\infty) = 0, F(+\infty) = 1$。允许平台与跳跃，不必严格递增。平台表示区间内部无概率质量。

记左极限 $F(a-) = P(X < a)$，则点概率为

$$P(X = a) = F(a) - F(a-).$$

离散型 $F(x) = \sum_{x_k \le x} p_k$，为右连续阶梯；反过来取跳跃量即可还原 PMF。

区间概率与端点

所求事件	概率
$a < X \le b$	$F(b) - F(a)$
$a < X < b$	$F(b-) - F(a)$
$a \le X \le b$	$F(b) - F(a-)$
$a \le X < b$	$F(b-) - F(a-)$
$X > a$	$1 - F(a)$
$X \ge a$	$1 - F(a-)$

右端排除该点用左极限；左端包含该点时减去左极限。只有相关端点概率为 0，开闭才可互换。

连续型与概率密度 PDF

若存在非负函数 f_X，使得所有 x 满足

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt,$$

则称 X 为连续型（此处指 absolutely continuous，绝对连续），f_X 为概率密度函数（probability density function）。

密度如何给出概率

一个函数为 PDF 的充要条件是

$$f_X(x) \ge 0, \qquad \int_{-\infty}^{+\infty} f_X(x)dx = 1.$$

密度曲线下的面积表示概率：

$$P(X \in A) = \int_A f_X(x)dx.$$

沿课件称 $\{x : f_X(x) > 0\}$ 为支撑（support）。积分区间先与支撑相交；分段密度的 CDF 要保留此前累计的概率。

小区间上 $P(x < X \le x + \Delta x) \approx f_X(x)\Delta x$。密度是单位长度上的概率集中程度，**密度值不是点概率，也不要求小于 1**。

连续型单点概率均为 0，区间端点开闭不影响概率。有限个点改变密度值不改变分布。

密度与 CDF 的互推

f_X 连续的点处，$F_X'(x) = f_X(x)$。分段光滑连续 CDF 可逐段求导；连接处不可导并不妨碍存在密度。

CDF 连续是存在密度的必要条件，却不是充分条件，不能将两者直接等同。

未知参数的密度先用总积分为 1；已知连续型 CDF 则先用连接处连续性，再检查单调性、两端极限与密度非负。

混合型 Mixed distribution

一个变量可同时有离散点质量和连续质量。CDF 的跳跃记录点质量，平滑上升部分记录连续质量。

计算事件概率时两部分相加；仅对 CDF 的光滑部分求导会漏掉跳点。将连续变量截断到固定上限，超过上限的概率会合并成上限处的点质量；取整则可使其变为离散型。

检查分布的顺序

PMF：取值列全、非负、和为 1。PDF：支撑明确、非负、积分为 1。CDF：全实轴、单调、右连续、两端极限为 0／1。

07　均匀、指数与计数—等待关系

均匀 Uniform $U(a,b)$，$a<b$

区间内按长度比例取值，密度在 (a,b) 内为 $1/(b-a)$，其他为 0。

$$F(x)=\begin{cases}0, & x<a,\\ (x-a)/(b-a), & a\le x<b,\\ 1, & x\ge b.\end{cases}$$

所求事件若为不等式，先解不等式，与支撑相交后取长度比。到达时间均匀不保证等待时间也均匀，须先写出二者的函数关系。

指数 Exponential $\mathrm{Exp}(\lambda)$

描述恒定事件速率下的连续等待时间，或恒定故障速率下的寿命。等待时间非负，没有固定上限。

本文用速率参数（rate）$\lambda>0$，单位为时间的倒数。λ 越大，等待通常越短，平均等待为 $1/\lambda$。

$$f_T(t)=\begin{cases}\lambda e^{-\lambda t}, & t>0,\\ 0, & t\le 0.\end{cases}$$

$t\ge 0$ 时 $F_T(t)=1-e^{-\lambda t}$；$t<0$ 时 $F_T(t)=0$。生存函数（survival function）：

$$S(t)=P(T>t)=e^{-\lambda t},\quad t\ge 0.$$

用尺度参数（scale）$\theta=1/\lambda$ 时，密度为 $\theta^{-1}e^{-t/\theta}$。速率与尺度不能混用。

无记忆性 Memorylessness

对 $s,t\ge 0$，

$$P(T>s+t\mid T>s)=\frac{e^{-\lambda(s+t)}}{e^{-\lambda s}}=e^{-\lambda t}.$$

已等 s 仍未发生，继续等待的分布与原来相同。$s+t$ 是总寿命阈值，t 是额外等待。

若寿命为多种指数分布的来源混合，通常不再无记忆；已存活会改变来源权重。此时用混合生存函数的比值 $S(s+t)/S(s)$。

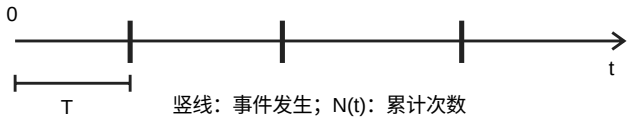
泊松过程 Poisson process

齐次泊松过程以恒定速率 λ 发生事件：不相交时间段计数独立，长度 t 的计数 $N(t)\sim\mathrm{Pois}(\lambda t)$。

泊松变量记录固定时间内发生几次；首次等待 T 记录多久发生。它们由同一事件连接：

$$\{T>t\}=\{N(t)=0\},$$
$$P(T>t)=P(N(t)=0)=e^{-\lambda t}.$$

因此 $T\sim\mathrm{Exp}(\lambda)$。各相邻事件的间隔也相互独立、同服从此指数分布。仅知道一个时段计数服从泊松分布，不足以推出这些过程性质。



几何分布是离散等待

伯努利试验前 n 次成功数 $S_n\sim B(n,p)$。首次成功次数 G 满足

$$\{G>n\}=\{S_n=0\},\quad P(G>n)=(1-p)^n.$$

同样有无记忆性：

$$P(G>m+n\mid G>m)=(1-p)^n.$$

几何按整数次试验等待，指数按连续时间等待。共同依据是“尚未首次发生，等价于此前次数为零”。

从离散分段到连续时间

将泊松过程按 Δt 分段，每段至少有事件的概率 $p=1-e^{-\lambda\Delta t}$，各段独立。首次有事件的段号 G 为几何分布：

$$P(G\Delta t>n\Delta t)=(1-p)^n=e^{-\lambda n\Delta t}.$$

$G\Delta t$ 是将等待时间向上取整到段末的值，误差小于 Δt；当 $\Delta t\to 0$，趋于指数等待时间。

参数与单位

泊松计数参数 λt 是整个窗口的平均次数；指数等待参数 λ 是发生速率。时间从分钟换到小时，速率必须同步换算。

08 正态分布与标准化

正态分布 Normal / Gaussian

$X \sim N(\mu, \sigma^2)$，$\sigma > 0$，密度为

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad x \in \mathbb{R}.$$

μ 是位置参数，σ 是尺度参数；记号第二项是 σ^2，而不是 σ。

曲线关于 μ 对称，峰值 $1/(\sqrt{2\pi}\sigma)$，拐点为 $\mu \pm \sigma$。增大 μ 使曲线平移；增大 σ 使其变宽、变低。两端密度趋于零。

高斯积分 $\int_{\mathbb{R}} e^{-x^2/2}dx = \sqrt{2\pi}$；平方后用极坐标，积分变为 $\int_0^{2\pi}\int_0^{\infty} e^{-r^2/2} r\, dr d\theta = 2\pi$，由此验证归一性。

标准化 Standardization

减去均值将中心移到 0，再除以标准差，将偏离量表示为“几个标准差”：

$$Z = \frac{X-\mu}{\sigma} \sim N(0,1).$$

标准正态密度 $\varphi(z) = e^{-z^2/2}/\sqrt{2\pi}$；CDF 为 $\Phi(z) = P(Z \le z)$。

所有正态概率都可转为同一标准正态表。两个端点同时变换，因 $\sigma > 0$，不等号方向不变：

$$\begin{aligned} P(a < X < b) &= P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right). \end{aligned}$$

标准化除以标准差，不是方差；连续型端点开闭不影响概率。

左尾、右尾与对称区间

$$P(X \le b) = \Phi\left(\frac{b-\mu}{\sigma}\right),$$

$$P(X > b) = 1 - \Phi\left(\frac{b-\mu}{\sigma}\right).$$

利用对称性，$\Phi(-z) = 1 - \Phi(z)$。对 $k > 0$，

$$P(|X-\mu| < k\sigma) = 2\Phi(k) - 1.$$

$k = 1, 2, 3$ 时约为 $0.6827, 0.9545, 0.9973$。

查表先确认给的是左尾 $\Phi(z)$，还是 $P(0 < Z < z)$（$z \ge 0$）；后者须加 0.5 才是左尾。

z	0	0.8	1	1.645
$\Phi(z)$	.5	.7881	.8413	.95
z	1.96	2	2.576	3
$\Phi(z)$	.975	.9772	.995	.99865

由概率反推参数

$\Phi^{-1}(q)$ 是标准正态的 q 分位点。若 $P(X \le a) = q$，则

$$a = \mu + \sigma\Phi^{-1}(q).$$

两个不同分位点可联立求 μ, σ。若要求 $P(|X-\mu| < d) \ge q$，其中 $d > 0, 0 < q < 1$，则

$$\sigma \le \frac{d}{\Phi^{-1}((1+q)/2)}.$$

尺度越小，落在固定中心区间内的概率越大；这也可用来检查不等号方向。

线性变换保持正态

若 $a \neq 0$，则

$$aX + b \sim N(a\mu + b, a^2\sigma^2).$$

均值按一次变换，方差乘 a^2。若 $a = 0$，得到常数即点质量，不再有普通正态密度。

09 随机变量函数的分布

等价事件与 CDF 法

已知 X 分布，求 $Y=g(X)$： 1. 根据原支撑确定输出范围。 2. 将 $\{g(X)\le y\}$ 改写为关于 X 的事件。 3. 求和或积分得到 F_Y，有密度时再求导。

CDF 法适用于单调与非单调变换。支撑外须补齐 CDF 的 0／1 与密度的 0。

离散型按相同输出合并：

$$P(Y=y)=\sum_{x:g(x)=y}P(X=x).$$

单调变换 Monotone transformation

g 在原支撑区间连续可微且导数恒正或恒负，反函数为 $h=g^{-1}$：

$$f_Y(y)=f_X(h(y))\,|h'(y)|.$$

导数绝对值是尺度修正因子（Jacobian factor）；递减时也必须取绝对值。

递增时 $F_Y(y)=F_X(h(y))$；递减且 X 连续时 $F_Y(y)=1-F_X(h(y))$。只需在原支撑上单调，不要求全实轴单调。

线性变换 $Y=aX+b$、$a\neq 0$：

$$f_Y(y)=\frac{1}{|a|}f_X\left(\frac{y-b}{a}\right).$$

多对一变换 Many-to-one

将原支撑拆成单调区间，对各有效反解 $x_i(y)$ 的贡献求和：

$$f_Y(y)=\sum_i f_X(x_i(y))|x_i'(y)|.$$

必须先检查反解是否仍在原支撑内。反解数量改变时，密度通常分段；分段点来自边界或单调区间，不凭经验猜测。

绝对值与平方

X 有密度时，$y>0$ 有

$$F_{|X|}(y)=P(-y\le X\le y)=F_X(y)-F_X(-y),$$
$$f_{|X|}(y)=f_X(y)+f_X(-y).$$

原支撑不对称时，两项不一定同时非零，不能直接写成 $2f_X(y)$。

平方变换在 $z>0$ 时：

$$f_{X^2}(z)=\frac{f_X(\sqrt{z})+f_X(-\sqrt{z})}{2\sqrt{z}}.$$

原变量可取正负，输出却非负，须合并正负分支。标准正态平方的密度为 $e^{-z/2}/\sqrt{2\pi z}$（$z>0$），称自由度 1 的卡方分布（chi-square，χ^2_1）。

指数与对数变换

$X\sim N(0,1)$，$Y=e^X$ 的范围为 $(0,\infty)$：

$$F_Y(y)=\Phi(\ln y),\qquad f_Y(y)=\frac{\varphi(\ln y)}{y}\quad(y>0).$$

这是对数正态分布（lognormal）；范围外密度为 0。

$Z=\ln|X|$ 的范围为全实轴：

$$F_Z(z)=2\Phi(e^z)-1,\qquad f_Z(z)=2e^z\varphi(e^z).$$

$X=0$ 的概率为 0，在该点任意补定义不改变分布。

概率积分变换 Probability integral transform

若 X 的 CDF 连续，则

$$F_X(X)\sim U(0,1).$$

当 CDF 在有效范围严格递增，可由 $P(X\le F_X^{-1}(u))=u$ 证明。即使有平台结论仍成立，证明可用广义分位数。离散型有跳跃时不能直接使用此结论。

10 二维离散分布与联合分布函数

联合、边际与条件

同一样本空间上的 (X,Y) 构成随机向量（random vector）。联合描述共同取值，边际描述单个分量，条件描述给定信息后一个分量的分布。

二维离散型：$p_{ij}=P(X=x_i,Y=y_j)\ge 0$，$\sum_{i,j}p_{ij}=1$。区域概率为相应格子相加。不可能的取值对填 0。

边际（marginal）：

$$P(X=x_i)=p_{i\cdot}=\sum_j p_{ij},$$

$$P(Y=y_j)=p_{\cdot j}=\sum_i p_{ij}.$$

X 为行、Y 为列时，分别取行和、列和。

条件（conditional）：

$$P(X=x_i\mid Y=y_j)=\frac{p_{ij}}{p_{\cdot j}},\quad p_{\cdot j}>0.$$

指定行或列后重新归一化。条件也可为取值区域事件，先与所求事件求交再相除。

构造联合：$p_{ij}=P(X=x_i)P(Y=y_j\mid X=x_i)$，不需要独立。联合决定边际；仅边际通常不能决定联合。

分层与累计计数模型

泊松总数 $N\sim\text{Pois}(\lambda)$，各项独立以概率 p 分类且与计数机制独立，则 $K\mid N=n\sim B(n,p)$：

$$P(N=n,K=k)=e^{-\lambda}\frac{\lambda^n}{n!}\binom{n}{k}p^k(1-p)^{n-k}.$$

$n\ge 0,0\le k\le n$。对 $n\ge k$ 求和，得 $K\sim\text{Pois}(p\lambda)$，称泊松筛选。

泊松过程 $0<t_1<t_2$，前段与后段增量独立：

$$\begin{aligned}&P(N(t_1)=i,N(t_2)=j)\\&=e^{-\lambda t_2}\frac{(\lambda t_1)^i}{i!}\frac{[\lambda(t_2-t_1)]^{j-i}}{(j-i)!},\quad j\ge i\ge 0.\end{aligned}$$

累计数 $N(t_1),N(t_2)$ 本身不独立。

独立伯努利试验成功率 p，首次与第二次成功序号为 U,V，则 $P(U=m,V=n)=p^2(1-p)^{n-2}$，$1\le m<n$；给定 $V=n$，U 在 $1,\dots,n-1$ 均匀。

联合 CDF Joint cumulative distribution function

$$F(x,y)=P(X\le x,Y\le y).$$

它是左下矩形的概率；分别对各坐标单调不减、右连续，取值在 $[0,1]$。任一坐标趋负无穷时为 0；两坐标都趋正无穷时为 1。

还须有矩形增量非负（不能由分别单调替代）：

$$\begin{aligned}&P(x_1<X\le x_2,y_1<Y\le y_2)\\&=F(x_2,y_2)-F(x_1,y_2)\\&\quad-F(x_2,y_1)+F(x_1,y_1)\ge 0.\end{aligned}$$

这是右闭矩形的四角公式。$F(x,y)=1$ 只表示该左下矩形含全部概率；固定一坐标的切片不一定归一到 1。

CDF 与 PMF、边际的转换

离散时 $F(x,y)=\sum_{x_i\le x,y_j\le y}p_{ij}$；反向取双跳跃：

$$\begin{aligned}p_{ij}&=F(x_i,y_j)-F(x_i-,y_j)\\&\quad-F(x_i,y_j-)+F(x_i-,y_j-).\end{aligned}$$

边际 $F_X(x)=F(x,+\infty)$、$F_Y(y)=F(+\infty,y)$，不是将另一坐标取 0。

条件 CDF 与混合质量

$P(Y=y)>0$ 时，$F_{X|Y}(x\mid y)=P(X\le x,Y=y)/P(Y=y)$。固定 y 后是关于 x 的一元 CDF。

若 Y 连续，在适当正则条件下用邻域概率比的极限：

$$\lim_{\varepsilon\downarrow 0}\frac{P(X\le x,y<Y\le y+\varepsilon)}{P(y<Y\le y+\varepsilon)}.$$

不能直接除以 $P(Y=y)=0$。有连续联合密度且 $f_Y(y)>0$ 时，等于 $\int_{-\infty}^{x}f(u,y)/f_Y(y)\,du$。

混合分布的 CDF 须将已覆盖点质量与连续区域积分相加；只求二阶导会漏掉点质量。

11 二维密度、区域积分与条件分布

联合密度 Joint PDF

若非负函数 f 满足

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,dv\,du,$$

则称联合密度。必须有 $\iint_{\mathbb{R}^2} f=1$；适当光滑处 $f=\partial^2 F/(\partial x\partial y)$。

区域概率为

$$P((X,Y)\in A)=\iint_A f(x,y)\,dx\,dy.$$

有联合密度则两个边际均有密度；反过来不一定，两个连续分量可能都集中在同一条曲线上。

单点和通常曲线的二维面积为 0，因此相应概率为 0。边界开闭通常不改变积分。

支撑与积分限

先画支撑区域 D，再与所求事件求交。若 $D=\{a<x<b, g_1(x)<y<g_2(x)\}$，则固定 x 后，内层 y 从 $g_1(x)$ 到 $g_2(x)$。

换积分次序须重新确定截线；上下界切换处拆分积分，不可机械交换 dx, dy。

边际密度 Marginal PDF

消去另一个变量：

$$f_X(x)=\int_{-\infty}^{+\infty} f(x,y)\,dy,\quad f_Y(y)=\int_{-\infty}^{+\infty} f(x,y)\,dx.$$

实际积分只取支撑内的截线；变量范围外边际为 0。求边际时保留目标变量，不要将两变量都积掉。

条件密度 Conditional PDF

$f_Y(y)>0$ 时，

$$f_{X|Y}(x\mid y)=\frac{f(x,y)}{f_Y(y)}.$$

固定 y 后，以 x 为变量；有效范围由该水平截线决定，积分为 1。条件互换时分母与研究变量都变。

从条件密度构造联合

$$f(x,y)=f_X(x)f_{Y|X}(y\mid x)=f_Y(y)f_{X|Y}(x\mid y).$$

无需独立；先将边际范围与条件范围相交，得到联合支撑，再求另一边际。

条件密度不是联合密度；“先均匀选第一变量，再条件均匀选第二变量”通常不产生均匀联合或均匀边际。

二维均匀 Bivariate uniform

在有限正面积区域 D 上均匀：

$$f(x,y)=\frac{1}{\mathrm{Area}(D)}\mathbf{1}_D(x,y).$$

$\mathbf{1}_D$ 为指示函数，区域内为 1，外为 0。于是区域概率等于 $\mathrm{Area}(A\cap D)/\mathrm{Area}(D)$。

边际与相应截线长度成正比，因此一般不均匀。给定一个坐标后，另一个坐标在有效截线内均匀（该截线长度为正时）。

单值条件与范围条件

给定 $Y=y$，看水平截线，使用密度比；给定 $Y>y$，看正面积区域，用交集概率除条件区域概率：

$$P(X\in A\mid Y>y)=\frac{P(X\in A, Y>y)}{P(Y>y)}.$$

单值条件不能直接套正概率事件的相除公式；范围条件不能只截取一条线。

条件分布的适用范围

只在正边际密度处使用密度比。一般条件分布在边际分布几乎处处意义下确定；光滑题目使用连续版本，不能据单个点上任意改变密度值来改变整套条件计算。

区域题的计算顺序

画支撑 → 与事件相交 → 选外层变量 → 看截线求上下界 → 判断是否分段 → 积分。求条件密度时再除相应边际，并检查条件积分为 1。

12　随机变量独立性与二维正态

随机变量独立 Independence

X,Y 独立，当且仅当所有 x,y 都满足

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

等价地，对任意适当集合 A,B，$P(X\in A,Y\in B)=P(X\in A)P(Y\in B)$。

离散型须所有格满足 $p_{ij}=p_{i\cdot}p_{\cdot j}$；有联合密度时须 $f=f_Xf_Y$ 几乎处处成立。

证明独立要验证全部；否定独立只需一个反例格或正概率区域。密度在单个不连续点不同不够；在连续点不同则其邻域也不同，可用于否定独立。

因式分解 Factorization

若联合密度在全平面几乎处处可写成非负单变量函数乘积 $m(x)n(y)$，则归一化后就是 f_Xf_Y，从而独立。

代数表达式与支撑范围都必须可分离。矩形支撑不是充分条件；三角支撑上的正密度即使代数可分，也通常不独立。

联合已知可求边际；边际已知不能默认相乘。条件分布等于边际分布（对适当条件值几乎处处）才对应独立。

二维正态 Bivariate normal

记 $(X,Y)\sim N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$。非退化时 $\sigma_1,\sigma_2>0,|\rho|<1$。令

$$z_1=\frac{x-\mu_1}{\sigma_1},\quad z_2=\frac{y-\mu_2}{\sigma_2},$$

则联合密度为

$$f(x,y)=\frac{\exp\left[-\dfrac{z_1^2-2\rho z_1z_2+z_2^2}{2(1-\rho^2)}\right]}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}.$$

ρ 为相关参数（correlation）；$|\rho|=1$ 时退化到直线，不适用上述普通二维密度。

边际与条件正态

边际仍正态：$X\sim N(\mu_1,\sigma_1^2)$、$Y\sim N(\mu_2,\sigma_2^2)$。边际不含 ρ，不能仅凭边际还原联合。

给定 $X=x$，Y 的条件均值与方差为

$$m_{Y|x}=\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1),$$

$$v_{Y|x}=\sigma_2^2(1-\rho^2).$$

因此 $Y\mid X=x\sim N(m_{Y|x},v_{Y|x})$。交换下标即可得到 $X\mid Y=y$。

条件标准差为 $\sigma_2\sqrt{1-\rho^2}$。求条件概率时用条件均值与条件标准差进行标准化，而不是原边际参数。

密度指数配方可写成边际正态密度乘条件正态密度；条件因子对其变量积分为 1，便得到边际及条件公式。

零相关与独立的适用前提

联合二维正态中，独立当且仅当 $\rho=0$。

充分性：$\rho=0$ 使交叉项消失，密度分成边际乘积。
必要性：独立且密度连续，在中心比较联合与边际密度乘积，得 $\sqrt{1-\rho^2}=1$，故 $\rho=0$。

仅边际分别正态不足以推出联合正态；不能把零相关等价独立套到任意分布。

最后检查

- 公式的模型条件、取值范围、正分母。
- 概率在 $[0,1]$，分布非负且归一。
- 离散端点用左极限；连续端点可换开闭。
- 指数用速率，泊松用窗口平均次数；单位一致。
- 变量变换不漏有效分支与绝对导数。
- 条件独立不替代无条件独立；有共同来源先分类。

资料依据：黄炜第一至三章课件、课堂记录、原始课件及本课程章节笔记；示意图为本册自绘。范围至第三章 §3.4，2026-11-05。