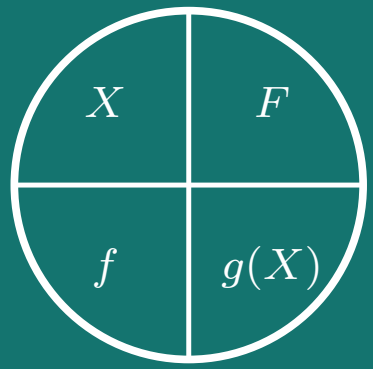


概率论与数理统计

第二章：随机变量及其分布

随机变量　分布律　分布函数　概率密度　函数分布

叶文杰

yewenjie@fjnu.edu.cn

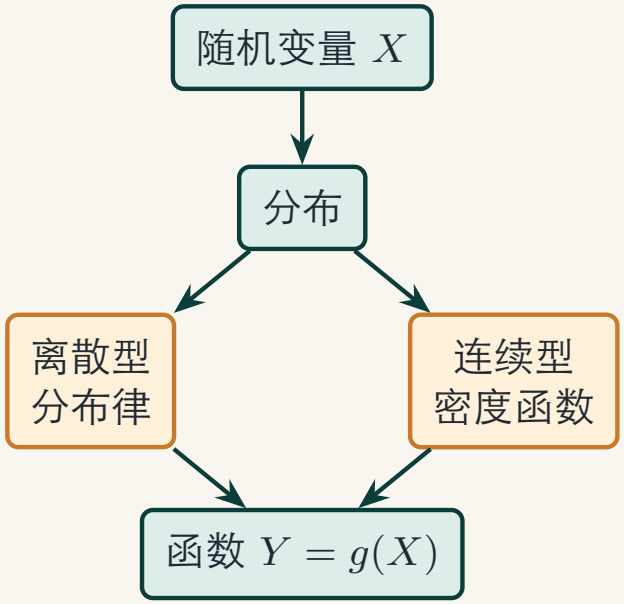
yewenjie.org

2026 年秋季学期

第二章学习内容

主要内容

1. 把试验结果数量化，得到随机变量；
2. 用分布律或分布函数描述取值规律；
3. 对连续型变量，用概率密度计算区间概率；
4. 已知 X 的分布，进一步求 $g(X)$ 的分布。



1

随机变量

第二章　随机变量及其分布

为什么要引入随机变量?

动机

随机试验的结果有时本身就是数量，有时不是数量。引入随机变量，能够把随机结果与实数建立对应关系。

结果本身是数

- 掷骰子的点数;
- 一天进入某楼的人数;
- 昆虫的产卵数;
- 某地某月最高温度。

结果需要数值化

- 球的颜色：红色、白色;
- 硬币的面：正面、反面;
- 产品状态：合格、不合格。

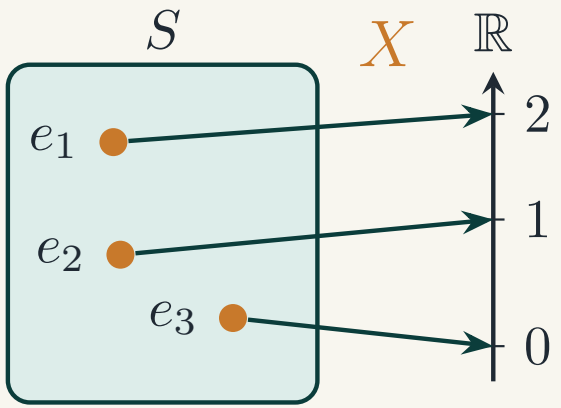
随机变量：把样本点映到实数的 (可测) 映射

定义

设 S 是样本空间。随机变量 X 是定义在样本空间上的实值函数：

$$X : S \longrightarrow \mathbb{R}, \qquad e \longmapsto X(e).$$

每个试验结果 e 经过 X 后，得到一个数 $X(e)$。



随机性在哪里?

函数关系本身确定；随机性来自样本点 e 的随机性。

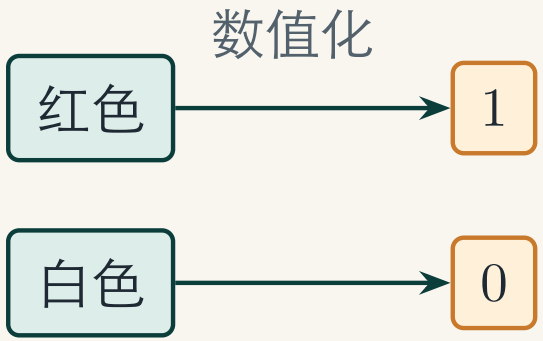
例：把颜色数值化

袋中有红球、白球，任取一球，观察颜色。样本空间为

$$S = \{\text{红色}, \text{白色}\}.$$

可以定义

$$X(e) = \begin{cases} 1, & e = \text{红色}, \\ 0, & e = \text{白色}. \end{cases}$$



这样便把非数量结果变成了数量结果。

注

数值编码不唯一。可以令红色为 1、白色为 0，也可以采用其他编码；关键是要明确对应关系。

随机变量与普通函数有什么区别?

相同点

随机变量是函数：它有定义域，值域和映射关系

不同点

随机变量的定义域是随机试验结果。试验前，我们不知道哪个样本点会出现，因此也不知道 X 会取哪个值。

取值不确定

试验之前，只知道 X 可能取哪些值；不能预先确定它实际取哪一个值。

取值有概率

样本点以一定概率出现，于是 X 取每个值也有相应概率。

记号约定

随机变量

通常用大写字母表示：

$$X, Y, Z, W, N, \ldots$$

例如，X 可以表示一次试验中得到的数值结果。

随机变量的取值

通常用小写字母表示：

$$x, y, z, w, n, \ldots$$

例如，事件 $\{X = x\}$ 表示随机变量 X 取值等于 x。

注意区分：X 是随机变量，x 是一个具体数值。

事件可以通过随机变量表达

随机变量的方便之处

在有了随机变量 X 后，随机试验中的许多事件可以写成关于 X 的关系式。

电话交换台

令 X 表示某电话交换台一分钟内接到的呼叫次数。

$$\{\text{没有收到呼叫}\} = \{X = 0\},$$

$$\{\text{收到不少于一次呼叫}\} = \{X \geq 1\}.$$

随机变量的方便之处

概率论的研究对象从事件发生的概率扩展为研究随机变量的取值规律。

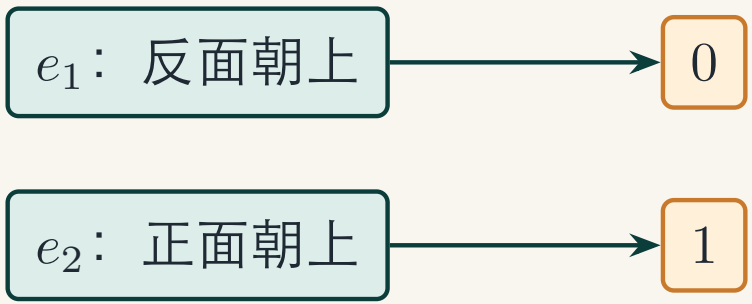
例：抛一枚硬币

抛一枚硬币，设

$$e_1 = \text{反面朝上}, \qquad e_2 = \text{正面朝上}.$$

令 X 表示出现正面的次数，则

$$X(e) = \begin{cases} 0, & e = e_1, \\ 1, & e = e_2. \end{cases}$$



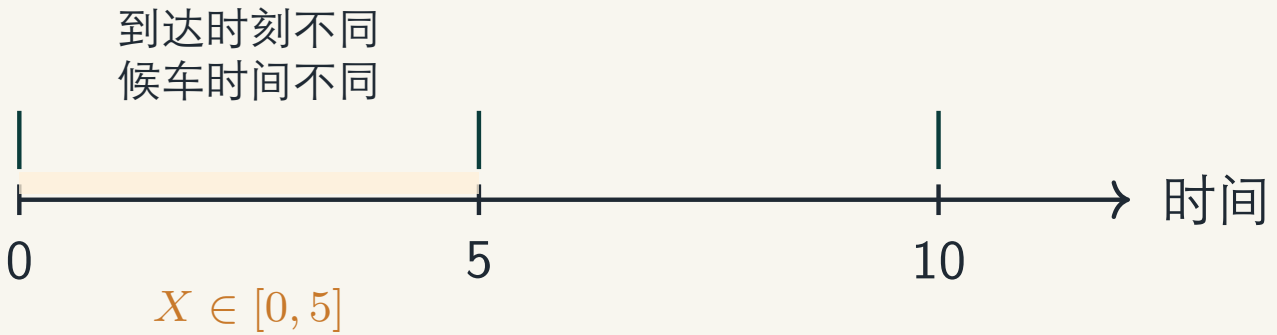
例：候车时间

某公共汽车站每隔 5 分钟有一辆车通过。若某人到达车站的时刻是随机的，令

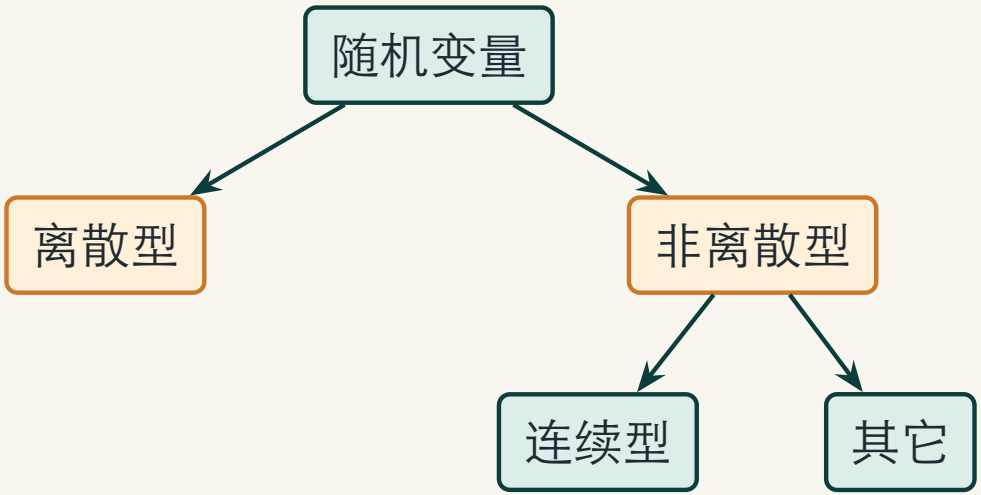
$$X = \text{此人的候车时间}.$$

则 X 是随机变量，并且可能取值范围为

$$0 \le X \le 5.$$



随机变量的分类



本章主要研究离散型随机变量和具有概率密度的连续型随机变量。寿命、长度、候车时间等连续量，常用连续型分布建模。

离散型与连续型的典型例子

离散型随机变量

- 掷骰子的点数：$1, 2, 3, 4, 5, 6$；
- 连续射击直到命中时的射击次数：$1, 2, 3, \ldots$；
- 射击 30 次命中目标的次数：$0, 1, \ldots, 30$。

连续型随机变量

- 灯泡寿命：$[0, +\infty)$；
- 测量误差：通常落在某个区间；
- 候车时间、温度、长度等连续量。

本节小结

随机变量是定义在样本空间上的实值函数；分类的标准是它可能取值的形态。

2

离散型随机变量及其分布律

第二章　随机变量及其分布

离散型随机变量

定义

若随机变量 X 的所有可能取值为有限多个或可列多个，则称 X 为**离散型随机变量**。

例：取球数

袋中有若干红球和白球，从中任取 3 个球。令

$$X = \text{取到的白球数}.$$

若袋中白球数为 2，则

$$X \in \{0, 1, 2\}.$$

一个取球例子：白球数的分布

袋中共有 5 个球，其中 2 个白球、3 个红球。从中任取 3 个球，令 X 为取到的白球数。

$$X = 0, 1, 2.$$

由古典概型，

$$\mathbb{P}(X=0) = \frac{\binom{3}{3}}{\binom{5}{3}} = \frac{1}{10},$$

$$\mathbb{P}(X=1) = \frac{\binom{2}{1}\binom{3}{2}}{\binom{5}{3}} = \frac{6}{10}, \qquad \mathbb{P}(X=2) = \frac{\binom{2}{2}\binom{3}{1}}{\binom{5}{3}} = \frac{3}{10}.$$

x	0	1	2
$\mathbb{P}(X=x)$	1/10	6/10	3/10

离散型随机变量的分布律

定义

设离散型随机变量 X 的所有可能取值为

$$x_1, x_2, \ldots, x_k, \ldots$$

若

$$\mathbb{P}(X = x_k) = p_k, \qquad k = 1, 2, \ldots,$$

则称这些概率给出了 X 的**分布律**。

X 的分布律满足

$$p_k \geq 0, \qquad \sum_k p_k = 1.$$

分布律的两种表示方法

公式法

直接给出

$$\mathbb{P}(X = x_k) = p_k, \qquad k = 1, 2, \ldots.$$

适合取值有规律的情况。

列表法

$$X \sim \begin{pmatrix} x_1 & x_2 & \cdots & x_n & \cdots \\ p_1 & p_2 & \cdots & p_n & \cdots \end{pmatrix}.$$

也可写成表格。

注

无论使用哪种方法，都必须确认所有概率非负，并且总和为 1。

例：由总概率为 1 确定常数

设随机变量 X 的分布律为

$$\mathbb{P}(X=k)=a\frac{\lambda^k}{k!},\qquad k=0,1,2,\ldots,\qquad \lambda>0.$$

求常数 a。

解

由分布律性质，

$$\sum_{k=0}^{\infty}\mathbb{P}(X=k)=1.$$

因此

$$a\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}=a\mathrm{e}^{\lambda}=1,$$

从而

$$a=\mathrm{e}^{-\lambda}.$$

例：两次投篮命中次数

某篮球运动员每次投篮命中概率为 0.9，两次独立投篮。令

$$X = \text{两次投篮中的命中次数.}$$

则 X 的可能取值为 $0, 1, 2$。

$$\mathbb{P}(X = 0) = 0.1^2 = 0.01,$$
$$\mathbb{P}(X = 1) = 2 \cdot 0.9 \cdot 0.1 = 0.18,$$
$$\mathbb{P}(X = 2) = 0.9^2 = 0.81.$$

x	0	1	2
$\mathbb{P}(X = x)$	0.01	0.18	0.81

例：交通信号灯

一辆汽车到目的地需经过 4 组信号灯。每组信号灯以概率 $1/2$ 禁止汽车通过，且各组独立。令 X 表示汽车首次停下时已经通过的信号灯组数。

$$p = \mathbb{P}(\text{禁止通过}) = \frac{1}{2}.$$

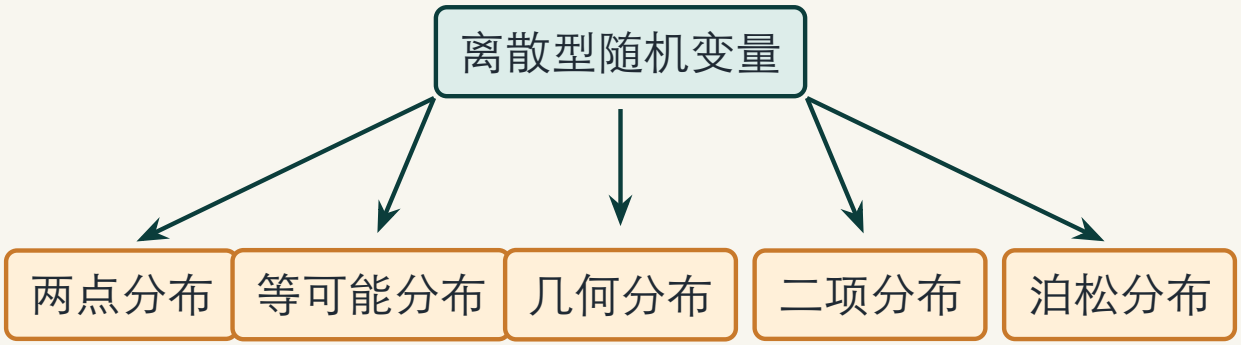
X	0	1	2	3	4
$\mathbb{P}(X=k)$	p	$(1-p)p$	$(1-p)^2p$	$(1-p)^3p$	$(1-p)^4$

代入 $p = 1/2$ 得

$$0.5,\ 0.25,\ 0.125,\ 0.0625,\ 0.0625.$$

最后一项对应四组信号灯都允许通过。

常见离散分布:



要求: 能识别每种分布对应的典型情境。

两点分布

定义

若随机变量 X 只可能取 0 和 1，且

$$\mathbb{P}(X=1)=p, \qquad \mathbb{P}(X=0)=1-p, \qquad 0<p<1,$$

则称 X 服从**两点分布**或 Bernoulli 分布。

X	0	1
p_k	$1-p$	p

适用情况

只有两个可能结果：成功/失败，是/否，合格/不合格，正面/反面。

两点分布的例子

抛硬币

令

$$X=\begin{cases}1, & \text{正面},\\ 0, & \text{反面}.\end{cases}$$

均匀硬币下

$$\mathbb{P}(X=0)=\mathbb{P}(X=1)=\frac{1}{2}.$$

抽检产品

200 件产品中 190 件合格、10 件不合格。若规定

$$X=\begin{cases}1, & \text{取得不合格品},\\ 0, & \text{取得合格品},\end{cases}$$

则

$$\mathbb{P}(X=1)=\frac{10}{200},\quad \mathbb{P}(X=0)=\frac{190}{200}.$$

等可能分布

定义

若随机变量 X 可能取 n 个不同值 $a_1, \ldots, a_n$，且

$$\mathbb{P}(X = a_i) = \frac{1}{n}, \qquad i = 1, \ldots, n,$$

则称 X 服从**等可能分布**。

掷骰子

若 X 为掷一颗均匀骰子的点数，则

$$\mathbb{P}(X = k) = \frac{1}{6}, \qquad k = 1, 2, \ldots, 6.$$

重复独立试验与伯努利试验

重复独立试验

将同一个试验重复进行 n 次，若各次试验的结果互不影响，则称为 n 次重复独立试验。

n 重伯努利试验

若每次试验只有两个互逆结果 A 与 $\overline{A}$，且

$$\mathbb{P}(A) = p, \qquad \mathbb{P}(\overline{A}) = 1 - p,$$

并独立重复 n 次，则称为 n 重伯努利试验。

典型例子：独立抛掷硬币 n 次，每次只观察“是否出现正面”；独立掷骰子 n 次，每次只观察“是否出现 1 点”。

二项分布：发生次数的分布

在 n 重伯努利试验中，令

$$X = \text{事件} A \text{ 发生的次数}.$$

则

$$X = 0, 1, 2, \ldots, n.$$

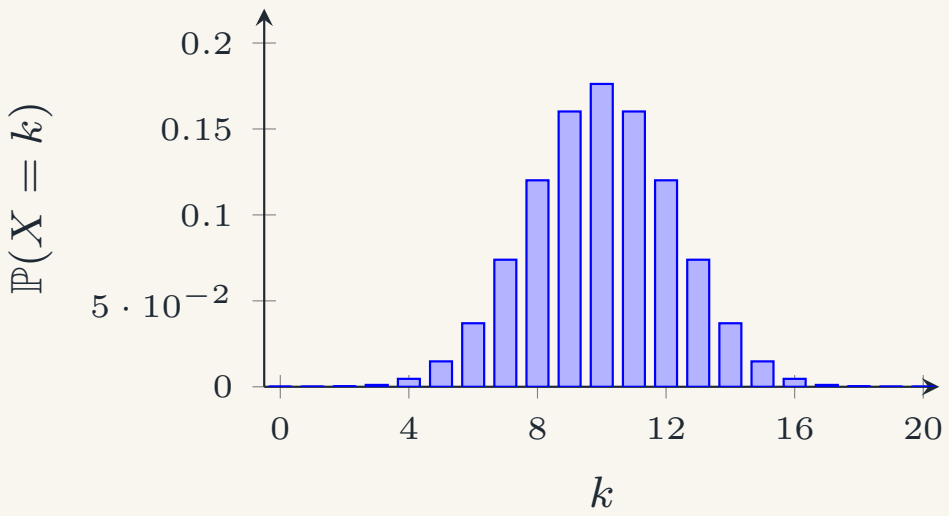
若 $\mathbb{P}(A) = p$，记 $q = 1 - p$。当 $X = k$ 时，A 在 n 次试验中发生 k 次，方式共有 $\binom{n}{k}$ 种；每一种方式的概率为 $p^k q^{n-k}$。

概率公式

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

记为 $X \sim \mathrm{Bin}(n, p)$。

二项分布的图形直觉



例

$X \sim \mathrm{Bin}(20, 0.5)$ 时，概率集中在 10 附近。

影响形状的因素

n 控制试验次数，p 控制成功概率。np 附近常是概率较集中的位置。

例：有放回抽样与二项分布

已知 100 个产品中有 5 个次品。从中有放回地取 3 次，每次任取 1 个。求所取 3 个中恰有 2 个次品的概率。

解

每次取到次品的概率相同且各次独立，故这是伯努利试验。令

$$X = \text{取到的次品数}.$$

则

$$X \sim \mathrm{Bin}(3, 0.05).$$

所以

$$\mathbb{P}(X = 2) = \binom{3}{2}(0.05)^2(0.95) = 0.007125.$$

若改为无放回

各次试验条件改变，不再是伯努利试验，应使用古典概型。

伯努利试验的识别条件

三条要求

1. 每次试验条件相同；
2. 每次试验只考虑两个互逆结果 A 与 $\overline{A}$；
3. 各次试验相互独立。

二项分布描述什么？

二项分布描述的是 n 重伯努利试验中事件 A 出现次数 X 的分布律。

例：灯泡寿命

某类灯泡使用时数在 1000 小时以上的概率是 0.2。求三个灯泡在使用 1000 小时以后最多只有一个坏了的概率。

解

把观察一个灯泡是否在 1000 小时前坏掉看作一次试验。令

$$X = \text{三个灯泡中在 1000 小时已坏的个数.}$$

则每个灯泡“已坏”的概率为 0.8，所以

$$X \sim \mathrm{Bin}(3, 0.8).$$

所求为

$$\mathbb{P}(X \leq 1) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1).$$

因此

$$\mathbb{P}(X \leq 1) = (0.2)^3 + \binom{3}{1}(0.8)(0.2)^2 = 0.104.$$

大批产品抽查：近似二项分布

某型号电子元件中，规定使用寿命超过 1500 小时的为一级品。已知一大批该型号元件的一级品率为 0.2。从中随机抽查 20 只，求其中恰有 k 只一级品的概率，其中 $k = 0, 1, \ldots, 20$。

近似思想

严格地说这是无放回抽样。但若总体很大，抽查数量相对很小，可以近似看作有放回抽样。

令 X 为 20 只元件中一级品的只数，则近似有

$$X \sim \mathrm{Bin}(20, 0.2).$$

于是

$$\mathbb{P}(X = k) = \binom{20}{k} (0.2)^k (0.8)^{20-k}, \qquad k = 0, 1, \ldots, 20.$$

泊松分布

定义

若随机变量 X 的可能取值为 $0, 1, 2, \ldots$，且

$$\mathbb{P}(X = k) = \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!}, \qquad k = 0, 1, 2, \ldots, \qquad \lambda > 0,$$

则称 X 服从参数为 λ 的泊松分布，记为

$$X \sim \mathrm{Poi}(\lambda).$$

典型背景

泊松分布常用于描述：在固定时间或空间范围内，某类稀有事件发生的次数。

泊松分布的应用背景

常见例子

- 电话交换台的呼叫次数；
- 某路段交通事故次数；
- 某时段顾客到达人数；
- 放射性粒子释放次数。

关键词

- 次数型随机变量；
- 固定时间或空间；
- 试验次数很多，单次事件发生的概率很小。

注

当试验次数很多、单次事件发生概率很小时，事件发生的总次数常近似服从泊松分布。

二项分布与泊松分布的关系

历史上，泊松分布可作为二项分布的近似。若

$$X \sim \mathrm{Bin}(n, p),$$

且 n 较大、p 较小、$np = \lambda$ 适中，则

$$\binom{n}{k} p^k (1-p)^{n-k} \approx \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!}.$$

Poisson 近似

$$\mathrm{Bin}(n, p) \approx \mathrm{Poi}(np).$$

何时可以用泊松近似?

若

$$X \sim \mathrm{Bin}(n, p), \qquad \lambda = np,$$

则可用 $\mathrm{Poi}(\lambda)$ 近似 X。一个定量误差控制为

$$|\mathbb{P}(X \in E) - \mathbb{P}(Y \in E)| \le np^2, \qquad Y \sim \mathrm{Poi}(np).$$

因此，np^2 越小，泊松近似通常越精确。

例：汽车站事故次数

某繁忙汽车站每天某段时间内有 1000 辆汽车通过。每辆车在该段时间内出事故的概率为 0.0001。求出事故次数不少于 2 的概率。

二项分布：精确计算

令 X 为出事故的汽车数，则

$$X \sim \mathrm{Bin}(1000, 0.0001), \qquad np = 0.1.$$

于是

$$\mathbb{P}(X \geq 2) = 1 - (0.9999)^{1000} - 1000(0.0001)(0.9999)^{999} \approx 0.0046748.$$

即精确概率约为 0.4675%。

泊松近似

由于 n 很大、p 很小，可取

$$X \approx \mathrm{Poi}(0.1).$$

因此

$$\mathbb{P}(X \geq 2) \approx 1 - \mathrm{e}^{-0.1}(1 + 0.1) \approx 0.0046788.$$

即泊松近似约为 0.4679%。

两者非常接近：$0.4675\% \approx 0.4679\%$。

维修人员配置问题

工厂共有 300 台设备，每台独立工作，发生故障概率为 0.01。一台设备故障由一人处理。问至少配备多少维修人员，才能保证不能及时维修的概率小于 0.01。

建模

令

$$X = \text{同时发生故障的设备台数.}$$

则

$$X \sim \mathrm{Bin}(300, 0.01), \qquad np = 3.$$

若配备 N 个维修人员，需要

$$\mathbb{P}(X > N) < 0.01.$$

泊松近似

$$X \approx \mathrm{Poi}(3), \qquad \mathbb{P}(X \le N) > 0.99.$$

查表可得至少需要 8 个维修人员。

几何分布

定义

若随机变量 X 的分布律为

$$\mathbb{P}(X=k)=q^{k-1}p, \qquad k=1,2,\ldots, \qquad p+q=1,$$

则称 X 服从**几何分布**。

首次成功的概率

有放回抽样检查某批产品，次品率为 p。一直抽到第一次出现次品为止。若 X 为抽到的产品数，则

$$\mathbb{P}(X=k)=(1-p)^{k-1}p.$$

几何分布的本质

设独立重复进行同一种试验，每次成功概率为 p，失败概率为 $q = 1 - p$。令

$$X = \text{首次成功所需要的试验次数.}$$

首次成功

$X = k$ 表示前 $k-1$ 次失败、第 k 次成功：

$$\underbrace{\overline{A}_1 \cdots \overline{A}_{k-1}}_{\text{前 } k-1 \text{ 次失败}} A_k.$$

因此

$$\mathbb{P}(X = k) = q^{k-1}p, \qquad k = 1, 2, \ldots.$$

几何分布描述的是**首次成功所需要的试验次数。**

无记忆性

$X > k$ 表示前 k 次均失败，因此

$$\mathbb{P}(X > k) = q^k.$$

于是

$$\begin{aligned}&\mathbb{P}(X > m + n \mid X > m)\\ =&q^n = \mathbb{P}(X > n).\end{aligned}$$

这说明，在已经连续失败若干次的条件下，未来还需等待多少次才首次成功，其条件分布与从头开始等待相同。

注

几何分布是离散型的**首次成功等待时间模型**，其突出特征是**无记忆性。**

离散型分布小结

分布	分布律	典型情境
两点分布	$\mathbb{P}(X=1)=p$	一次成功/失败
等可能分布	$\mathbb{P}(X=a_i)=1/n$	有限个等可能结果
二项分布	$\binom{n}{k}p^k(1-p)^{n-k}$	n 次独立重复中的成功次数
泊松分布	$\mathrm{e}^{-\lambda}\lambda^k/k!$	固定时间内稀有事件次数
几何分布	$(1-p)^{k-1}p$	首次成功所需试验次数

关系

二项分布是两点分布的推广；当 np^2 较小时，二项分布 $\mathrm{Bin}(n,p)$ 可用参数为 np 的泊松分布近似，且其总变差误差不超过 np^2。

本节作业

第五版课后习题 P55: 1，5，7，11;
第六版课后习题 P63: 1，5，7，11;

3

随机变量的分布函数

第二章　随机变量及其分布

为什么要引入分布函数?

离散型：点概率可以描述分布

知道所有可能取值及其概率 $\mathbb{P}(X = x_k)$，就知道了随机变量 X 的分布。

连续型：仅有点概率还不够

在 $[0, 1]$ 上均匀选取一个数，记为 X。虽然每个指定点的概率都是 0，但区间概率可以不同：

$$\mathbb{P}(X = 0.5) = 0, \qquad \mathbb{P}(0 < X \leq 0.5) = \frac{1}{2}.$$

因此，我们需要能够描述区间概率的工具。

统一的工具：分布函数

对任意随机变量，考察“取值不超过 x”的概率：

$$F(x) = \mathbb{P}(X \leq x).$$

知道分布函数 F，就能计算任意左开右闭区间的概率：

分布函数的定义

定义

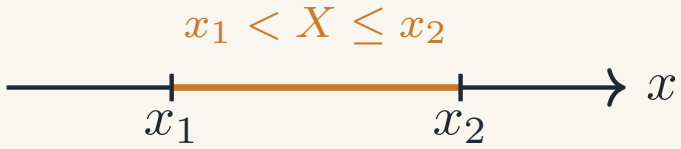
设 X 是随机变量。对任意实数 x，定义

$$F(x) = \mathbb{P}(X \leq x).$$

称 $F(x)$ 为随机变量 X 的**分布函数**。

对于任意 $x_1 < x_2$，有

$$\mathbb{P}(x_1 < X \leq x_2) = F(x_2) - F(x_1).$$



分布函数的两个注意点

区分随机变量 X 与阈值 x

在 $F(x) = \mathbb{P}(X \le x)$ 中，X 是随机变量，x 是用于比较的阈值。x 可以是任意实数，不必是 X 的可能取值。

区分累计概率与点概率

$F(x)$ 表示“X 取值不超过 x”的概率，不要与点概率 $\mathbb{P}(X = x)$ 混淆。固定 X 的分布后，F 是一个确定的函数。

例：掷一颗均匀骰子

令 X 表示所得点数，则

$$F(3) = \mathbb{P}(X \le 3) = \frac{1}{2}, \qquad \mathbb{P}(X = 3) = \frac{1}{6}.$$

虽然 X 不会取到 3.5，但 $F(3.5)$ 仍有意义：

$$F(3.5) = \mathbb{P}(X \le 3.5) = \mathbb{P}(X \le 3) = \frac{1}{2}.$$

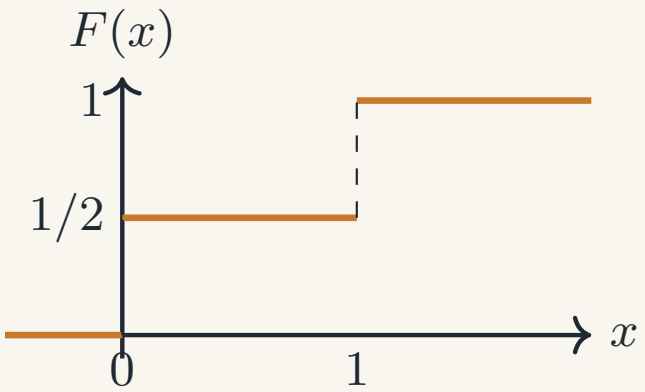
例：抛均匀硬币的分布函数

令

$$X = \begin{cases} 1, & \text{出正面}, \\ 0, & \text{出反面}. \end{cases}$$

且 $\mathbb{P}(X=0) = \mathbb{P}(X=1) = 1/2$。则

$$F(x) = \mathbb{P}(X \le x) = \begin{cases} 0, & x < 0, \\ \dfrac{1}{2}, & 0 \le x < 1, \\ 1, & x \ge 1. \end{cases}$$



离散型随机变量的分布函数

若离散型随机变量 X 的分布律为

$$\mathbb{P}(X = x_k) = p_k, \qquad k = 1, 2, \ldots,$$

则其分布函数为

$$F(x) = \mathbb{P}(X \le x) = \sum_{x_k \le x} \mathbb{P}(X = x_k) = \sum_{x_k \le x} p_k.$$

直觉

$F(x)$ 是所有不超过 x 的取值对应概率之和。

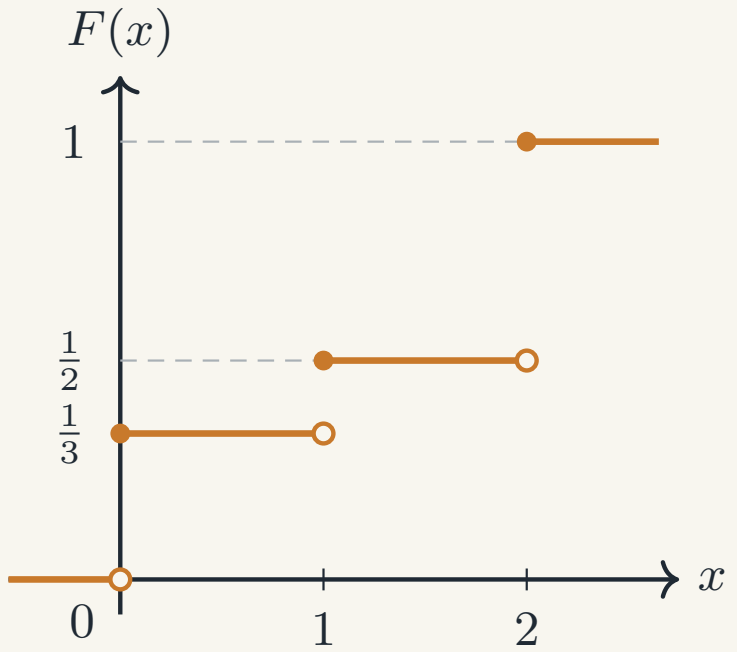
练习：由分布律求分布函数

设 X 的分布律为

X	0	1	2
p_k	1/3	1/6	1/2

则

$$F(x)=\begin{cases}0, & x<0,\\ \frac{1}{3}, & 0\le x<1,\\ \frac{1}{2}, & 1\le x<2,\\ 1, & x\ge 2.\end{cases}$$



检查

在每个取值点，分布函数向上跳跃；跳跃高度正是该点的概率。

分布函数的基本性质

任何随机变量的分布函数 F 都满足：

1. $0 \le F(x) \le 1$；
2. 若 $x_1 < x_2$，则 $F(x_1) \le F(x_2)$；
3. $\lim\limits_{x \to -\infty} F(x) = 0$，$\lim\limits_{x \to +\infty} F(x) = 1$；
4. F 右连续：

$$\lim_{x \to x_0 +} F(x) = F(x_0).$$

重要事实：一个特殊情形

设 F 连续且严格递增，并且

$$\lim_{x \to -\infty} F(x) = 0, \qquad \lim_{x \to +\infty} F(x) = 1.$$

此时 F 存在连续的反函数 $F^{-1} : (0,1) \to \mathbb{R}$。取 $U \sim \mathrm{U}(0,1)$，令 $X = F^{-1}(U)$，则对任意 $x \in \mathbb{R}$，

$$\mathbb{P}(X \le x) = \mathbb{P}(F^{-1}(U) \le x) = \mathbb{P}(U \le F(x)) = F(x).$$

因此，F 就是随机变量 X 的分布函数。

重要公式

设 F 是随机变量 X 的分布函数。

区间概率

对 $a < b$,

$$\mathbb{P}(a < X \leq b) = F(b) - F(a).$$

尾部概率

$$\mathbb{P}(X > a) = 1 - F(a).$$

证明思路

因为

$$\{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\},$$

且两部分互斥。

例：判断是否为分布函数

设

$$F(x)=\begin{cases}\sin x, & 0\le x\le \pi,\\ 0, & \text{其它.}\end{cases}$$

问 F 能否作为某个随机变量的分布函数?

不能

因为 $\sin x$ 在 $[\pi/2,\pi]$ 上下降，不满足分布函数的单调不减性。

另外，

$$\lim_{x\to+\infty} F(x)=0\neq 1,$$

也不满足分布函数的极限性质。

不同随机变量可以有相同分布函数

抛均匀硬币，令

$$X_1 = \begin{cases} 1, & \text{正面}, \\ -1, & \text{反面}, \end{cases} \qquad X_2 = \begin{cases} -1, & \text{正面}, \\ 1, & \text{反面}. \end{cases}$$

X_1 与 X_2 作为样本空间上的对应法则不同，是不同的随机变量。

但它们分布相同

二者都满足

$$\mathbb{P}(X = -1) = \mathbb{P}(X = 1) = \frac{1}{2}.$$

因此分布函数相同：

$$F(x) = \begin{cases} 0, & x < -1, \\ \dfrac{1}{2}, & -1 \le x < 1, \\ 1, & x \ge 1. \end{cases}$$

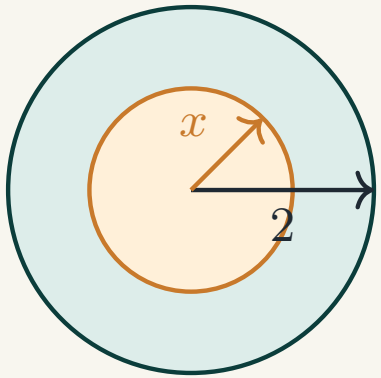
例：射击圆盘的分布函数

一个靶子是半径为 2 米的圆盘。设击中任一同心圆盘的概率与该圆盘面积成正比，并且射击都能中靶。令

$$X = \text{弹着点与圆心的距离}.$$

求 X 的分布函数。
当 $0 \le x \le 2$ 时，

$$\mathbb{P}(X \le x) = \frac{\pi x^2}{\pi 2^2} = \frac{x^2}{4}.$$



故

$$F(x) = \begin{cases} 0, & x < 0, \\ \dfrac{x^2}{4}, & 0 \le x < 2, \\ 1, & x \ge 2. \end{cases}$$

从分布函数到连续型变量

上一例中，X 表示弹着点到圆心的距离。当 $0 \le x \le 2$ 时，$F(x) = x^2/4$。

区间概率可以写成积分

对 $0 \le a < b \le 2$，

$$\mathbb{P}(a < X \le b) = F(b) - F(a) = \frac{b^2 - a^2}{4} = \int_a^b \frac{t}{2}\,\mathrm{d}t.$$

这个积分表达式提示我们：可以通过一个非负函数来计算区间概率。

将被积函数在 $(0, 2)$ 外定义为 0，得到

$$f(t) = \begin{cases} \dfrac{t}{2}, & 0 < t < 2, \\ 0, & \text{其它}, \end{cases} \qquad F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t.$$

本节作业

课后练习

第五版课后习题 P57：18，19。
第六版课后习题 P65：18，19。

复习重点

理解分布函数的定义、四条基本性质，以及如何由分布函数计算区间概率。

4

连续型随机变量及其概率密度

第二章　随机变量及其分布

连续型随机变量：为什么需要密度?

离散型随机变量

可以用分布律给出每个点的概率。

连续型随机变量

可能取值充满一个区间，不能再通过“指定每个点的概率”来给出分布。

新的工具

用**概率密度函数**描述概率在区间上的分布。

概率密度的定义

定义

若随机变量 X 的分布函数 $F(x)$ 可以写成

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t,$$

其中 f 是非负可积函数，则称 X 为连续型随机变量，f 为 X 的**概率密度函数**，简称密度。

必要性质

$$f(x) \geq 0, \qquad \int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1.$$

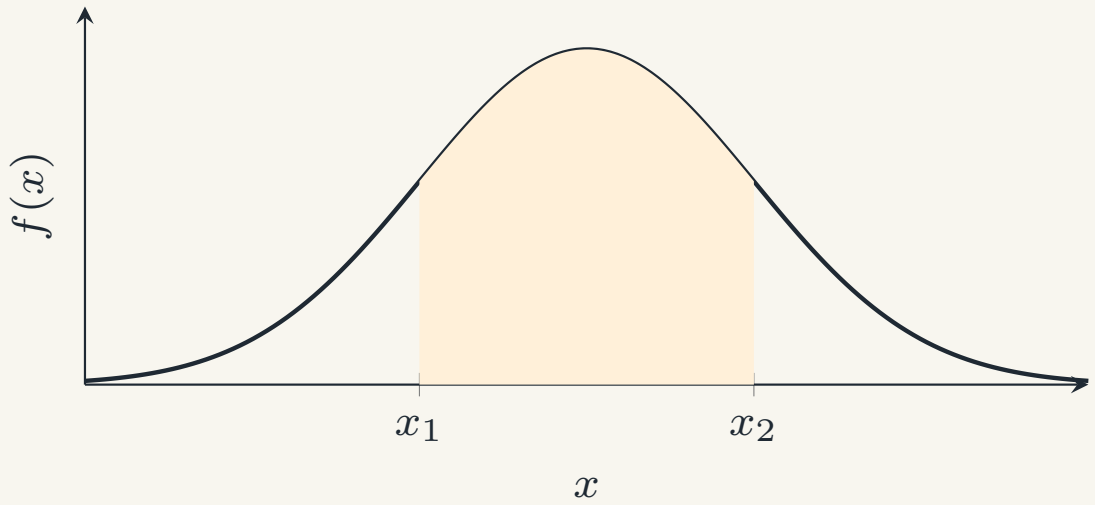
密度与区间概率

若 X 的密度为 f，则

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t.$$

对 $x_1 < x_2$，

$$\mathbb{P}(x_1 < X \le x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)\,\mathrm{d}x.$$



密度函数不是概率

重要提醒

密度值 $f(a)$ 不是概率。不能写成

$$\mathbb{P}(X = a) = f(a).$$

正确理解

$f(a)$ 越大，只能说明 X 落在 a 附近小区间内的概率越大。若 f 在 x 处连续，则

$$\mathbb{P}(x < X \leq x + \Delta x) \approx f(x)\Delta x.$$

与离散型类比

离散型用点概率 $\mathbb{P}(X = x_k)$；连续型用密度曲线下的面积。

连续型随机变量取某一点的概率

若 X 是连续型随机变量，则对任意实数 a，

$$\mathbb{P}(X = a) = 0.$$

因此区间端点是否包含不影响概率：

$$\mathbb{P}(a \le X \le b) = \mathbb{P}(a < X \le b) = \mathbb{P}(a \le X < b) = \mathbb{P}(a < X < b).$$

注意

$\mathbb{P}(X = a) = 0$ 不等于说事件 $\{X = a\}$ 是“不可能事件”。在连续型情形中，单点概率为零是正常现象。

例：由密度确定常数

设 X 的密度为

$$f(x)=\begin{cases} kx, & 0\le x<3,\\ 2-\dfrac{x}{2}, & 3\le x\le 4,\\ 0, & \text{其它}. \end{cases}$$

求 k。

解

由密度总面积为 1，

$$\int_0^3 kx\,\mathrm{d}x+\int_3^4\left(2-\frac{x}{2}\right)\mathrm{d}x=1.$$

计算得

$$\frac{9k}{2}+\frac{1}{4}=1,$$

故

$$k=\frac{1}{6}.$$

例：由密度求分布函数

若上一例中 $k = 1/6$，则

$$f(x) = \begin{cases} \dfrac{x}{6}, & 0 \le x < 3, \\ 2 - \dfrac{x}{2}, & 3 \le x \le 4, \\ 0, & \text{其它}. \end{cases}$$

由

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$$

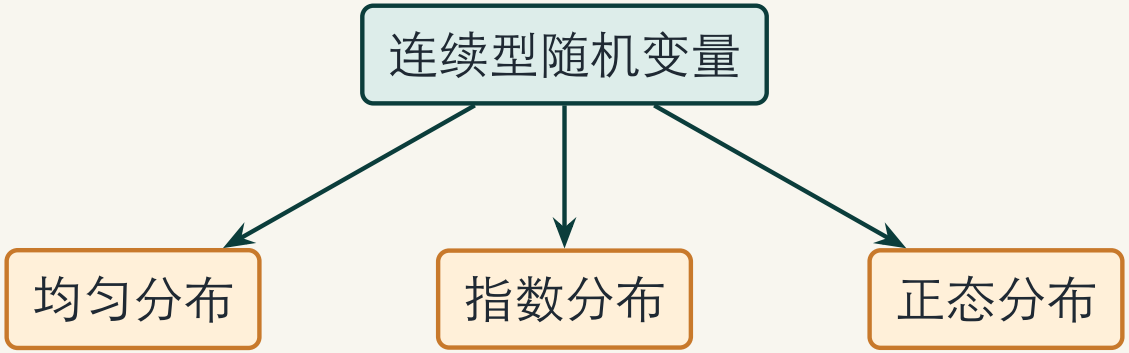
得

$$F(x) = \begin{cases} 0, & x < 0, \\ \dfrac{x^2}{12}, & 0 \le x < 3, \\ -3 + 2x - \dfrac{x^2}{4}, & 3 \le x < 4, \\ 1, & x \ge 4. \end{cases}$$

例如

$$\mathbb{P}(1 < X \le 7/2) = F(7/2) - F(1) = \frac{41}{48}.$$

常见连续型分布：总览



本节重点

掌握每种分布的密度、背景和典型概率计算。

均匀分布

定义

若连续型随机变量 X 的密度为

$$f(x)=\begin{cases}\dfrac{1}{b-a}, & a<x<b,\\ 0, & \text{其它},\end{cases}$$

则称 X 在区间 (a,b) 上服从**均匀分布**，记为

$$X\sim \mathrm{U}(a,b).$$

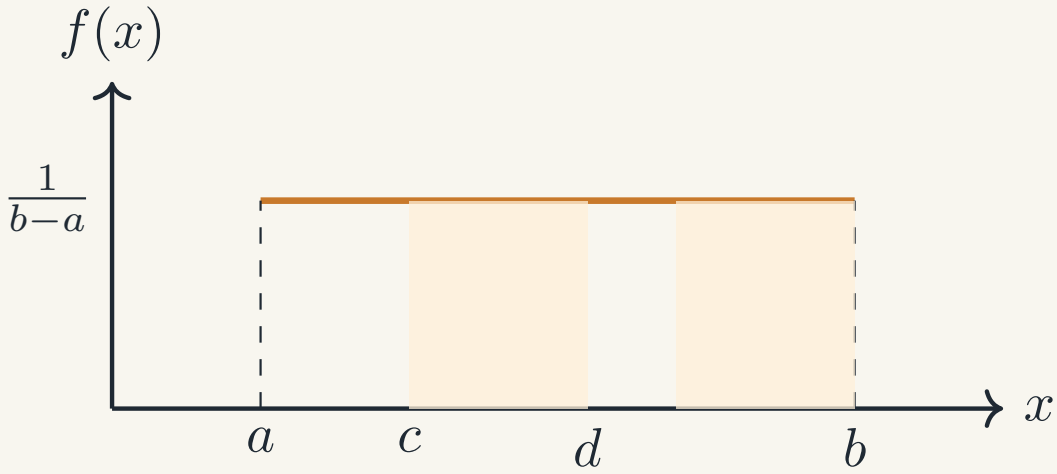
直觉

在区间 (a,b) 内，等长度子区间的概率相同。

均匀分布的意义

若 $X \sim \mathrm{U}(a,b)$，则对 $a < c < d < b$，

$$\mathbb{P}(c \le X \le d) = \int_c^d \frac{1}{b-a}\,\mathrm{d}x = \frac{d-c}{b-a}.$$

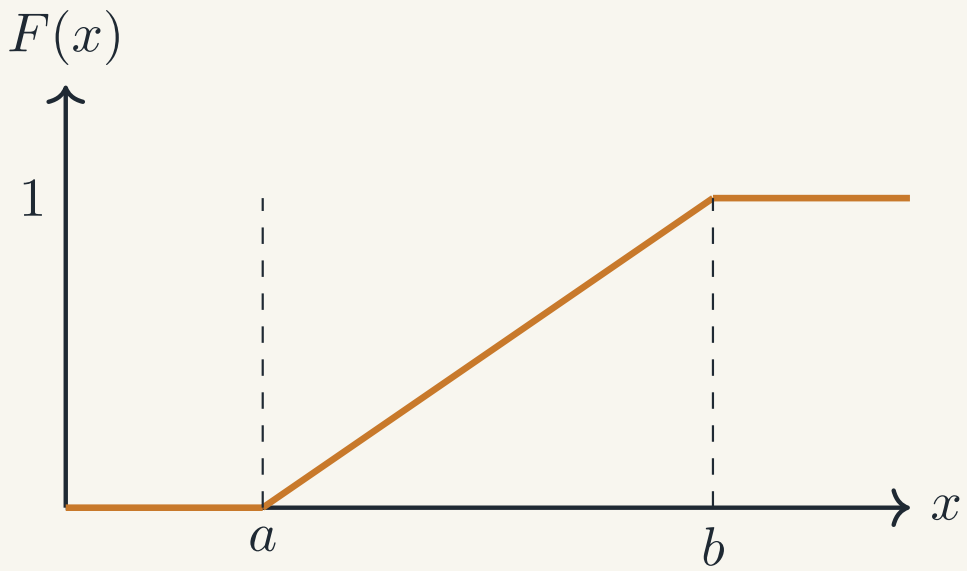


概率等于曲线下的面积；在均匀分布中，面积只由区间长度决定。

均匀分布的分布函数

若 $X \sim \mathrm{U}(a,b)$，则

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



例：公交候车时间

公交车从上午 7 时起每隔 15 分钟来一班。某乘客到达车站的时间在 7 : 10 到 7 : 30 之间均匀分布。求候车时间不超过 5 分钟的概率。

建模

设乘客于 7 时 X 分到达，则

$$X \sim \mathrm{U}(10, 30).$$

候车不超过 5 分钟发生在

$$10 \le X \le 15 \qquad \text{或} \qquad 25 \le X \le 30.$$

于是

$$\mathbb{P}(\text{候车不超过5 分钟}) = \int_{10}^{15} \frac{1}{20}\,\mathrm{d}x + \int_{25}^{30} \frac{1}{20}\,\mathrm{d}x = \frac{1}{2}.$$

例：均匀分布与二项分布结合

设 $X \sim \mathrm{U}(2,5)$。对 X 进行三次独立观测，求至少有两次观测值大于 3 的概率。

单次观测

$$\mathbb{P}(X > 3) = \int_3^5 \frac{1}{3}\,\mathrm{d}x = \frac{2}{3}.$$

令 Y 为三次观测中大于 3 的次数，则

$$Y \sim \mathrm{Bin}\left(3, \frac{2}{3}\right).$$

所求为

$$\mathbb{P}(Y \geq 2) = \binom{3}{2}\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right) + \left(\frac{2}{3}\right)^3 = \frac{20}{27}.$$

指数分布

定义

若连续型随机变量 X 的密度为

$$f(x)=\begin{cases}\dfrac{1}{\theta}\mathrm{e}^{-x/\theta}, & x>0,\\ 0, & x\le 0,\end{cases}\qquad \theta>0,$$

则称 X 服从参数为 θ 的**指数分布**。

应用背景

常用于刻画某些元件或设备的寿命，如无线电元件、电力设备等。

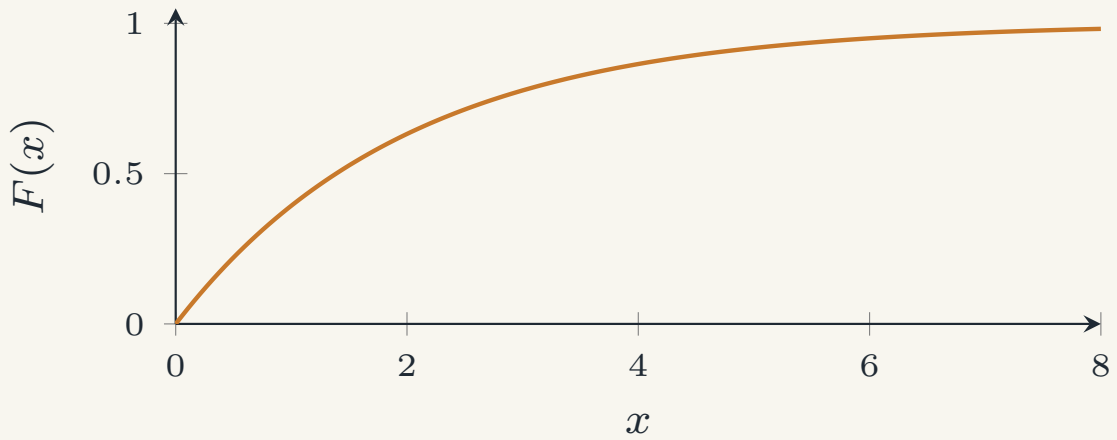
指数分布的分布函数

若 X 服从参数为 θ 的指数分布，则

$$F(x) = \begin{cases} 0, & x \le 0, \\ 1 - \mathrm{e}^{-x/\theta}, & x > 0. \end{cases}$$

因此尾部概率为

$$\mathbb{P}(X > x) = \mathrm{e}^{-x/\theta}, \qquad x > 0.$$



例：日光灯管寿命

设某类日光灯管的使用寿命 X 服从参数 $\theta = 2000$ 的指数分布，单位为小时。

正常使用超过 1000 小时

$$\mathbb{P}(X > 1000) = \mathrm{e}^{-1000/2000} = \mathrm{e}^{-1/2} \approx 0.607.$$

已使用 1000 小时后还能再用 1000 小时

$$\mathbb{P}(X > 2000 \mid X > 1000) = \frac{\mathbb{P}(X > 2000)}{\mathbb{P}(X > 1000)} = \frac{\mathrm{e}^{-1}}{\mathrm{e}^{-1/2}} = \mathrm{e}^{-1/2}.$$

结果仍约为 0.607。

指数分布的无记忆性

指数分布的重要性质是：对 $s, t > 0$，

$$\mathbb{P}(X > s + t \mid X > s) = \mathbb{P}(X > t).$$

证明

$$\mathbb{P}(X > s + t \mid X > s) = \frac{\mathbb{P}(X > s + t)}{\mathbb{P}(X > s)} = \frac{\mathrm{e}^{-(s+t)/\theta}}{\mathrm{e}^{-s/\theta}} = \mathrm{e}^{-t/\theta}.$$

而

$$\mathbb{P}(X > t) = \mathrm{e}^{-t/\theta}.$$

直觉

已经使用了多久，不影响未来还能使用多久的概率规律。

正态分布

定义

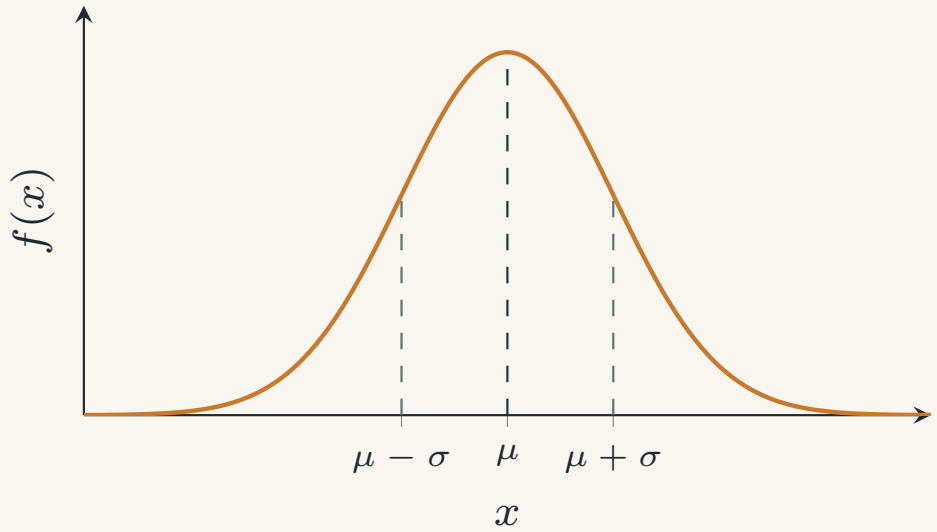
若连续型随机变量 X 的密度为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\},\qquad -\infty<x<+\infty,$$

其中 $\sigma>0$，则称 X 服从参数为 μ,σ 的**正态分布**，记为

$$X\sim \mathrm{N}(\mu,\sigma^2).$$

正态密度的几何特征

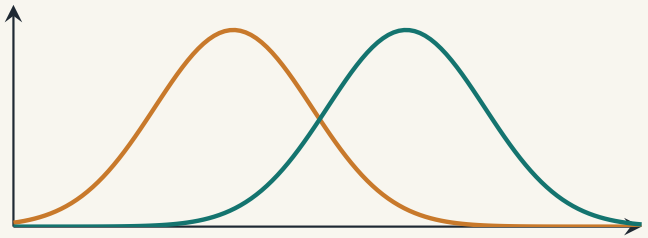


- 关于 $x=\mu$ 对称；
- 在 $x=\mu$ 处取得最大值；
- $x\to\pm\infty$ 时密度趋近于 0；
- 在 $x=\mu\pm\sigma$ 处有拐点。

参数 μ 与 σ 的作用

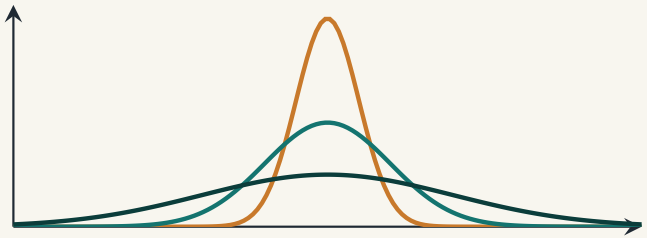
改变 μ

固定 σ，改变 μ，图形形状不变，只沿 x 轴平移。



改变 σ

固定 μ，σ 越小图形越高越瘦；σ 越大图形越矮越胖。



正态分布的背景

实际背景

正态分布常用于拟合测量误差、人的身高体重、产品尺寸误差、弹着点偏差等。

为什么重要?

很多变量受到大量微小、独立随机因素共同影响时，整体结果往往近似服从正态分布。

后续课程联系

中心极限定理将解释为什么大量独立随机变量的和在适当条件下近似服从正态分布。

标准正态分布

当 $\mu = 0, \sigma = 1$ 时，称为**标准正态分布**，记为

$$X \sim \mathrm{N}(0, 1).$$

密度函数为

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}, \qquad -\infty < x < +\infty.$$

分布函数为

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

重要性

一般正态分布都可以通过标准化转化为标准正态分布。

一般正态分布的标准化

若

$$X \sim \mathrm{N}(\mu, \sigma^2),$$

则

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

于是

$$\mathbb{P}(X \leq x) = \mathbb{P}\left(Z \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right).$$

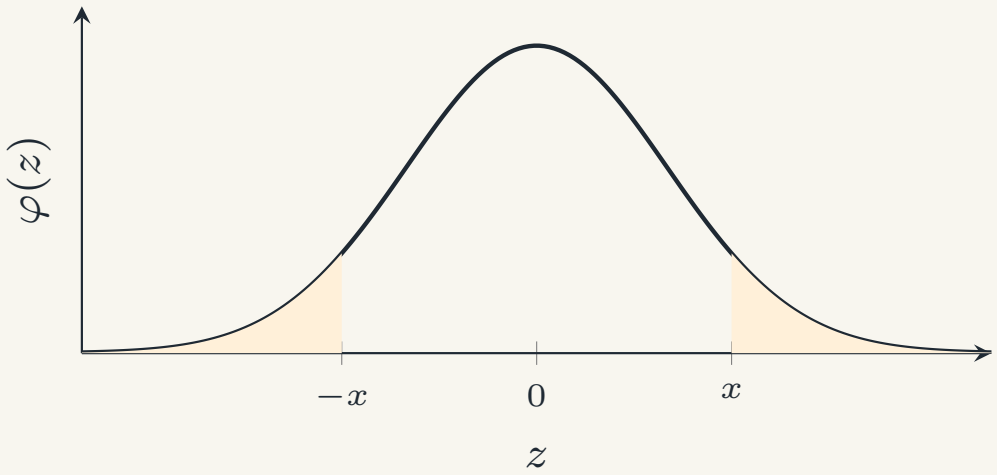
区间概率

$$\mathbb{P}(c \leq X \leq d) = \Phi\left(\frac{d - \mu}{\sigma}\right) - \Phi\left(\frac{c - \mu}{\sigma}\right).$$

标准正态分布的对称性

由于标准正态密度关于 0 对称，

$$\Phi(-x) = 1 - \Phi(x).$$



这条公式常用于查标准正态分布表。

例：温度调节器

温度调节器整定在 $d^\circ\mathrm{C}$，液体温度 X 满足

$$X \sim \mathrm{N}(d, 0.5^2).$$

问题 1

若 $d = 90$，求 $X < 89$ 的概率。

$$\mathbb{P}(X < 89) = \Phi\left(\frac{89 - 90}{0.5}\right) = \Phi(-2) = 1 - \Phi(2) = 0.0228.$$

问题 2

若要求保持液体温度至少为 $80^\circ\mathrm{C}$ 的概率不低于 0.99，即

$$\mathbb{P}(X > 80) \geq 0.99,$$

则

$$\Phi\left(\frac{80 - d}{0.5}\right) \leq 0.01.$$

查表得 $\frac{80-d}{0.5} \leq -2.327$，故 $d \geq 81.1635$。

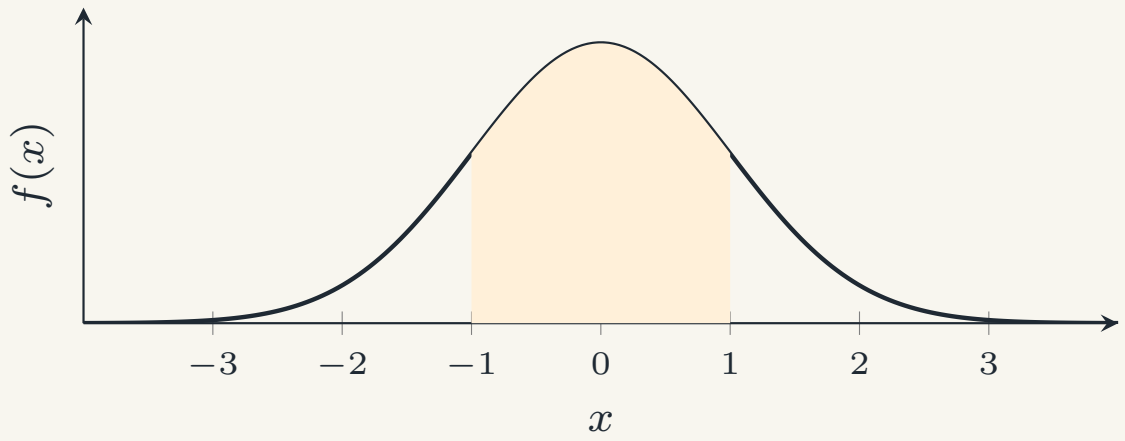
正态分布的 3σ 法则

若 $X \sim \mathrm{N}(\mu, \sigma^2)$，则

$$\mathbb{P}(|X-\mu| \le \sigma) = 2\Phi(1) - 1 \approx 0.6826,$$

$$\mathbb{P}(|X-\mu| \le 2\sigma) = 2\Phi(2) - 1 \approx 0.9544,$$

$$\mathbb{P}(|X-\mu| \le 3\sigma) = 2\Phi(3) - 1 \approx 0.9974.$$

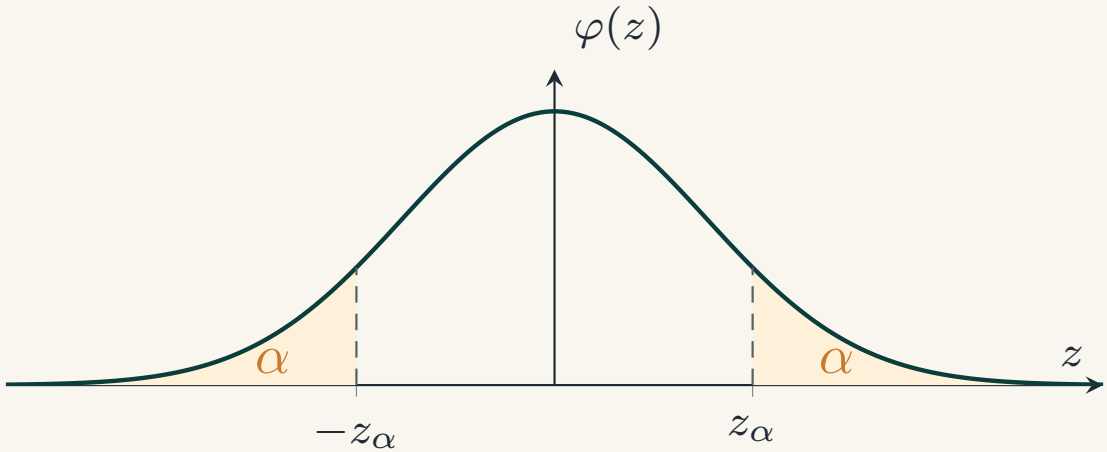


几乎全部取值集中在 $[\mu - 3\sigma, \mu + 3\sigma]$ 内。

标准正态分布的上 α 分位点

设 $Z \sim \mathrm{N}(0,1)$。其**上 α 分位点** z_α 定义为

$$\mathbb{P}(Z > z_\alpha) = \alpha \quad \Longleftrightarrow \quad \Phi(z_\alpha) = 1-\alpha, \qquad 0 < \alpha < 1.$$



图示为 $0 < \alpha < 1/2$ 的情形，此时 $z_\alpha > 0$；两侧阴影面积各为 α。

由标准正态分布的对称性，

$$\mathbb{P}(Z < -z_\alpha) = \mathbb{P}(Z > z_\alpha) = \alpha.$$

连续型随机变量小结

密度与分布函数

$$F(x)=\int_{-\infty}^{x} f(t)\,\mathrm{d}t, \qquad \mathbb{P}(a<X\le b)=\int_{a}^{b} f(x)\,\mathrm{d}x.$$

分布	密度	典型应用
均匀分布	$1/(b-a)$	等可能落在区间内
指数分布	$\theta^{-1}\mathrm{e}^{-x/\theta}$	寿命、等待时间，无记忆性
正态分布	钟形曲线	测量误差、自然波动、中心极限定理

本节作业

课后练习

第五版课后习题 P58：21，24，27。
第六版课后习题 P56：21，24，27。

复习重点

明确密度不是概率；会由密度求分布函数和区间概率；掌握均匀、指数、正态三种常见连续型分布。

5

随机变量的函数的分布

第二章　随机变量及其分布

问题的提出

实际问题中，我们常常更关心随机变量的函数。

几何量

已知圆轴截面直径 D 的分布，求截面面积

$$A = \frac{\pi}{4} D^2$$

的分布。

物理量

已知某时刻噪声电压 V 的分布，求功率

$$W = \frac{V^2}{R}$$

的分布。

一般问题

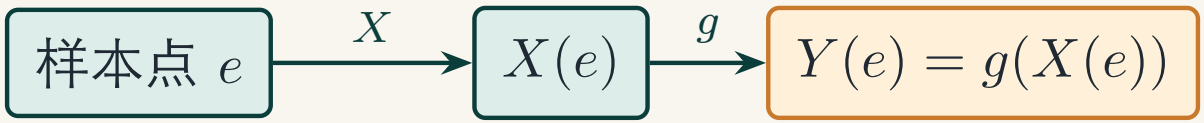
已知随机变量 X 的分布，如何求 $Y = g(X)$ 的分布?

随机变量的函数

设 g 是定义在随机变量 X 可能取值集合上的函数。若

$$Y = g(X),$$

则称 Y 为随机变量 X 的函数。



目标

由 X 的分布求 Y 的分布。

离散型随机变量的函数

若 X 是离散型随机变量，且分布律为

X	x_1	x_2	$\cdots$	x_k
p	p_1	p_2	$\cdots$	p_k

则 $Y = g(X)$ 也是离散型随机变量。初步写出

Y	$g(x_1)$	$g(x_2)$	$\cdots$	$g(x_k)$
p	p_1	p_2	$\cdots$	p_k

关键

若 $g(x_i)$ 中有相同值，应把对应概率相加合并。

例：$Y = X^2$

设 X 的分布律为

X	-1	0	1	2
p	$1/4$	$1/4$	$1/4$	$1/4$

求 $Y = X^2$ 的分布律。

解

Y 的可能取值为 $0, 1, 4$。

$$\mathbb{P}(Y = 0) = \mathbb{P}(X = 0) = \frac{1}{4},$$

$$\mathbb{P}(Y = 1) = \mathbb{P}(X = -1) + \mathbb{P}(X = 1) = \frac{1}{2},$$

$$\mathbb{P}(Y = 4) = \mathbb{P}(X = 2) = \frac{1}{4}.$$

$$Y \sim \begin{pmatrix} 0 & 1 & 4 \\ 1/4 & 1/2 & 1/4 \end{pmatrix}.$$

例：奇偶函数

设离散型随机变量 X 的分布律为

$$\mathbb{P}(X=n)=\frac{1}{2^n},\qquad n=1,2,\ldots.$$

定义

$$Y=g(X)=\begin{cases}-1, & X \text{ 为奇数},\\ 1, & X \text{ 为偶数}.\end{cases}$$

求分布律

$$\mathbb{P}(Y=-1)=\sum_{k=0}^{\infty}\mathbb{P}(X=2k+1)=\sum_{k=0}^{\infty}\frac{1}{2^{2k+1}}=\frac{2}{3}.$$

$$\mathbb{P}(Y=1)=1-\mathbb{P}(Y=-1)=\frac{1}{3}.$$

$$Y\sim\begin{pmatrix}-1 & 1\\ 2/3 & 1/3\end{pmatrix}.$$

连续型随机变量函数的分布：分布函数法

设 X 是连续型随机变量，密度为 $f_X(x)$。令

$$Y = g(X).$$

方法 1：先求分布函数

$$F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(g(X) \le y).$$

把事件 $\{g(X) \le y\}$ 转化为关于 X 的不等式，再利用 X 的已知分布求概率。

最后一步

若 F_Y 可导，则

$$f_Y(y) = F_Y'(y).$$

例：线性变换 $Y = 2X + 8$

设 X 的密度为

$$f_X(x) = \begin{cases} \dfrac{x}{8}, & 0 < x < 4, \\ 0, & \text{其它.} \end{cases}$$

求 $Y = 2X + 8$ 的密度。

分布函数法

$$F_Y(y) = \mathbb{P}(2X + 8 \le y) = \mathbb{P}\left(X \le \frac{y-8}{2}\right).$$

因此

$$f_Y(y) = f_X\left(\frac{y-8}{2}\right) \cdot \frac{1}{2}.$$

由于 $0 < \frac{y-8}{2} < 4$ 等价于 $8 < y < 16$，故

$$f_Y(y) = \begin{cases} \dfrac{y-8}{32}, & 8 < y < 16, \\ 0, & \text{其它.} \end{cases}$$

连续型函数的单调变换公式

设 X 的密度为 f_X，g 处处可导并且严格单调。令

$$Y = g(X).$$

设 h 是 g 的反函数，即 $x = h(y)$。若

$$\alpha < y < \beta$$

为 Y 的取值区间，则

$$f_Y(y) = \begin{cases} f_X(h(y))|h'(y)|, & \alpha < y < \beta, \\ 0, & \text{其它}. \end{cases}$$

使用条件

这个公式要求 g 在相关区间上一一对应；若不是单调，需要分段或用分布函数法。

例：正态变量的线性函数

若

$$X \sim \mathrm{N}(\mu, \sigma^2), \qquad Y = aX + b, \qquad a \neq 0,$$

则反函数为

$$x = h(y) = \frac{y-b}{a}, \qquad |h'(y)| = \frac{1}{|a|}.$$

由变换公式，

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

化简可得

$$\boxed{Y = aX + b \sim \mathrm{N}(a\mu + b, a^2\sigma^2)}.$$

结论

高斯变量的线性函数仍是高斯变量。

例：非单调变换 $Y = X^2$

设 $X \sim \mathrm{U}(-1,1)$，求 $Y = X^2$ 的分布函数与概率密度。

分布函数法

由 $0 \le Y \le 1$，先确定分布函数的分段区间。当 $0 \le y < 1$ 时，

$$F_Y(y) = \mathbb{P}(X^2 \le y) = \mathbb{P}(-\sqrt{y} \le X \le \sqrt{y}) = \frac{\sqrt{y} - (-\sqrt{y})}{2} = \sqrt{y}.$$

写出分布函数，并在 $0 < y < 1$ 上求导，得到

$$F_Y(y) = \begin{cases} 0, & y < 0, \\ \sqrt{y}, & 0 \le y < 1, \\ 1, & y \ge 1, \end{cases} \qquad f_Y(y) = \begin{cases} \dfrac{1}{2\sqrt{y}}, & 0 < y < 1, \\ 0, & \text{其它}. \end{cases}$$

关键

平方函数在 $(-1,1)$ 上不单调。解不等式 $X^2 \le y$ 时，必须同时考虑负数和正数两侧，不能只考虑正根。

连续函数作用于随机变量时要小心

设 g 是连续函数。若 X 是离散型随机变量，则

$$Y = g(X)$$

仍是离散型随机变量，因为 Y 的取值至多仍是有限或可列无限个。

但连续型不一定保持连续型

若 X 是连续型随机变量，$Y = g(X)$ 不一定是连续型随机变量。

例如，设 $X \sim \mathrm{U}(0,2)$，令

$$g(x) = \begin{cases} x, & 0 \le x \le 1, \\ 1, & 1 < x \le 2. \end{cases}$$

则 $Y = g(X)$ 在 1 处有正概率质量，因此不是连续型随机变量。

函数分布小结

离散型

列出 $g(x_k)$，再把相同取值对应的概率相加。

连续型：分布函数法

$$F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(g(X) \le y).$$

把关于 Y 的事件转化为关于 X 的事件。

连续型：单调变换公式

若 g 严格单调，$h = g^{-1}$，则

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

本节作业

课后练习

第五版课后习题 P58：36，37。
第六版课后习题 P66：36，37。

复习重点

遇到 $Y = g(X)$，先判断 X 是离散型还是连续型；再决定用列表合并、分布函数法或单调变换公式。

第二章总小结

随机变量

把随机试验结果数量化：

$$X : S \to \mathbb{R}.$$

分布函数

统一描述任意随机变量：

$$F(x) = \mathbb{P}(X \le x).$$

离散型

用分布律描述：

$$\mathbb{P}(X = x_k) = p_k, \qquad \sum p_k = 1.$$

连续型

用密度描述：

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t.$$

进一步

已知 X 的分布后，可以研究 $Y = g(X)$ 的分布，这是后续概率模型计算的重要基础。