概率统计 B
第二章 随机变量与概率分布

原著：陈家鼎、刘婉如、汪仁官
制作：李东风，邓明华

2025 春季学期

本章目录

本节目录

随机变量概念引入

- 一些事件的结果用变量值表述比较容易，如:
- 独立重复投掷分币 5 次，正面朝上的次数;
- 无放回抽取产品，抽取到的次品个数;
- 数轴上的随机游动，第 i 步移动到的点 K，等等。
- 这些结果值有随机性（取值不能预先确定，但发生的可能性大小可以预估）。

随机变量定义

- **例 1.1** 100 件产品中有 5 件次品，95 件正品。随机抽取 20 件，“抽取出的次品件数”在抽取前是一个随机数值，可能在 0，1，. . ., 5 中取值。
- 但是，每次抽取完毕以后必有一个确定的抽取结果，此抽取结果有一个确定的次品件数。
- **定义 1.1** 对于条件组 S 下的每一个可能结果 ω 都唯一地对应到一个实数值 $X(\omega)$, 则称实值变量 $X(\omega)$ 为一个随机变量，简记为 X。
- 随机变量实际是从结果到数值的一个函数（映射），以试验结果为自变量。

例 1.2

- **例 1.2** 盒中 5 个球，2 白 3 黑。从中随机抽取 3 个。“抽得的白球数” X 是一个随机变量。
- 把 5 个球编号，1，2，3 号为黑球，4，5 号为白球。
- 所有结果有 $C_5^2 = 10$ 种，与 X 值的对应关系可列表。
- X 只能取 0，1，2。“$X=0$“, “$X=1$“, “$X=2$“都是随机事件。
- 由古典概型：

$$P(X=0) = \frac{C_3^3}{C_5^3} = \frac{1}{10}$$

$$P(X=1) = \frac{C_3^2 C_2^1}{C_5^3} = \frac{6}{10}$$

$$P(X=2) = \frac{C_3^1 C_2^2}{C_5^3} = \frac{3}{10}$$

- **例 1.3** 单次射击击中概率 0.8，射击 30 次，“击中目标的次数”X 是随机变量。
- 可以取值 $0, 1, 2, \ldots, 30$。
- “$X = 0$”, “$X = 1$”, “$X = 2$”, $\cdots\cdots$, “$X = 30$”都是随机事件。
- **例 1.4** 单次射击击中概率 0.8，连续射击直到第一次击中未知，所需的“射击次数”X 是随机变量。X 可以取 $1, 2, \ldots$。
- “$X = k(k = 1, 2, 3, \ldots)$”是随机事件。
- **例 1.5** 出租车 400 辆。每天每辆出租车故障概率 0.02。
- 一天内有故障的出租车辆数 X 是随机变量，取值在 $0, 1, 2, \ldots, 400$ 中。

- **例 1.6** 某公共汽车站每隔 5 分钟有一辆汽车通过。
- 一位乘客对于汽车通过的规律完全不知情，所以在任一时刻到达车站都是等可能的。
- 其候车时间 X 是随机变量，取值 $0 \le X < 5$。
- “$X > 2$”, “$X \le 3$”都是随机事件。
- 这里 X 的取值范围是一个区间内的所有实数值，取值是“连续的”。
- **例 1.7** 一门大炮瞄准某个地面目标射击，以目标为坐标原点建立坐标系，y 轴指向从大炮到目标的方向。
- 弹着点与目标的距离 ρ 是一个随机变量，$\rho > 0$。
- 弹着点的直角坐标 (X, Y) 的两个分量 X, Y 是随机变量。
- X, Y 连续取值，还可以取负值。

- 随机变量是重要的概率模型。
- 实际中的随机变量：工业生产中随机抽取的一件产品的质量指标（强度、硬度、光洁度、粘合力、纤度等），医学检验的测量值，问卷调查汇总的答卷选择比例，等等。
- 随机变量按取值范围分为两类：取有限个可能取值或可数个可能取值的，叫做**离散型**随机变量；可以在一个区间或若干个区间取值的，叫做**连续型**随机变量。虽然除此之外还可以有其它类型但本课程不考虑。

本节目录

概率分布

- 离散型随机变量取值范围是有限个值或可数个值。
- 设 X 可取的值为 $x_1, x_2, \ldots, x_k, \ldots$。
- 随机变量作为古典概型、独立试验序列概型等的推广，关键是它能用取值作为事件，并可以计算取值概率。
- 离散型随机变量 X 的每个取值的概率可列表如下

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

称为 X 的概率分布表。

- 概率分布表可简写为:

$$P(X = x_i) = p_i, \quad i = 1, 2, \ldots \tag{2.1}$$

这是一个函数，自变量取值于 $x_i, i = 1, 2, \ldots$，函数值取值在 $p_1, p_2, \ldots$ 中。

- (2.1) 称为 X 的概率分布 (probability distribution), 或概率质量函数 (probability mass function，简记为 PMF)。
- 概率分布的直观理解：X 有一系列不同取值，在每个取值上的概率大小，总数为 1。
- 性质:

$$(1) p_k \geq 0 \quad (k = 1, 2, \ldots)$$
$$(2) \sum_k p_k = 1$$

- 事件组 $\{X = x_k\}, k = 1, 2, \ldots$ 构成完备事件组。

例 1.2（续）

- **例 1.2（续）** X 为 3 黑 2 白共 5 个球中任取 3 个，结果白球个数。
- 其概率分布表为

X	0	1	2
p	0.1	0.6	0.3

- 概率质量函数（PMF）为

$$
\begin{aligned}
P(X=0) &= 0.1 \\
P(X=1) &= 0.6 \\
P(X=2) &= 0.3
\end{aligned}
$$

两点分布

- 设 $0 < p < 1$, 设随机变量 X 只能取 1，0，其分布为

$$\begin{aligned} P(X = 1) &= p \\ P(X = 0) &= 1 - p \end{aligned}$$

称 X 服从**两点分布**，p 称为其**分布参数**。也叫做伯努利分布(Bernoulli)。可记作 $\mathrm{b}(1, p)$ 分布。

例 2.1

- **例 2.1** 100 件产品中，95 正品，5 次品。随机抽取一件。

$$X = \begin{cases} 1 & \text{当取得正品} \\ 0 & \text{当取得次品} \end{cases}$$

- 则

$$P(X = 1) = 0.95$$
$$P(X = 0) = 1 - 0.95$$

$X \sim \mathrm{b}(1, p)$。

- 两点分布仅适用于描述只有两个结果的情况。

二项分布

- 考虑独立试验序列概型问题。设每次试验只有“成功”和“失败”两种可能结果，成功概率为 $p(0 < p < 1), q = 1 - p$，独立重复试验 n 次。
- 令 X 表示 n 次试验中成功的次数。则 X 的取值范围为 $\{0, 1, 2, \ldots, n\}$。
- 在 §1.7 中已经证明

$$P(X = k) = C_n^k p^k q^{n-k}, \quad k = 0, 1, 2, \ldots, n \tag{2.3}$$

- 如果随机变量 X 的分布为 (2.3), 则称 X 服从二项分布（参数为 n, p），记为 $X \sim \mathrm{B}(n, p)$。

- 二项分布所有概率之和等于 1:

$$1 = (p+q)^n = \sum_{k=0}^{n} C_n^k p^k q^{n-k} \quad (\text{二项式定理})$$

- 两点分布就是 $n=1$ 时的二项分布。

泊松分布

- 设 $\lambda > 0$, 若 X 分布为

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (X = 0, 1, 2, \ldots) \tag{2.4}$$

则称 X 服从泊松分布 (参数为 λ)，记为 $X \sim \text{Poisson}(\lambda)$。

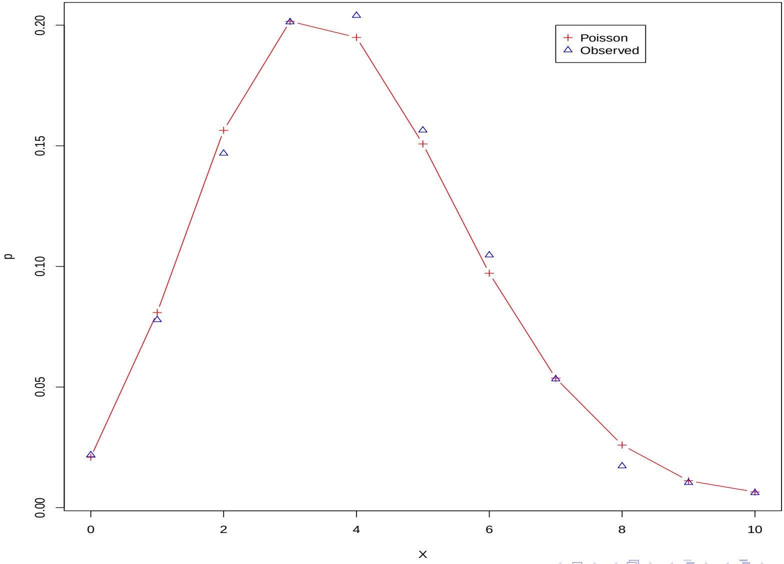
放射粒子数分布

- **例 2.2** 放射性物质在某一段时间放射的粒子数 X 服从泊松分布。
- 每次观察 7.5 秒，共观察 2608 次。设 X 表示每次观察记录下的放射粒子个数。
- 观测结果的频率和 Poisson(3.87) 的概率很接近。

放射粒子数试验的结果表格

X	频数	频率	p_k
0	57	0.022	0.021
1	203	0.078	0.081
2	383	0.147	0.156
3	525	0.201	0.202
4	532	0.204	0.195
5	408	0.156	0.151
6	273	0.105	0.097
7	139	0.053	0.054
8	45	0.017	0.026
9	27	0.010	0.011
≥ 10	16	0.006	0.007
总计	2608	1.000	1.000

放射粒子数试验的结果图形



泊松分布的应用

- 泊松分布还用于:
- 生物学、医学、工业、排队等问题。
- 如：容器内细菌数；铸件或布匹的疵点数；交换台的接入电话数；某路口经过的汽车数。

粒子数问题建模

- 放射粒子数为何服从泊松分布?
- 体积为 V 的一块分割为 n 份相同体积 $\Delta V = \frac{V}{n}$ 的小块，假定:
- (1) 每个特定的小块在 7.5 秒内放出两个以上 α 粒子的概率为 0（实际是很小到可忽略）;
- 小块放出一个 α 例子的概率为

$$p_n = \mu \Delta V = \mu \frac{V}{n}$$

- (2) 各小块放出粒子与否相互独立。

- 则 7.5 秒内体积 V 的大块放射出 k 个粒子，可近似看作在 n 个独立的小块中共有 k 个小块放射出例子：

$$P(X=k)\approx C_n^k p_n^k q_n^{n-k}\quad (q_n=1-p_n)$$

- 令 $n\to\infty$ 可以逼近概率精确值：

$$P(X=k)=\lim_{n\to\infty} C_n^k p_n^k q_n^{n-k}$$

极限的推导

- 记 $\lambda = \mu V$, 则 $p_n = \frac{\lambda}{n}$。

$$
\begin{aligned}
C_n^k p_n^k q_n^{n-k} &= \frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \frac{1}{k!}\cdot\frac{n(n-1)\dots(n-k+1)}{n^k}\cdot\lambda^k\cdot\frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k}
\end{aligned}
$$

- $n\to\infty$ 时, k 不变，第二个因子

$$
\frac{n(n-1)\dots(n-k+1)}{n^k} = \frac{n}{n}\cdot\frac{n-1}{n}\dots\frac{n-(k-1)}{n}\to 1
$$

- 第四个因子中

$$
\begin{aligned}
\left(1-\frac{\lambda}{n}\right)^k &\to 1 \\
\left(1-\frac{\lambda}{n}\right)^n &= \left[\left(1-\frac{\lambda}{n}\right)^{-\frac{n}{\lambda}}\right]^{-\lambda}\to e^{-\lambda}
\end{aligned}
$$

- 于是

$$C_n^k p_n^k q_n^{n-k} \to \frac{\lambda^k}{k!} e^{-\lambda}, \quad \text{当} n \to \infty \tag{2.5}$$

- 即

$$P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (k = 0, 1, 2, \dots)$$

- 具体问题如果能符合类似于（1），（2）的条件，也可能会服从泊松分布。

用泊松分布近似二项分布

- **推论** 对二项分布 $\mathrm{B}(n,p)$，如果 $np \to \lambda > 0(n \to \infty)$, 则

$$C_n^k p_n^k q_n^{n-k} \to \frac{\lambda^k}{k!} e^{-\lambda}, \quad 当 n \to \infty$$

即 p 很小而 n 较大时可以用泊松分布近似计算二项概率。

- $\sum_{k=0}^{\infty} p_k = 1$?
- 用指数函数的泰勒展开：

$$e^{\lambda} = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}$$

- 两边乘以 $e^{-\lambda}$ 即可。

超几何分布

- 设有 N 个同类产品，其中 M 个次品。从中任取 n 个（假定 $n \le N - M$）。则这 n 个中的次品数 X 是离散型随机变量，由第一章例 2.5

$$P(X = m) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n} \quad (m = 0, 1, 2, \ldots, \min(M, n))$$

- 称 X 服从超几何分布（N, M, n 为参数）。

超几何分布与二项分布的关系

- 超几何分布是无放回抽样结果；二项分布可以看成有放回抽样结果。
- 当产品总数 N 很大时，两者分布近似相等。
- 设 $N \to \infty$ 时 $M/N \to p(n, m$ 不变$)$, 则

$$\frac{C_M^m C_{N-M}^{n-m}}{C_N^n} = C_n^m p^m q^{n-m} \quad (N \to \infty)$$

证明

$$
\begin{aligned}
&\frac{C_M^m C_{N-M}^{n-m}}{C_N^n} \\
=&\frac{M!}{(M-m)!m!}\cdot\frac{(N-M)!}{[N-M-(n-m)]!(n-m)!}\cdot\frac{(N-n)!n!}{N!} \\
=&\frac{n!}{m!(n-m)!}\cdot\left(\frac{M\cdot(M-1)\ldots[M-(m-1)]}{N^m}\right)\cdot \\
&\left(\frac{[(N-M)]\cdot[(N-M)-1]\ldots[(N-M)-(n-m-1)]}{N^{n-m}}\right)\cdot \\
&\left(\frac{N^n}{N\cdot(N-1)\cdots[N-(n-1)]}\right)
\end{aligned}
$$

- 第二个因子趋于 p^m; 第三个因子趋于 $(1-p)^{n-m}$; 第四个因子趋于 1。

本节目录

连续型随机变量与概率密度函数

- 例：弹着点与目标间距离；等车时间。
- 随机变量取值不确定，但取每个值的可能性大小可以确定。
- 随机变量的优势：可以更准确地描述事件结果；可以更准确地描述事件概率（概率分布）。
- 一般地考虑 $\{a < X < b\}$ 这样的事件，对连续型随机变量不考虑 $\{X = a\}$ 这样的事件。
- **定义 3.1** 对于随机变量 X，如果存在非负可积函数 $p(x)(-\infty < x < \infty)$，使对任意 $a, b(a < b)$ 都有

$$P(a < X < b) = \int_a^b p(x)dx \tag{3.1}$$

则称 X 为**连续型随机变量**；称 $p(x)$ 为 X 的概率密度函数（probability density funcition, PDF），简称概率密度或密度。

概率密度的解释

- 若 $p(x)$ 在 x_0 处连续，则对很小的 δ,

$$P(X \in (x_0 - \delta, x_0 + \delta)) = \int_{x_0-\delta}^{x_0+\delta} p(x)dx \approx 2\delta p(x_0)$$

  即 $p(x_0)$ 的大小代表了 X 在 x_0 附近取值的概率大小的一个比例。
- 图示。
- 概率本身还与邻域大小 2δ 有关，所以
- 概率密度不是概率！
- 概率密度非负，但不需要小于 1。

连续型随机变量单点概率为零

- 连续型随机变量至少在一个区间内可以取到任意实数值，所以取每个值的概率应该等于零。
- 对正整数 n,

$$
\begin{aligned}
\{X=a\} &\subset \left\{a-\frac{1}{n}<X<a+\frac{1}{n}\right\} \\
P(X=a) &\leq P\left(a-\frac{1}{n}<X<a+\frac{1}{n}\right) \\
&= \int_{a-\frac{1}{n}}^{a+\frac{1}{n}} p(x)dx \to 0 \quad (n\to\infty)
\end{aligned}
$$

所以

$$
P(X=a)=0
$$

概率密度函数积分等于 1

- 按定积分定义

$$\int_{-\infty}^{\infty} p(x)\,dx = \lim_{a\to\infty}\int_{-a}^{a} p(x)\,dx$$

- 对正整数 n, 事件 $A_n = \{-n < X < n\}$ 构成单调递增事件列:

$$\begin{aligned} A_n \subset & A_{n+1}, n = 1, 2, \ldots \\ \bigcup_{n=1}^{\infty} A_n = & \{-\infty < X < \infty\} \\ & P(-\infty < X < \infty) = 1 \end{aligned}$$

由概率的单调极限性质

$$\begin{aligned} 1 = & P(-\infty < X < \infty) = P\left(\bigcup_{n=1}^{\infty} A_n\right) \\ = & \lim_{n\to\infty} P(A_n) = \lim_{n\to\infty}\int_{-n}^{n} p(x)dx = \int_{-\infty}^{\infty} p(x)dx \end{aligned}$$

概率密度性质

- 概率密度为非负可积函数，在 $(-\infty, \infty)$ 积分等于 1。
- 实际中的非离散型的随机变量一般是连续型的，而且 $p(x)$ 至多有有限多个间断点，在其它地方连续。
- 概率密度改变单个点的数值仍为 X 的密度。
- 但是若 $p_1(x), p_2(x)$ 都是 X 的密度且都在 x_0 连续，则 $p_1(x_0) = p_2(x_0)$。

均匀分布

- 若 X 有概率密度

$$p(x) = \begin{cases} \frac{1}{b-a} & \text{当} x \in [a, b] \\ 0 & \text{其它} \end{cases}$$

则称 X 服从 $[a, b]$ 区间上的**均匀分布** (uniform distribution)，记为 $X \sim \mathrm{U}[a, b]$。

- 对 $a \leq c < d \leq b$,

$$P(c < X < d) = \int_c^d p(x) = \frac{d-c}{b-a}$$

- X 取值于 $[a, b]$ 中任一区间的概率与该区间长度成正比。
- 概率密度是常数，说明 X 取 $[a, b]$ 内任何一点附近的值的可能性大小都是相同的，所以叫“均匀”分布。
- 例：§2.1 例 1.6 等车时间服从均匀分布。

指数分布

- 若 X 有概率密度

$$p(x)=\begin{cases}\lambda e^{-\lambda x} & 当 x\geq 0\\ 0 & 当 x<0\end{cases}\quad(\lambda>0)$$

则称 X 服从**指数分布** (exponential distribution)（参数为 λ），记作 $X\sim \mathrm{E}(\lambda)$。

- 对 $0\leq a<b$ 有

$$\begin{aligned}P(a<X<b)=&\lambda\int_a^b e^{-\lambda x}dx=\int_{\lambda a}^{\lambda b}e^{-t}dt\\ =&e^{-\lambda a}-e^{-\lambda b}\\ P(X>a)=&\lambda\int_a^{\infty}e^{-\lambda x}dx=e^{-\lambda a}\\ P(X<b)=&\lambda\int_0^b e^{-\lambda x}dx=1-e^{-\lambda b}\end{aligned}$$

$$\int_{-\infty}^{\infty}p(x)dx=\int_0^{\infty}p(x)dx=1$$

正态分布

- 若 X 有概率密度

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \quad (-\infty < x < \infty) \tag{3.4}$$

$$(-\infty < \mu < \infty, \sigma > 0)$$

  则称 X 服从正态分布 $\mathrm{N}(\mu, \sigma^2)$（normal distribution, 或 Gaussian distribution），简记为 $X \sim \mathrm{N}(\mu, \sigma^2)$。
- 概率密度图形演示。μ, σ^2 变化时曲线的变化。
- $p(x)$ 曲线呈钟形，最大值点在 $x = \mu$，关于 $x = \mu$ 轴对称；
- 在 $x = \mu \pm \sigma$ 处有拐点（二阶导数等于零的点，是曲线由凹变凸或由凸变凹的点）。
- 当 $x \to \pm\infty$ 时曲线以 x 轴为渐近线。
- μ 决定曲线中心位置；σ 越大，曲线越平缓，σ 越小，曲线越陡峭。

正态分布密度性质

- N(0,1) 叫做标准正态分布，其分布密度为

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

- 用微积分的二重积分极坐标变换可以证明 $\int_{-\infty}^{\infty} \phi(x)dx = 1$。

$$I = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dxdy = \left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \right)^2$$

作极坐标变换，令 $x = r\cos\varphi, y = r\sin\varphi$,

$$\begin{vmatrix} \frac{dx}{dr} & \frac{dy}{dr} \\ \frac{dx}{d\varphi} & \frac{dy}{d\varphi} \end{vmatrix} = \begin{vmatrix} \cos\varphi & \sin\varphi \\ -r\sin\varphi & r\cos\varphi \end{vmatrix} = r$$

$$I = \int_{0}^{\infty} \int_{0}^{2\pi} e^{-\frac{r^2}{2}} rdrd\varphi = 2\pi$$

- 一般的 $N(\mu, \sigma^2)$ 密度可以写成

$$\frac{1}{\sigma}\phi(\frac{x-\mu}{\sigma})$$

- 由 $\int_{-\infty}^{\infty}\phi(x)dx = 1$, 作变换 $\frac{x-\mu}{\sigma} = t$ 可以证明 $N(\mu, \sigma^2)$ 密度积分等于 1。
- 测量误差和许多质量指标，如一批产品的长度、强度等，可以看作或近似看作服从正态分布。
- 正态分布在实际中最为常用，以至于一些不服从正态分布的数据也用正态分布来处理，这也是不合适的。

正态随机变量在一个区间的取值概率

- 设 $X \sim \mathrm{N}(0,1), a < b$。记

$$\Phi(x) = \int_{-\infty}^{x} \phi(t)dt$$

则

$$P(a < X < b) = \int_{a}^{b} \phi(t)dt = \Phi(b) - \Phi(a)$$

(演示)

- 由 $\phi(x)$ 的对称性

$$\Phi(-x) = 1 - \Phi(x), \quad x \in (-\infty, \infty)$$

(图示)

- 函数 $\Phi(x)(x>0)$ 已制成表格，见 P.431 附表 1。
- 现在一般用统计软件计算，在 R 软件中为 `pnorm(x)`。
- 如：

$$\begin{aligned}
P(1<X<2) =& \Phi(2)-\Phi(1)=0.9773-0.8413=0.1360 \\
P(-1<X<1) =& \Phi(1)-\Phi(-1)=\Phi(1)-[1-\Phi(1)] \\
=& 2\Phi(1)-1=2\times 0.8413-1=0.6826
\end{aligned}$$

- 对 $N(\mu,\sigma^2)$,

$$
\begin{aligned}
P(a<X<b) &= \int_a^b \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)dx \\
&= \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \phi(t)dt \quad (\text{令}t=\frac{x-\mu}{\sigma}) \\
&= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \\
P(X<b) &= \Phi\left(\frac{b-\mu}{\sigma}\right) \\
P(X>a) &= 1-\Phi\left(\frac{a-\mu}{\sigma}\right)
\end{aligned}
$$

例 3.3

- **例 3.3** 设 $X \sim \mathrm{N}(2, 0.3^2)$, 求 $P(X > 2.4)$。
- **解**

$$P(X > 2.4) = 1 - \Phi\left(\frac{2.4 - 2}{0.3}\right) = 1 - \Phi(1.33)$$

- 附表中 $\Phi(1.32) = 0.90658, \Phi(1.35) = 0.91149$。
- 线性插值公式:

$$\begin{aligned} f(x) &\approx f(a) + \frac{f(b) - f(a)}{b - a} \cdot (x - a) \\ \Phi(1.33) &\approx 0.90658 + \frac{0.91149 - 0.90658}{1.35 - 1.32} \cdot (1.33 - 1.32) \\ &= 0.9082167 \end{aligned}$$

- $P(X > 2.4) = 1 - 0.9082167 = 0.0918$。
- 用 R 软件直接计算: `1 - pnorm(2.4, 2, 0.3)`, 结果为 0.09121122。

正态分布的经验规则

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

$$\begin{aligned} P(X \in (\mu - \sigma, \mu + \sigma)) =& 0.6827 \\ P(X \in (\mu - 2\sigma, \mu + 2\sigma)) =& 0.9545 \\ P(X \in (\mu - 3\sigma, \mu + 3\sigma)) =& 0.9973 \end{aligned}$$

- 基本在 $(\mu - 2\sigma, \mu + 2\sigma)$ 内取值（超过 95%）。
- 几乎不在 $(\mu - 3\sigma, \mu + 3\sigma)$ 之外取值（不到千分之三）。

伽玛分布

- 若 X 有概率密度

$$p(x)=\begin{cases}\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}, & x>0\\ 0, & x\le 0\end{cases}\ (\alpha>0,\beta>0) \tag{3.6}$$

  则称 X 服从伽玛分布 (gamma distribution)，简记为 $X\sim\Gamma(\alpha,\beta)$。
- 其中

$$\Gamma(\alpha)=\int_0^\infty x^{\alpha-1}e^{-x}dx \quad (\alpha>0)$$

- $\Gamma(1)=1,\Gamma(\frac{1}{2})=\sqrt{\pi}$。
- $\Gamma(\alpha+1)=\alpha\Gamma(\alpha)(\alpha>0)$； $n!=\Gamma(n+1)$。

- $\Gamma(1, \beta)$ 是指数分布 $\mathrm{Exp}(\beta)$。
- $\Gamma(\alpha, 1)$ 叫做标准伽玛分布。
- $\Gamma(\frac{n}{2}, \frac{1}{2})$ 为 $\chi^2(n)$ 分布。
- 密度演示。

威布尔分布

- 若 X 有概率密度

$$p(x) = \begin{cases} m\frac{x^{m-1}}{\eta^m}\exp\left\{-\left(\frac{x}{\eta}\right)^m\right\}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

  则称 X 服从威布尔分布 (Weibull distribution), 简记为 $X \sim \mathrm{W}(m, \eta)$, 其中 $m > 0, \eta > 0$, m 称为形状参数 (shape parameter)， η 称为尺度参数 (scale parameter)。
- 许多机电产品的寿命服从威布尔分布。
- $\mathrm{W}(1, \eta)$ 是参数为 $\frac{1}{\eta}$ 的指数分布。
- 威布尔分布和指数分布在工业产品的寿命与可靠性研究中有广泛应用。
- 密度演示。

本节目录

分布函数

- 分布密度不利于直接计算概率。比如计算连续型随机变量 X 的概率 $P(a < X < b)$ 需要对密度积分。
- 引入"分布函数"。
- **定义 4.1** 设 X 是一个随机变量，称函数

$$F(x) = P(X \leq x) \quad (-\infty < x < \infty) \tag{4.1}$$

为 X 的分布函数 (distribution function, 或 cumulative distribution function, CDF)。

- 任何一个随机变量都有分布函数。

分布函数的性质

(1) $0 \le F(x) \le 1\ (-\infty < x < \infty)$;
(2) $F(x)$ 是 x 的单调递增函数;
(3) $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$;
(4) $F(x)$ 是 x 右连续函数;
(5) $P(a < X \le b) = F(b) - F(a)$。
(6) 若 X 为连续型，则 $P(a < X < b) = F(b) - F(a)$。

- 性质 (1), (2) 由定义和 $A \subset B \Longrightarrow P(A) \leq P(B)$ 可得。
- 性质 (3), (4) 需要概率的公理化定义中的性质：完全可加性，单调事件的概率极限。
- 性质 (5):

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

且不相容所以

$$\begin{aligned} P(X \leq b) =& P(X \leq a) + P(a < X \leq b) \\ F(b) =& F(a) + P(a < X \leq b) \end{aligned}$$

- 随机变量 X 的分布函数可记作 $F_X(x)$, 随机变量 Y 的分布函数可记作 $F_Y(x)$ 或 $F_Y(y)$ (注意函数的自变量符号的选用不影响函数本身)。

例 4.1

- **例 4.1** 设 $X \sim \mathrm{b}(1, p)$，$q = 1 - p$, 则

$$F(x) = \begin{cases} 0 & x < 0 \\ q & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

- 作图并说明 $F(x)$ 的分段表达式。

连续型随机变量的分布函数

- 设 X 是连续型随机变量，有密度 $p(x)$, 分布函数 $F(x)$，则

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} p(t)dt \tag{4.2}$$

- $F(x)$ 是 $p(x)$ 的可变上限的定积分，也是 $p(x)$ 的一个原函数。
- (1) $F(x)$ 是 $x \in (-\infty, \infty)$ 的连续函数;
- (2) 对于 $p(x)$ 的连续点 x_0 而言有

$$F'(x_0) = p(x_0)$$

(注意密度函数可以修改单个点的函数值而仍为原随机变量密度)

- 若 $p(x)$ 只有至多有限个间断点，则对非间断点的 x

$$p(x) = F'(x) \tag{4.3}$$

例 4.2

- **例 4.2** 已使用了 t 小时的电子管在以后的 Δt 小时内损坏概率为 $\lambda\Delta t + o(\Delta t)$, $\lambda > 0$ 不依赖于 t。
- 电子管寿命为 0 的概率是 0。求电子管在 T 小时内损坏的概率。
- 令 X 为电子管的寿命，对于成批电子管 X 是一个随机变量，设分布函数为 $F(x)$。求 $P(X \leq T)$。
- “已使用了 t 小时的电子管在以后的 Δt 小时内损坏的概率”:

$$P(t < X \leq t + \Delta t | X > t) = \lambda\Delta t + o(\Delta t)$$

- 左边等于

$$\frac{P(t < X \leq t + \Delta t)}{P(X > t)} = \frac{F(t + \Delta t) - F(t)}{1 - F(t)}$$

- 于是

$$
\begin{aligned}
\frac{F(t+\Delta t)-F(t)}{1-F(t)} &= \lambda\Delta t + o(\Delta t) \\
\frac{F(t+\Delta t)-F(t)}{\Delta t} &= [1-F(t)]\left[\lambda + \frac{o(\Delta t)}{\Delta t}\right]
\end{aligned}
$$

- 令 $\Delta t \to 0$

$$
\begin{aligned}
F'(t) &= \lambda[1-F(t)] \\
\frac{F'(t)}{1-F(t)} &= \lambda \\
\frac{d}{dt}\log[1-F(t)] &= -\lambda \\
F(t) &= 1 - c_1 e^{-\lambda t}
\end{aligned}
$$

- 联系 $F(0)=0$ 有

$$
\begin{aligned}
F(t) &= 1-e^{-\lambda t}, \quad (t>0) \\
F(t) &= 0, \quad (t \le 0)
\end{aligned}
$$

- 电子管子 T 小时内损坏的概率为

$$P(X \leq T) = F(T) = 1 - e^{-\lambda T}$$

- 分布密度为

$$p(t) = F'(t) = \begin{cases} \lambda e^{-\lambda t} & \text{当} t > 0 \\ 0 & \text{当} t \leq 0 \end{cases}$$

- 即 X 服从参数为 λ 的指数分布。

随机变量函数

- 设 $f(x)$ 是一个函数，随机变量 $f(X)$ 是随机变量 Y, 当 $X=x$ 时，Y 取值 $y=f(x)$。记作 Y = f(X)
- 相当于复合函数 $Y=f(X)=f(X(\omega))$。
- 如：设 X 是分子的速率，Y 是分子的动能，则

$$Y=\frac{1}{2}mX^2 \quad (m \text{ 为分子的质量})$$

- 已知 X 的分布后如何确定 $Y=f(X)$ 的分布？

离散型随机变量的函数

- 若 X 是离散型随机变量，取值 $x_1, x_2, \dots, x_k, \dots,$ $P(x_k) = p_k(k = 1, 2, \dots),$
- 则 $Y = f(X)$ 取值 $f(x_1), f(x_2), \dots, f(x_k), \dots,$ 若各 $f(x_k), k = 1, 2, \dots$ 互不相同则

$$P(Y = f(x_k)) = p_k \tag{*}$$

(*) 即为 Y 的概率分布。

- 若 $f(x_k), k = 1, 2, \dots$ 有重复值，设所有互不相等的值为 $y_1, y_2, \dots,$ 则

$$P(Y = y_k) = \sum_{f(x_j)=y_k} p_j, \quad k = 1, 2, \dots$$

例 4.3

- **例 4.3** X 概率分布为

X	0	1	2	3	4	5
$P(X=x_i)$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{12}$	$\frac{2}{9}$	$\frac{1}{9}$

- 则 $Y=2X+1$ 的概率分布为:

Y	1	3	5	7	9	11
$P(Y=y_i)$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{12}$	$\frac{2}{9}$	$\frac{1}{9}$

例 4.4

- **例 4.4** X 同例 4.3, 但 $Y=(X-2)^2$。
- X 概率分布:

X	0	1	2	3	4	5
$P(X=x_i)$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{12}$	$\frac{2}{9}$	$\frac{1}{9}$

- 对应到每个（可重复）$f(x_i)$ 的概率:

Y	4	1	0	1	4	9
$P(Y=f(x_i))$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{12}$	$\frac{2}{9}$	$\frac{1}{9}$

- 合并:

Y	0	1	4	9
$P(Y=y_i)$	$\frac{1}{3}$	$\frac{1}{6}+\frac{1}{12}$	$\frac{1}{12}+\frac{2}{9}$	$\frac{1}{9}$

连续型随机变量函数

- **例 4.5** $X \sim \mathrm{N}(\mu, \sigma^2)$, $Y = \frac{X-\mu}{\sigma}$。
- 设 $Y \sim F_Y(y)$,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(\frac{X-\mu}{\sigma} \leq y) \\ &= P(X \leq \mu + \sigma y) \\ &= F_X(\mu + \sigma y) \end{aligned}$$

- 求导得

$$p_Y(y) = p_X(\mu + \sigma y)\sigma$$

-

$$p_Y(y) = p_X(\mu + \sigma y)\sigma$$

- 而

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

- 代入得

$$p_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{[(\mu + \sigma y) - \mu]^2}{2\sigma^2}\right\} = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$$

- 即 $Y \sim \mathrm{N}(0, 1)$。

求连续型随机变量函数的密度的一般方法

- 证明用到：
- 若连续型随机变量的密度在除去几个点之外连续，则 $F'(x) = p(x)$ 对除去几个点后；
- **定理（习题七第 16 题）** 若 $F'(x)$ 连续，则 X 是连续型随机变量，$F'(x)$ 为密度。
- 以上通过计算 Y 的分布函数并用 X 的分布函数表示，然后求导得到密度函数的方法是一般性的。

例 4.6

- **例 4.6** $X \sim \mathrm{N}(\mu, \sigma^2)$, $Y = a + bX$(a, b 为常数，$b \neq 0$)。
- 设 $b > 0$,

$$
\begin{aligned}
F_Y(y) =& P(Y \leq y) = P(a + bX \leq y) \\
=& P\left(X \leq \frac{y-a}{b}\right) = F_X\left(\frac{y-a}{b}\right)
\end{aligned}
$$

求导得

$$
p_Y(y) = p_X\left(\frac{y-a}{b}\right)\frac{1}{b}
$$

- 当 $b < 0$ 时，

$$\begin{aligned} F_Y(y) =& P(Y \le y) = P(a + bX \le y) \\ =& P\left(X \ge \frac{y-a}{b}\right) = 1 - F_X\left(\frac{y-a}{b}\right) \end{aligned}$$

求导得

$$p_Y(y) = -\,p_X(\frac{y-a}{b})\frac{1}{b}$$

- 于是

$$
\begin{aligned}
p_Y(y) =& \frac{1}{|b|} p_X\left(\frac{y-a}{b}\right) \\
=& \frac{1}{\sqrt{2\pi}|b|\sigma} \exp\left\{-\frac{(y-a-b\mu)^2}{2b^2\sigma^2}\right\}
\end{aligned}
$$

- 即 $Y \sim \mathrm{N}(a+b\mu, b^2\sigma^2)$。
- 正态分布经线性变换（斜率不为零）仍服从正态分布。
- 特别地，正态分布 $X \sim \mathrm{N}(\mu, \sigma^2)$ 可标准化:

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

- 反之，若 $X \sim \mathrm{N}(0,1)$, $Y = \mu + \sigma X$, 则 $Y \sim \mathrm{N}(\mu, \sigma^2)$。

例 4.7

- **例 4.7** 圆片直径服从 U[5, 6]。求圆片面积概率分布。
- $Y = \frac{1}{4}\pi X^2$。易见 Y 取值范围是 $[\frac{\pi}{4}5^2, \frac{\pi}{4}6^2]$。
- 对于均匀分布 $\mathrm{U}[a, b]$, 其密度为

$$p(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & \text{其它} \end{cases}$$

- 所以分布函数为

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & x \in [a, b] \\ 1 & x > b \end{cases}$$

- 这里 $a=5, b=6$。当 $y\in[\frac{\pi}{4}5^2,\frac{\pi}{4}6^2]$ 时

$$
\begin{aligned}
F_Y(y) =&P(\frac{\pi}{4}X^2\le y) = P(|X|\le\sqrt{\frac{4y}{\pi}})\\
=&P(X\le\sqrt{\frac{4y}{\pi}}) = \sqrt{\frac{4y}{\pi}}-5\\
p_Y(y) =&F'(y) = \frac{1}{\sqrt{\pi y}}
\end{aligned}
$$

- 所以

$$
p_Y(y)=\begin{cases}\frac{1}{\sqrt{\pi y}} & \text{当 } \frac{25}{4}\pi\le y\le 9\pi\\ 0 & \text{其它}\end{cases}
$$

分布函数的导数

- **定理 (习题七第 17 题)** 如果随机变量 X 的分布函数 $F(x)$ 满足以下条件:
  1. $F(x)$ 连续;
  2. 存在 $x_1 < x_2 < \cdots < x_n (n \geq 1)$, 在区间 $(-\infty, x_1), (x_1, x_2), \ldots, (x_{n-1}, x_n), (x_n, \infty)$ 上 $F'(x)$ 存在且连续。
- 令

$$f(x) = \begin{cases} F'(x) & \text{当 } F'(x) \text{ 存在时} \\ 0 & \text{当 } F'(x) \text{ 不存在时} \end{cases}$$

- 则 $f(x)$ 是 X 的分布密度。

例子

- 若 $X \sim N(0,1)$, 求 $Y = X^2$ 的概率密度
- 考察 $X^2 = y$ 的反函数时，需要分区域考虑。当 $x > 0$ 时，反函数 $x = \sqrt{y}$, 当 $x < 0$ 时，反函数 $x = -\sqrt{y}$. 于是事件 $\{Y = X^2 \le y\}$ 等价于事件 $\{0 \le X \le \sqrt{y}\} \cup \{-\sqrt{y} \le X \le 0\}$. 考虑到对称性，于是 Y 的分布函数

$$
\begin{aligned}
F_Y(y) &= P(Y \le y) = P(X^2 \le y) \\
&= 2P(0 \le X \le \sqrt{y}) = 2\int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt
\end{aligned}
$$

- 于是密度函数

$$
\begin{aligned}
p_Y(y) &= 2\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\sqrt{y})^2} \frac{1}{2} y^{-\frac{1}{2}} \\
&= \frac{1}{2\pi} y^{-\frac{1}{2}} e^{-\frac{1}{2}y} = \frac{\left(\frac{1}{2}\right)^{1/2}}{\Gamma\left(\frac{1}{2}\right)} y^{\frac{1}{2}-1} e^{-\frac{1}{2}y}
\end{aligned}
$$

即为自由度为 1 的卡方分布 χ_1^2, 或者 Gamma$(1/2, 1/2)$.

一般情况

- **定理:** 设 X 有密度函数 $f(x)$, $D \subset \mathcal{R}$, $Y = g(X)$, $P(Y \in D) = 1$. 如果存在函数 $h_i(y)$ 使得
  1. 对 $y \in D$, $\{Y = y\} = \bigcup_{i=1}^{n} \{X = h_i(y)\}$,
  2. 每个 $h_i(y)$ 是 D 到其值域 D_i 的可逆映射, 有连续的导数,
  3. 值域 $D_1, D_2, \cdots, D_n$ 互不相交,

  则 Y 有密度函数

$$f_Y(y) = \begin{cases} \sum_{i=1}^{n} f(h_i(y))|h_i'(y)|, & y \in D, \\ 0, & y \in \overline{D}. \end{cases}$$

例 4.8

- **例 4.8** 设 X 的密度函数为 $p_X(x)$, 仅有至多有限个不连续点，函数 $f(x)$ 导数 $f'(x)$ 连续且处处大于零。求 $Y=f(X)$ 的密度 $p_Y(y)$。
- **解** 这时 $f(x)$ 是严格单调递增连续函数，设其值域为 $(A,B)(-\infty \le A < B \le \infty)$, 反函数为 $g(y)$, 在 (A,B) 有定义。
- $g'(y)$ 存在:

$$g'(y) = \frac{1}{f'(g(y))}$$

- $\forall y \in (A,B)$,

$$\begin{aligned}
\{f(X) \le y\} &= \{X \le g(y)\} \\
F_Y(y) &= P(Y \le y) = P(f(X) \le y) \\
&= P(X \le g(y)) = F_X(g(y)) \\
p_Y(y) &= F'_Y(y) = p_X(g(y))g'(y) = p_X(g(y))\frac{1}{f'(g(y))}
\end{aligned}$$

- 对 $y \leq A, F_Y(y) = P(Y \leq A) = 0, p_Y(y) = F'_Y(y) = 0;$
- 对 $y \geq B, F_Y(y) = P(Y \leq B) = 1, p_Y(y) = F'_Y(y) = 0;$
- 所以

$$F_Y(y) = \begin{cases} 0 & y \leq A \\ F_X(g(y)) & y \in (A, B) \\ 1 & y \geq B \end{cases}$$

$$p_Y(y) = \begin{cases} p_X(g(y))\frac{1}{f'(g(y))} & y \in (A, B) \\ 0 & y \notin (A, B) \end{cases}$$

- 如果 $f(x)$ 是连续可导严格单调下降函数（$f'(x) < 0$），则推导中有一步改变：

$$P(f(X) \leq y) = P(X \geq g(y)) = 1 - F_X(g(y))$$

- 于是

$$F_Y(y) = \begin{cases} 0 & y \leq A \\ 1 - F_X(g(y)) & y \in (A, B) \\ 1 & y \geq B \end{cases}$$

$$p_Y(y) = \begin{cases} -p_X(g(y))\frac{1}{f'(g(y))} & y \in (A, B) \\ 0 & y \notin (A, B) \end{cases}$$

- 如果 $f(x)$ 不是连续可导且严格单调函数，则这样的方法不适用。

例 4.9

- **例 4.9** 设随机变量 $X \sim F(x)$ 且 $F(x)$ 是连续函数，则 $Y = F(X) \sim \mathrm{U}[0,1]$。
- $F(x)$ 的值域为 $[0,1]$ 所以 Y 取值于 $[0,1]$。
- 对 $y \in (0,1)$，必有 x_0 使 $F(x_0) = y$。
- 这样的 x_0 不一定唯一，取

$$x_0 = \sup\{x : F(x) \le y\}$$

应有 $x_0 \in (-\infty, \infty)$, 否则与 $\lim_{x\to\infty} F(x) = 1$ 矛盾；
- 且 $F(x_0) = y$,

$$F(x) \le y \Leftrightarrow x \le x_0$$

- 则

$$F_Y(y) = P(Y \le y) = P(X \le x_0) = F(x_0) = y$$

- 还要证明 $F_Y(y)=0(y\le 0), F_Y(y)=1(y\ge 1)$。
- 第二条显然 (必然事件)；第一条只要证 $F_Y(0)=0$。
- 因为 $F(x)$ 连续，设

$$a=\sup\{x:F(x)=0\}$$

- 若 a 不存在 (当 $F(x)$ 处处为正时)，则 $\{F(X)\le 0\}$ 是不可能事件。
- 若 a 存在，由 $F(x)$ 连续知 $F(a)=0$, 于是

$$\begin{aligned} F(x)\le 0 &\Longleftrightarrow x\le a \\ F_Y(0) &= P(F(X)\le 0) \\ &= P(X\le a)=F(a)=0 \end{aligned}$$

例 4.10

- **例 4.10** 设函数 $F(x)$ 具有下列性质：
  1. $0 \le F(x) \le 1,\ \forall x \in (-\infty, \infty)$;
  2. $F(x)$ 是 x 的单调递增函数；
  3. $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$;
  4. $F(x)$ 是右连续函数；
- 令

$$g(y) = \inf\{x : F(x) \ge y\} \quad (0 < y < 1) \qquad (*)$$

- 若 $U \sim \mathrm{U}(0,1)$, 则 $X = g(U) \sim F(x)$。

- (*) 保证 $\forall y \in (0,1)$

$$F(x) \geq y \Longleftrightarrow x \geq g(y)$$

- 于是

$$P(X \leq x) = P(g(U) \leq x) = P(U \leq F(x)) = F(x)$$