工程概率统计 Probability and Statistics for Engineering

第二章随机变量与分布 Chapter 2 Random Variables and Distributions

Chapter 2 Random Variables and Distributions

2.1 引言

- 2.1.1 随机变量的定义 - 2.1.2 概率分布的描述 - 2.1.3 期望与方差

2.2 常见离散分布

- 2.3 常见连续分布 - 2.4 随机变量的变换

2.1 引言

- 在第1章中，我们学习了基本概念，如随机试验、样本空间、随机事件、概率的古典模型、概率的几何模型等。

- 处理复杂问题时，我们需要一种工具，能够统一研究随机事件。

- 目标是将问题简化为函数运算，从而能够使用微积分等数学工具。

- 随机变量就是这样的工具：它连接了现实世界与数学世界。 - 简而言之，随机变量将随机试验的每个可能结果转换为一个数字。 - 例如，假设某组织想调查公众对某项提案的意见，通过随机抽取50个人，询问他们是支持还是反对。 - 根据古典概率，有 2^{50} 种可能结果。 - 然而，如果我们用 X 表示支持该提案的人数，那么 X 就是一个随机变量，其取值范围仅限于 $\{0, 1, 2, ..., 50\}$。

- 因此，随机变量帮助我们以精确和简洁的方式描述问题。

2.1.1 随机变量的定义

- 这里我们给出**随机变量**的数学定义。

随机变量

随机变量（简称 r.v.）是定义在样本空间上的**实值函数**，通常用大写字母如 X, Y, Z 表示。

例 2.1

- 情人节当天，一家餐厅提供"幸运情侣"折扣，可以为情侣们的浪漫晚餐省钱。如何将"幸运情侣"折扣定义为随机变量： - 当服务员拿来账单时，他还会拿来一副扑克牌中的四张A。他会洗牌并将它们面朝下放在桌上。然后情侣可以翻一张牌。 - 如果是黑桃A或梅花A，他们需要支付全额；但如果是红桃A，服务员将给他们20美元的"幸运情侣"折扣。 - 如果他们先翻到方块A（嘿——至少是红色的!），然后他们可以再翻剩下的一张牌，如果这次翻到红桃A，将获得10美元的折扣。

2.1.1 随机变量的定义

解

- 不难得到"幸运情侣"游戏的样本空间：

$$\Omega = \{[\overline{\Delta}], [\overline{\Delta}], [\overline{\Delta}], (\overline{\Delta}, [\overline{\Delta}]), (\overline{\Delta}, [\overline{\Delta}]), (\overline{\Delta}, [\overline{\Delta}])\}$$

- "幸运情侣"折扣可以定义为：

$$X(\omega) = \begin{cases} 20 & 若\ \omega = [\overline{\Delta}], \\ 10 & 若\ \omega = ([\overline{\Delta}], [\overline{\Delta}]), \\ 0 & 其他情况. \end{cases}$$

- 这是一个从样本空间中的结果到实数的映射。

- 定义了随机变量后，对随机事件的研究可以简化为对随机变量的研究。

- 对随机变量的研究本质上涉及检查随机变量可以取的所有可能值以及与每个值相关的概率，这称为**概率分布**。

2.1.1 随机变量的定义

- 有了概率分布，我们可以掌握随机事件的整体确定性，为进一步探索其潜在规律奠定基础。

- 根据随机变量可以取的值，可以将其分类为： - **离散型随机变量**：取有限个或可数个值，例如，访问餐厅的顾客数量。 - **连续型随机变量**：取连续值（不可数个值），例如，等公交车的时间。

- 描述这两类随机变量概率分布的异同： - 离散型随机变量可以使用**概率质量函数**描述。 - 连续型随机变量可以使用**概率密度函数**描述。 - 两种类型的随机变量都可以使用**累积分布函数**描述。

2.1.2 概率分布的描述

概率质量函数

设 X 为离散随机变量，$S = \{a_1, a_2, ...\}$ 为 X 的支撑集，即 X 可以取的值的集合。则 X 的概率质量函数是 $p : S \to [0, 1]$，定义为

$$p(a_i) = p_t = P(X = a_i), i = 1, 2, ...,$$

满足： - 非负性：$p(a_i) \geq 0, i = 1, 2, ...$; - 规范性：$\sum_i p(a_i) = 1$.

概率质量函数也可以用表格形式显示：

取值	a_1	a_2	...	a_i	...
概率	p_1	p_2	...	p_i	...

概率质量函数可以通过在 y 轴上绘制 $p(a_i)$ 对 x 轴上的 a_i 以图形形式显示。

2.1.2 概率分布的描述

例 2.1 （续）

求”幸运情侣”折扣的概率质量函数。

解

- 不难得到：

$$P(\mathbf{A}) = P(\mathbf{A}) = P(\mathbf{A}) = \frac{1}{4},$$

$$P(\mathbf{A}, \mathbf{A})) = P(\mathbf{A}, \mathbf{A})) = P(\mathbf{A}, \mathbf{A})) = \frac{1}{4} \times \frac{1}{3} = \frac{1}{12}.$$

- 那么：

$$P(X = 20) = P(\mathbf{A}) = \frac{1}{4}, \quad P(X = 10) = P(\mathbf{A}, \mathbf{A})) = \frac{1}{12}, \quad P(X = 0) = \frac{1}{4} + \frac{1}{4} + \frac{1}{12} + \frac{1}{12} = \frac{2}{3}.$$

- 用表格形式显示概率质量函数：

取值	0	10	20
概率	0.667	0.083	0.250

2.1.2 概率分布的描述

例 2.2

假设随机变量 X 的支撑集为 $\{0,1,2,...\}$，且 X 的概率质量函数由 $P(X=k)=c\lambda^k/k!$ 给出，其中 λ 是某个正常数。请用 λ 表示 c 的值。

解

- 由概率质量函数的规范性，有

$$1=\sum_{k=0}^{\infty}P(X=k)=c\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}$$

这意味着

$$1=ce^{\lambda}\Rightarrow c=e^{-\lambda}$$

2.1.2 概率分布的描述

连续随机变量与概率密度函数

如果存在一个非负函数 f，定义在所有 $x\in\mathbb{R}$ 上，满足对于任意 $-\infty<a\leq b<\infty$：

$$P(a\leq X\leq b)=\int_a^b f(x)dx$$

则称 X 为连续随机变量。函数 f 称为 X 的**概率密度函数**。

不难看出，概率密度函数与概率质量函数一样具有以下性质：

- **非负性**：$f(x)\geq 0, \forall x\in\mathbb{R}$;
- **规范性**：$\int_{-\infty}^{\infty}f(x)dx=1$.
- 注意，对于任意 $x\in\mathbb{R}$，有

$$P(X=x)=\int_x^x f(u)du=0.$$

- 因此，与概率质量函数不同，$f(x)$ 并不反映 X 取值为 x 的概率。

2.1.2 概率分布的描述

累积分布函数

对于随机变量 X，无论是离散型还是连续型，其累积分布函数定义为

$$F(x) = P(X \leq x), \forall x \in \mathbb{R}.$$

- 对于离散随机变量，其累积分布函数是阶梯函数，即

$$F(x) = \sum_{a_i \leq x} p(a_i).$$

- 对于连续随机变量，其累积分布函数是连续函数，即

$$F(x) = \int_{-\infty}^{x} f(u)du.$$

因此，

$$f(x) = F'(x).$$

- 累积分布函数是非递减且右连续的。
- 累积分布函数的最大值为

$$F(\infty) = \lim_{x \to \infty} F(x) = 1.$$

- 累积分布函数的最小值为

$$F(-\infty) = \lim_{x \to -\infty} F(x) = 0.$$

- 对于任意实数 $a < b$，

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

2.1.2 概率分布的描述

例 2.3

- 假设某种家用电器使用寿命（年）是一个随机变量，其概率密度函数为

$$f(x) = \begin{cases} ce^{-x/10}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

- 一件电器使用寿命在1到2年之间的概率是多少？

解

- 由概率密度函数的规范性，有

$$1=\int_{-\infty}^{\infty} f(x)dx = c\int_0^{\infty} e^{-x/10}dx \implies 1=10c \implies c=\frac{1}{10}$$

- 令 X 表示计算机使用寿命（年）的随机变量，则

$$P(1\le X\le 2)=\int_1^2 f(x)dx=\frac{1}{10}\int_1^2 e^{-x/10}dx=e^{-1/10}-e^{-1/5}\approx 0.086.$$

2.1.3 期望与方差

- 概率分布提供了对随机变量全面完整的描述，但有时我们可能只想从特定方面用几个数字来描述随机变量。 - 这些数字应该简单、清晰、独特且直观。 - 最常用的数字特征是数学期望和方差。

数学期望

随机变量 X 的数学期望记为 $E(X)$：

- 如果 X 是离散随机变量，其概率质量函数为 $p(x_k)=p_k=P(X=x_k), k=1,2,...,$ 给定 $\sum_{k=1}^{\infty}|x_k|p_k<\infty$，则

$$E(X)\triangleq\sum_{k=1}^{\infty}x_kP(X=x_k)=\sum_{k=1}^{\infty}x_kp_k.$$

- 如果 X 是连续随机变量，其概率密度函数为 $f(x)$，给定 $\int_{-\infty}^{\infty}|x|f(x)dx<\infty$，则

$$E(X)\triangleq\int_{-\infty}^{\infty}xf(x)dx.$$

也称为均值、期望、期望值。

2.1.3 期望与方差

- 简而言之，期望是随机变量所有可能值的加权平均。 - 期望代表我们可以从一系列随机事件中预期的平均结果或长期价值，在现实生活中有广泛应用。 - 在保险行业，期望可用于设定保费。 - 在金融领域，投资者使用期望来预测投资回报。

例 2.3 （续）

- 一家商店采用"先用后付"的方式销售该家用电器，付款金额根据电器的使用寿命 X 确定。

使用寿命（年）	$X \leq 1$	$1 < X \leq 2$	$2 < X \leq 3$	$X > 3$
付款金额	1500	2000	2500	3000

- 商店销售一件随机电器的期望回报是多少？

2.1.3 期望与方差

解

- X 的累积分布函数为

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du = \frac{1}{10}\int_0^x e^{-u/10} = 1 - e^{-x/10}.$$

- 令 Y 为商店销售一件随机电器的回报，则 Y 的概率质量函数为

$$\begin{aligned}
&P(Y = 1500) = P(X \leq 1) = F_X(1) \approx 0.095,\\
&P(Y = 2000) = P(1 < X \leq 2) = F_X(2) - F_X(1) \approx 0.086,\\
&P(Y = 2500) = P(2 < X \leq 3) = F_X(3) - F_X(2) \approx 0.078,\\
&P(Y = 3000) = P(X > 3) = 1 - F_X(3) \approx 0.741.
\end{aligned}$$

- Y 的概率质量函数总结于下表：

取值	1500	2000	2500	3000
概率	0.095	0.086	0.078	0.741

- 因此，Y 的期望计算如下：

$$E(Y) = \sum_{k=1}^{4} y_k P(Y = y_k) = 1500\times0.095+2000\times0.086+2500\times0.078+3000\times0.741 = 2732.5.$$

2.1.3 期望与方差

- 虽然期望代表平均结果，但我们也关心结果的稳定性/波动性。 - 方差描述了随机变量围绕其期望的波动性。

如果随机变量 X 满足 $E(X^2) < \infty$，则 $\mathrm{Var(X)} \triangleq E(X - E(X))^2$ 称为 X 的**方差**，且 $SD(X) \triangleq \sqrt{}$

- 具体来说，根据期望的定义： - 如果 X 是**离散随机变量**，其概率质量函数为 $p(x_k) = p_k = P(X = x_k), k = 1, 2, ...,$，则

$$\mathrm{Var}(X) = \sum_{k=1}^{\infty} [x_k - E(X)]^2 \cdot p_k$$

。 - 如果 X 是**连续随机变量**，其概率密度函数为 $f(x)$，则

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} [x - E(X)]^2 f(x) dx$$

。

2.1.3 期望与方差

例 2.4

- 假设两个品牌手表的时间偏差（秒）总结如下

偏差	-3	-2	-1	0	1	2	3
概率（品牌 A）	0.10	0.15	0.15	0.20	0.15	0.15	0.10
概率（品牌 B）	0.15	0.10	0.10	0.30	0.10	0.10	0.15

- 哪个品牌质量更好？

解

- 令 X 和 Y 分别为两个品牌手表的时间偏差，不难得到

$$E(X) = E(Y) = 0.$$

- 因此，从平均时间偏差的角度看，两个品牌质量相同。 - 然后我们考虑方差：

$$\mathrm{Var}(X) = \sum [x_k - 0]^2 \cdot p_k = 3^2 \times 0.1 + 2^2 \times 0.15 + \cdots + 3^2 \times 0.1 = 3.3.1$$

- 因此，从稳定性的角度看，品牌 A 质量更好。

2.1.3 期望与方差

期望和方差有一些基本性质： - 对于任意常数 a, b，$E(aX+b) = aE(X)+b$。- 对于任意常数 a, b，$\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$，因此 $\mathrm{SD}(aX+b) = |a|\mathrm{SD}(X)$。 - $\mathrm{Var}(X) = E(X^2) - [E(X)]^2$.

证明：不失一般性，以离散随机变量为例，其概率质量函数为 $p(x_k) = p_k = P(X = x_k)$。

令 $\mu \triangleq E(X)$，则

$$\mathrm{Var}(X) = E(X - E(X))^2 = \sum_{k=1}^{\infty}(x_k - \mu)^2 p_k$$

$$= \sum_{k=1}^{\infty}(x_k^2 - 2\mu x_k + \mu^2)p_k = \sum_{k=1}^{\infty} x_k^2 p_k - 2\mu\sum_{k=1}^{\infty} x_k p_k + \mu^2$$

$$= E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - [E(X)]^2.$$

Chapter 2 Random Variables and Distributions

- 2.1 引言 - 2.2 常见离散分布 - 2.3 常见连续分布 - 2.4 随机变量的变换

2.2 常见离散分布

■伯努利分布是最简单的离散分布。

伯努利试验与伯努利分布

• 如果一个随机试验只有两个可能结果，A 和 $\overline{A}$，则该试验称为伯努利试验。• 如果一个随机变量 X 只取值 0 和 1，且 $P(X=1)=p$，$P(X=0)=1-p$，则称 X 服从参数为 p 的伯努利分布，记为 $X \sim \mathrm{Bernoulli}(p)$。

雅各布·伯努利 (1655-1705)

伯努利(p) 的概率质量函数可以表示为

$$p(x) = p^x(1-p)^{1-x}, x = 0, 1.$$

$X \sim \mathrm{Bernoulli}(p)$ 的期望和方差：

$$E(X) = p, \quad \mathrm{Var}(X) = p(1-p).$$

JAC. BERNOUILLI, MATH.BE

■伯努利分布是许多经典概率分布的基础，例如二项分布、几何分布等。

2.2 常见离散分布

n 重伯努利试验与二项分布

- n 重伯努利试验是指独立重复一个伯努利试验 n 次的实验。注意：- ”独立”意味着每个伯努利试验的结果互不影响。 - ”重复”意味着每个伯努利试验中事件 A 的概率，即 $P(A)=p$，保持不变。

- 令 X 为一个随机变量，记录 n 重伯努利试验中事件 A 发生的次数，则称 X 服从参数为 n 和 p 的二项分布，记为

$$X \sim \text{Binomial}(n,p) \text{ (或 } \text{Bin}(n,p), B(n,p)).$$

- Binomial(n,p) 的概率质量函数可以推导为

$$p(x)=\binom{n}{x}p^x(1-p)^{n-x}=C_n^x p^x(1-p)^{n-x}, x=0,1,2,...,n.$$

—

二项分布 n=10, p=0.1

二项分布 n=10, p=0.5

2.2 常见离散分布

例 2.5

- 某工厂有80台同类型设备，每台设备独立运行，故障概率为0.01。 - 一个维修工一次只能修理一台设备。工厂正在考虑两种分配维修工的策略：- 分配4个维修工，每人负责维护20台设备。 - 分配3个维修工，他们共同负责维护所有80台设备。 - 请比较这两种策略在设备发生故障时无法及时维修的概率。

2.2 常见离散分布

解

- 首先考虑第一种策略，令 X_1, X_2, X_3, X_4 分别表示四个维修工负责的20台设备中同时发生故障的设备数量。则

$$X_i \sim \text{Binomial}(20, 0.01), i = 1, 2, 3, 4.$$

- 因此，一件设备无法及时维修的概率为

$$P\left(\bigcup_{i=1}^{4}\{X_i \geq 2\}\right) \geq P(X_1 \geq 2) = 1 - P(X_1 = 0) - P(X_1 = 1) = 1 - \binom{20}{0} \times 0.01^0 \times 0.99^{20} - \binom{20}{1} \times 0.01^1$$

- 然后考虑第二种策略，令 X 表示80台设备中同时发生故障的设备数量，则

$$X \sim \text{Binomial}(80, 0.01).$$

- 因此，一件设备无法及时维修的概率为

$$P(X \geq 4) = 1 - \sum_{i=0}^{3} P(X = i) = 1 - \sum_{i=0}^{3} \binom{80}{i} \times 0.01^i \times 0.99^{80-i} \approx 0.0087.$$

- 因此，第二种策略更高效。

2.2 常见离散分布

■ $X \sim \text{Binomial}(n, p)$ 的期望和方差可以推导为

$$E(X) = np, \quad \text{Var}(X) = np(1-p).$$

证明：由期望的定义，我们有：

$$E(X) = \sum_{k=0}^{n} k \cdot P(X = k) = \sum_{k=1}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} k \cdot \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

$$= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} = np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-k-1}$$

二项式展开

$$= np[(1-p) + p]^{n-1} = np.$$

方差也可以类似推导，此处不详细展示。

一旦我们引入随机变量之间的独立性，就有更简单的推导方法。

2.2 常见离散分布

几何分布

- 假设独立重复伯努利试验直到事件 A 发生，令 X 为记录所需试验次数的随机变量，则称 X 服从参数为 p 的几何分布，记为

$$X \sim \mathrm{Geometric}(p).$$

- $\mathrm{Geometric}(p)$ 的概率质量函数可以推导为

$$p(x) = p(1-p)^{x-1}, x = 1, 2, \ldots.$$

- $X \sim \mathrm{Geometric}(p)$ 的期望和方差：

$$E(X) = \frac{1}{p}, \quad \mathrm{Var}(X) = \frac{1-p}{p^2}.$$

- 几何分布是唯一具有无记忆性的离散分布，即对于 $X \sim \mathrm{Geometric}(p)$ 和任意正整数 m, n，有

$$P(X > m + n | X > m) = P(X > n).$$

例 2.6

- 赌场中一个赌徒在玩"大小"游戏时押注"大"，但赌桌已经连续十轮开出"小"。

- 赌徒认为既然"小"已经出现了这么多次，再出现"小"的概率非常低。

- 他考虑将他所有的钱押在"大"上以试图赢回他的损失。你怎么看？

赌徒谬误

2.2 常见离散分布

- 一些概率分布可以直接从数据生成机制本身推导出来，例如伯努利分布、二项分布、几何分布。 - 其他一些概率分布是先通过数学推导或从数据中总结出来，后来发现可以描述许多现实世界的模式。 - 泊松分布是第二种类型的一个例子，它由西莫恩·德尼·泊松在研究某个国家的错误定罪数量时首次引入。

泊松分布

- 令 X 为一个支撑集为 $\{0,1,2,...\}$ 的离散随机变量，如果其概率质量函数为

$$p(x) = P(X = x) = \frac{\lambda^x}{x!}e^{-\lambda}, x = 0, 1, 2, ...,$$

其中 $\lambda > 0$ 是一个常数，则称 X 服从参数为 λ 的**泊松分布**，记为 $X \sim \text{Poisson}(\lambda)$。

西莫恩·德尼·泊松
(1781-1840)

2.2 常见离散分布

虽然泊松分布最初是为了解决法律和刑事司法问题而引入，但后来被广泛应用于各个领域。例子： - 1860年，一位美国天文学家将泊松分布拟合到在某个空间区域内发现的恒星数量。 - 1898年，一位俄罗斯经济学家表明，普鲁士军队中士兵被马踢死的频率可以用泊松分布很好地建模。

泊松分布用于描述在固定时间/空间间隔内事件发生的次数，如果事件以恒定速率独立发生。例子： - 每天进入商店的顾客数量。 - 每年在一个十字路口发生的交通事故数量。 - 一页书上的打字错误数量。 - 一场世界杯足球赛中的进球数。 - 单位长度DNA链上的突变数。

来源：人类指导位置与实际，n-2019/6，实验1/2015

年龄	认证（米）
500	300
400	200
450	100
600	150
700	200
800	250
900	200
1000	0

编号	数字
500	300
400	200
450	100
600	150
700	200
800	250
900	200
1000	0

编号	数字
500	300
400	200
450	100
600	150
700	200
800	250
900	200
1000	0

2.2 常见离散分布

■ $X \sim \text{Poisson}(\lambda)$ 的期望和方差可以推导为

$$E(X) = \lambda, \quad \text{Var}(X) = \lambda.$$

证明：由期望的定义，我们有：

$$E(X) = \sum_{k=0}^{\infty} k \cdot P(X = k) = \sum_{k=1}^{\infty} k \cdot \frac{\lambda^k}{k!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda.$$

$$E(X^2) = \sum_{k=0}^{\infty} k^2 \cdot P(X = k) = \sum_{k=1}^{\infty} k^2 \cdot \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \left[\sum_{k=1}^{\infty} \frac{(k-1)\lambda^k}{(k-1)!} + \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} \right] = \lambda^2 + \lambda.$$

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = \lambda.$$

■参数 λ 也称为泊松分布的强度。

2.2 常见离散分布

例 2.7

- 某市政局正在考虑是否在一条道路上部署一辆救援车辆，以尽快清除事故。 - 记录显示，早高峰期间平均事故数量为5。 - 如果早高峰期间发生超过5起事故的概率低于30%，市政局将不会在道路上部署救援车辆。 - 基于此信息，市政局应该提供车辆吗？

2.2 常见离散分布

解

- 令 X 为随机一天早高峰期间的事故数量，则 $X \sim \text{Poisson}(5)$。 - 目标是计算 $P(X > 5) = 1 - P(X \leq 5)$。由于：

$$P(X = 0) = \frac{5^0}{0!} e^{-5} = 0.00674, \quad P(X = 1) = \frac{5^1}{1!} e^{-5} = 5 \times P(X = 0) = 0.03369$$

$$P(X = 2) = \frac{5}{2} \times P(X = 1) = 0.08422, \quad P(X = 3) = \frac{5}{3} \times P(X = 2) = 0.14037,$$

$$P(X = 4) = \frac{5}{4} \times P(X = 3) = 0.17547, \quad P(X = 5) = \frac{5}{5} \times P(X = 4) = 0.17547.$$

- 所以，早高峰期间道路上发生超过5起事故的概率为

$$P(X>5)=1-P(X\le 5)=1-0.00674-0.03369-\cdots-0.17547=0.38403>30\%.$$

- 因此，市政局应该在道路上提供救援车辆。

2.2 常见离散分布

- 泊松分布可以作为二项分布的极限情况推导出来。

证明： 考虑单位时间区间 $[0,1]$ 内的事件数，并将其分成 n 个子区间：

$$\left[0,\frac{1}{n}\right),\left(\frac{1}{n},\frac{2}{n}\right),\ldots,\left(\frac{i-1}{n},\frac{i}{n}\right),\ldots,\left(\frac{n-1}{n},1\right).$$

做了几个假设：

1. 令 n 很大，使得每个子区间非常短，以至于不可能在任何一个子区间内发生两个或更多事件（即，要么没有事件发生，要么恰好发生一个事件）。

2. 事件发生的概率与子区间的长度成正比，即 λ/n。

3. 一个子区间中事件是否发生与其他子区间独立。

令 X 为 $[0,1]$ 内的事件数，由上述假设，我们有 $X\sim \text{Binomial}\left(n,\frac{\lambda}{n}\right)$。所以

$$P(X=x)=\binom{n}{x}\left(\frac{\lambda}{n}\right)^{x}\left(1-\frac{\lambda}{n}\right)^{n-x}.$$

令 $n\to\infty$，不难得到

$$\lim_{n\to\infty}P(X=x)=\lim_{n\to\infty}\binom{n}{x}\left(\frac{\lambda}{n}\right)^{x}\left(1-\frac{\lambda}{n}\right)^{n-x}=\frac{\lambda^x}{x!}e^{-\lambda}.$$

泊松定理

2.2 常见离散分布

该定理表明，对于大的 n 和小的 p 的 n 重伯努利试验，二项分布 $\text{Binomial}(n,p)$ 可以用泊松分布 $\text{Poisson}(np)$ 近似。

n 应该多大，p 应该多小？

一个经验法则是当 $n>100$ 且 $p<0.05$ 时，泊松分布可以很好地近似二项分布。

2.2 常见离散分布

例 2.8

- 一家保险公司推出了一种人寿保险政策，每个参与者需要在1月1日支付12美元的保费。 - 如果参与者在一年内死亡，他/她的家人可以从保险公司获得2,000美元的赔偿。 - 假设有2,500人参加这种保险，每个人在一年内的死亡概率为0.002。 - 保险公司从这种人寿保险政策中获得的利润不低于20,000美元的概率是多少？

2.2 常见离散分布

解

- 对于保险公司从此人寿保险政策中获得的利润不低于20,000美元，最大死亡人数为

$$\frac{2500 \times 12 - 20000}{2000} = 5.$$

- 令 X 为2500名参与者中的死亡人数，则 $X \sim \text{Binomial}(2500, 0.002)$。我们想计算

$$P(X \leq 5) = \sum_{k=0}^{5} \binom{2500}{k} 0.002^k 0.998^{2500-k}.$$

- 由于二项分布中 $n = 2500 > 100$ 且 $p = 0.002 < 0.05$，我们应用泊松定理，即用 $\text{Poisson}(2500 \times 0.002) = \text{Poisson}(5)$ 来近似 $\text{Binomial}(2500, 0.002)$：

$$P(X \leq 5) \approx \sum_{k=0}^{5} \frac{5^k}{k!} e^{-5} \approx 0.616.$$

2.2 常见离散分布

■总结一下，我们介绍了以下离散分布：

分布	概率质量函数	期望	方差
Bernoulli(p)	$p(x)=p^x(1-p)^{1-x}, x=0,1$	p	$p(1-p)$
Binomial(n,p)	$p(x)=\binom{n}{x}p^x(1-p)^{n-x}, x=0,1,2,...,n$	np	$np(1-p)$
Geometric(p)	$p(x)=p(1-p)^{x-1}, x=1,2,...$	$1/p$	$(1-p)/p^2$
Poisson(λ)	$p(x)=P(X=x)=\frac{\lambda^x}{x!}e^{-\lambda}, x=0,1,2,...$	λ	λ

■这里没有涵盖许多其他离散分布，例如：■超几何分布：描述在 n 重伯努利试验的前 m 次试验中成功的次数。■负二项分布：几何分布的推广，描述直到预定义的成功次数 r 发生所需的伯努利试验次数。

第2章随机变量与分布

- 2.1 引言 - 2.2 常见离散分布 - 2.3 常见连续分布 - 2.4 随机变量的变换

2.3 常见连续分布

- 均匀分布是最简单的连续分布。

均匀分布

- 如果随机变量 X 的概率密度函数为

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

- 则称 X 服从 (a,b) 上的均匀分布，记为 $X\sim \mathrm{Uniform}(a,b)$ 或简记为 $X\sim \mathrm{U}(a,b)$。

- $X\sim \mathrm{U}(a,b)$ 的累积分布函数为

$$F(x)=P(X\leq x)=\int_{-\infty}^{x} f(u)du=\begin{cases}0, & \text{若 } x\leq a\\ \frac{x-a}{b-a}, & \text{若 } a<x<b,\\ 1, & \text{若 } x\geq b\end{cases}$$

2.3 常见连续分布

- 对于 $\forall (c, c+L) \subset (a,b)$，$P(c < X < c+L) = L/(b-a)$，这仅取决于区间的长度而不取决于其位置，表明了一种"等可能性"。 - $X \sim U(a,b)$ 的期望和方差可以推导为

$$E(X) = \frac{a+b}{2}, \quad \mathrm{Var}(X) = \frac{(b-a)^2}{12}.$$

证明：

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2},$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_a^b \frac{x^2}{b-a} dx = \frac{a^2+ab+b^2}{3},$$

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \frac{a^2+ab+b^2}{3} - \frac{a^2+2ab+b^2}{4} = \frac{a^2-2ab+b^2}{12} = \frac{(b-a)^2}{12}.$$

2.3 常见连续分布

例 2.9

- 一根长度为 $2l$ 的木棍被随机切成两段。 - 这两段与另一根长度为 l 的木棍能够组成一个三角形的概率是多少？

解

- 令两段的长度分别为 X 和 $2l - X$，则 $X \sim U(0, 2l)$。 - 要使这两段与另一根长度为 l 的木棍组成三角形，我们需要

$$\begin{cases} X + l > 2l - X \\ 2l - X + l > X \end{cases} \iff \frac{l}{2} < X < \frac{3l}{2}.$$

- 因此，概率为

$$P\left(\frac{l}{2} < X < \frac{3l}{2}\right) = \int_{l/2}^{3l/2} \frac{1}{2l} dx = \frac{1}{2}.$$

2.3 常见连续分布

下一个常见的连续分布是指数分布。

指数分布

如果随机变量 X 的概率密度函数为

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{若 } x \geq 0 \\ 0, & \text{否则} \end{cases}$$

则称 X 服从参数为 λ 的指数分布，记为 $X \sim \mathrm{Exp}(\lambda)$。

$X \sim \mathrm{Exp}(\lambda)$ 的累积分布函数为

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du = \begin{cases} 1 - e^{-\lambda x}, & \text{若 } x \geq 0 \\ 0, & \text{否则} \end{cases}$$

在实践中，指数分布经常作为直到某个特定事件发生的时间长度的分布出现。

2.3 常见连续分布

- 你是否注意到指数分布和泊松分布使用相同的符号 λ 作为它们的参数。它们之间有什么关系吗？

- 事实上，指数分布可以用来描述泊松过程中事件之间时间间隔的分布。

- 泊松过程可以简单理解为一个在时间轴上以恒定速率独立发生随机事件的过程。

- 单位时间间隔内发生的事件数服从 $\mathrm{Poisson}(\lambda)$，则在 $[0, t]$ 内发生的事件数服从 $\mathrm{Poisson}(\lambda t)$。

- 令 X 为直到第一个事件发生的时间，则

$$P(X \leq t) = 1 - P(X > t) = 1 - P(\text{在 } [0, t] \text{ 内没有事件发生}) = 1 - \frac{(\lambda t)^0}{0!} e^{-\lambda t} = 1 - e^{-\lambda t},$$

这意味着 $X \sim \mathrm{Exp}(\lambda)$。

- 类似地，我们可以证明事件之间的时间间隔独立地服从 $\mathrm{Exp}(\lambda)$。

2.3 常见连续分布

- $X \sim \mathrm{Exp}(\lambda)$ 的期望和方差可以推导为

$$E(X) = \frac{1}{\lambda}, \quad \mathrm{Var}(X) = \frac{1}{\lambda^2}.$$

证明：

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_0^{\infty} x\lambda e^{-\lambda x}dx = -\int_0^{\infty} xd(e^{-\lambda x}) = -xe^{-\lambda x}\Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x}dx = -\frac{1}{\lambda}e^{-\lambda x}\Big|_0^{\infty} =$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^{\infty} x^2\lambda e^{-\lambda x}dx = -\int_0^{\infty} x^2 d(e^{-\lambda x}) = 2\int_0^{\infty} xe^{-\lambda x}dx = \frac{2}{\lambda^2},$$

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

- 对于强度为 λ 的泊松过程，即单位时间间隔内期望事件数为 λ，事件之间的期望时间间隔为 $1/\lambda$ 是很自然的。

2.3 常见连续分布

例 2.10

- 2023年9月7日香港发生了一场特大暴雨。 - 新闻报道称这是500年一遇的暴雨。 - 500年前，是明朝嘉靖二年。 - 这是否意味着这是自明朝以来最大的雨？

报纸：https://www.hk01.com/article/939036?utm_source=01articlecopy&utm_medium=referral

2.3 常见连续分布

答案

- 事实上，天文台将这场雨描述为”100年一遇的雨”，而渠务署则称其为”500年一遇的雨”。 - 天文台和渠务署有不同的计算方法。 - 渠务署的计算基于防洪标准。 - 天文台的”XXX年一遇”是描述相同规模降雨事件的

泊松过程的事件率。 - 因此，两次"100年一遇的雨"之间的时间间隔服从Exp(1/100)。

两次"100年一遇的雨"之间的期望时间间隔是100年，但这并不意味着我们需要正好等待100年才能遇到一次这样的雨。

其实新闻报道中常提及的「几多年一遇」是表示概率。天文台与渠务署的计算公式亦不一样。简单而言，天文台的「多少年一遇」，是基于以往录得的雨量及出现频率作为统计基础作出推算，是数据上的结论；而渠务署所指的「多少年一遇」，其实是防洪标准，一般市区排水干渠系统足以应付重现期为200年一遇的暴雨，而今回「500年一遇」的暴雨侵袭，最后便导致多区出现水浸。

2.3 常见连续分布

■指数分布是唯一具有**无记忆性**的连续分布，即若 $X \sim \mathrm{Exp}(\lambda)$，则对任意 $s,t>0$：

$$P(X>s+t|X>s)=P(X>t).$$

证明：对任意 $s,t>0$，由于 $\{X>s+t\}\subset\{X>s\}$，有

$$P(\{X>s+t\}\cap\{X>s\})=P(X>s+t)=1-F(s+t)=e^{-\lambda(s+t)}.$$

由条件概率定义：

$$P(X>s+t|X>s)=\frac{P(\{X>s+t\}\cap\{X>s\})}{P(X>s)}=\frac{e^{-\lambda(s+t)}}{e^{-\lambda s}}=e^{-\lambda t}=P(X>t).$$

■无记忆性大大简化了分析，但也使得指数分布不适用于许多现实应用。

■思考：回忆几何分布是唯一具有无记忆性的离散分布，几何分布与指数分布之间是否存在某种联系？

2.3 常见连续分布

- 正态分布是最重要的分布，毫无例外。

正态分布

- 若随机变量 X 的概率密度函数为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},-\infty<x<\infty,$$

- 则称 X 服从参数为 μ 和 $\sigma^2(\sigma>0)$ 的**正态分布**，记为 $X\sim N(\mu,\sigma^2)$。

- 特别地，$N(0,1)$ 称为**标准正态分布**，其概率密度函数为

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

- $X\sim N(0,1)$ 的累积分布函数没有显式表达式，但由于使用频繁，记为

$$\Phi(x):\Phi(x)=P(X\le x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du.$$

$N(\mu,\sigma^2)$ 的概率密度函数是一条优美的钟形曲线，关于参数 μ 对称。

$$68\%$$
$$95\%$$
$$99.7\%$$
$$\mu - 3\sigma$$
$$\mu - 2\sigma$$
$$\mu - \sigma$$
$$\mu$$
$$\mu + \sigma$$
$$\mu + 2\sigma$$
$$\mu + 3\sigma$$

2.3 常见连续分布

$$\sigma = 0.5$$
$$\sigma = 1$$
$$\sigma = 1.5$$

证明：

对于 $Z = \frac{X-\mu}{\sigma}$，考虑其累积分布函数：

$$\begin{aligned} F_z(x) &= P(Z \leq x) = P\left(\frac{X-\mu}{\sigma} \leq x\right) \\ &= P(X \leq \sigma x + \mu) = F_X(\sigma x + \mu). \end{aligned}$$

通过求导，Z 的概率密度函数为

$$\begin{aligned} f_z(x) &= \frac{dF_X(\sigma x + \mu)}{dx} = \sigma f_X(\sigma x + \mu) \\ &= \frac{\sigma}{\sqrt{2\pi}\sigma} e^{-\frac{(\sigma x + \mu - \mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}} e^{\frac{x^2}{2}}, \end{aligned}$$

这表明 $Z \sim N(0,1)$。

2.3 常见连续分布

- $X \sim N(\mu, \sigma^2)$ 的期望和方差可推导为

$$E(X) = \mu, \quad \mathrm{Var}(X) = \sigma^2.$$

证明： 令 $Z = \frac{X-\mu}{\sigma}$，只需证明 $E(Z) = 0, \mathrm{Var}(Z) = 1$。

$$E(Z) = \int_{-\infty}^{\infty} x\phi(x)dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} xe^{-\frac{x^2}{2}}dx = -\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \Big|_{-\infty}^{\infty} = 0.$$

$$E(Z^2) = \int_{-\infty}^{\infty} x^2\phi(x)dx = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} xd(e^{-x^2/2}) = -\frac{1}{\sqrt{2\pi}} \left(xe^{-\frac{x^2}{2}} \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}dx \right)$$

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

由概率密度函数的归一性

$$\mathrm{Var}(Z) = E(Z^2) - [E(Z)]^2 = 1.$$

2.3 常见连续分布

- 正态分布也称为**高斯分布**。

图表

- 许多人认为是高斯发现了正态分布。 - 虽然高斯在历史上对确立正态分布的重要性起到了关键作用，但他并非第一个提出该分布的人。 - 法国数学家庞加莱建议使用“正态分布”这一中性术语。

2.3 常见连续分布

- 正态分布在概率论与统计学中起着至关重要的作用，主要归功于**中心极限定理**，这将在下一章介绍。 - 中心极限定理指出，随机变量的和/平均值通常服从正态分布。 - 因此，各种波动和测量误差呈现正态分布。 - 此

外，正态分布常被发现是体重、身高、智力、温度、污染水平、学生成绩等的良好模型。

诺贝尔奖获得者年龄

- 相对频率 - 噪声 - 时间 - 范围 - 数值 - 拟合SAT分数数据的正态曲线

学生人数 - 0 - 20 - 40 - 60 - 80 - 100 - 120

性别：女性 - 15.0% - 12.5% - 10.0% - 7.5% - 5.0% - 2.5%

诺贝尔奖获得者年龄

2.3 常见连续分布

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

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

$$1 - \Phi(x)$$

$$-x$$

若 $X \sim N(\mu, \sigma^2)$ 且 $Z = (X-\mu)/\sigma$，则 $Z \sim N(0,1)$，因此，此外，由 $N(0,1)$ 概率密度函数的对称性：由于 $\Phi(x)$ 没有解析表达式，我们需要在概率表中查找其值。例如，$\Phi(1.96) = 0.975$。

2.3 常见连续分布

例 2.11

- 亲子诉讼案中的专家证人作证称，人类妊娠期长度（以天计）近似服从参数为 $\mu = 270$ 和 $\sigma^2 = 100$ 的正态分布。 - 诉讼中的被告能够证明自己在孩子出生前 290 天至 240 天期间不在国内。 - 如果被告确实是孩子的父亲，那么母亲经历如证词所述的极长或极短妊娠期的概率是多少？

解

- 设母亲的妊娠期长度为 X，则 $X \sim N(270, 100)$。 - 需计算的概率为 $P((X > 290) \cup \{X < 240\})$。考虑

$$P(X > 290) = P\left(\frac{X - 270}{10} > \frac{290 - 270}{10}\right) = P(Z > 2) = 1 - \Phi(2) = 1 - 0.9772 = 0.0228.$$

$$P(X < 240) = P\left(\frac{X - 270}{10} < \frac{240 - 270}{10}\right) = P(Z < -3) = 1 - \Phi(3) = 1 - 0.9987 = 0.0013.$$

- 因此，概率为 $0.0228 + 0.0013 = 0.0241$。

2.3 常见连续分布

例 2.12

- 某巴士制造商正在设计一款巴士。在确定车门高度时，必须确保其高度适中，使得 99% 的男性乘客无需弯腰即可通过。 - 假设所有男性身高（单位：cm）服从正态分布 $N(170, 36)$，为满足此要求，车门的最小高度应为多少？

解

- 设 X 表示随机选择的男性身高，h 为巴士车门高度。 - 则要求可表示为 $P(X \leq h) \geq 0.99$。 - 反查概率表，对于 $Z \sim N(0, 1)$，有 $P(Z \leq 2.33) = 0.9901 > 0.99$。所以

$$P(X \leq h) = P\left(\frac{X - 170}{6} \leq \frac{h - 170}{6}\right) = P\left(Z \leq \frac{h - 170}{6}\right) \geq 0.99.$$

$$\Rightarrow \frac{h - 170}{6} \geq 2.33 \implies h \geq 170 + 13.98 \approx 184\,\mathrm{cm}.$$

2.3 常见连续分布

- 总结一下，我们介绍了以下连续分布：

分布	概率密度函数	期望	方差
均匀分布 (a, b)	$f(x) = \begin{cases} \frac{1}{b-a}, & \text{若 } a < x < b \\ 0, & \text{否则} \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
指数分布 (λ)	$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{若 } x \geq 0 \\ 0, & \text{否则} \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
正态分布 (μ, σ^2)	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty$	μ	σ^2

- 还有许多其他连续分布未在此涵盖，例如： - **贝塔分布**：通常用于描述支撑集为 [0, 1] 的随机变量的分布，广泛应用于贝叶斯统计。 - **伽马分布**：指数分布的推广，广泛用于多阶段方案的总时间。

2.3 常见连续分布

问题：概率分布的本质是什么？通常，我们也可以根据收集的数据绘制分布，例如使用直方图。这与前面提到的概率分布有何关系？根据收集数据绘制的分布只是数据可视化的一种形式。概率分布不对应于特定的数据集，而是一个模型。我们可以使用这个模型来描述数据生成的机制或总结数据背后的模式。

2.3 常见连续分布

- 概率分布本质上是模型： - 我们通常为数据假设一个概率分布模型，然后用收集的数据验证该假设。 - 很难完全匹配，但一定范围的偏差是可以接受的。 - “所有模型都是错误的，但有些是有用的”。 - 事实上，模型本身并无对错；错误在于为特定问题选择了错误的模型。 - 概率分布就像一个工具箱，每个分布都是工具箱中的一个特定工具。 - 当面临问题时，我们需要从工具箱中找到合适的工具来解决它。

所有模型都是错误的但有些是有用的
嘿，那很刻薄！
我的朋友是模特！
等等，乔治不是那个意思！
统计学家乔治·博克斯

他谈论的是数学模型
方程和近似现实的东西
好吧，我们来看一个例子

第2章随机变量与分布

- **2.1** 引言 - **2.2** 常见离散分布 - **2.3** 常见连续分布 - **2.4** 随机变量的变换

2.4 随机变量的变换

- 有时，我们可能知道一个随机变量 X 的分布，并希望推导该随机变量的某个函数的分布，即 $Y = g(X)$。例如： - 假设您投资了 ℵ1000 到一个连续复利账户，利率为 R - R 是一个具有概率密度函数 $f(r)$ 的连续随机变量的实现。 - 一年后账户中的金额，即 $A = 1000e^R$，是如何分布的？ - 一位科学家测量圆的半径，由于测量误差，结果是一个随机变量（记为 R）。 - R 是一个具有概率密度函数 $f(r)$ 的连续随机变量的实现。 - 计算出的圆面积，即 $A = \pi R^2$，的分布是什么？

2.4 随机变量的变换

- 对于具有概率质量函数 $p_X(x)$ 的离散随机变量 X，确定 $Y = g(X)$ 的概率质量函数并不困难：

$$p_Y(y) = P(Y = y) = \sum_{x:g(x)=y} p_X(x),$$

该公式同时考虑了 g 是一一映射和非一一映射的情况。

例 2.13

- 随机变量 X 的概率质量函数如下，求 $Y = (X - 1)^2$ 的概率质量函数。

x	-1	0	1	2
$p_X(x)$	0.2	0.3	0.1	0.4

解

- 对 X 的每个可能取值 x，计算 $y = (x-1)^2$： - 合并相同 y 值的概率：

y	4	1	0	1
$p_Y(y)$	0.2	0.3	0.1	0.4

y	0	1	4
$p_Y(y)$	0.1	0.7	0.2

2.4 随机变量的变换

■对于具有概率密度函数 $f_X(x)$ 的连续随机变量 X，若 $Y = g(X)$ 是离散随机变量，则 Y 的概率质量函数为：

$$p_Y(y) = P(Y = y) = \int_{x:g(x)=y} f_X(x)dx.$$

连续到离散

例 2.14

■假设随机选择的人的智商测试分数为 $X \sim N(100, 225)$。

■定义随机变量 Y 为

$$Y = \begin{cases} 1, & \text{若 } X \le 85 \\ 2, & \text{若 } 85 < X \le 115 \\ 3, & \text{若 } X > 115 \end{cases}$$

■ Y 的概率质量函数是什么？

解

$$P(Y = 1) = P(X \le 85) = P\left(\frac{X - 100}{15} \le \frac{85 - 100}{15}\right)$$
$$= P(Z \le -1) = 1 - P(Z \le 1) = 0.1587.$$

■因此，Y 的概率质量函数为

y	1	2	3
$p_Y(y)$	0.1587	0.6826	0.1587

2.4 随机变量的变换

- 对于具有概率密度函数 $f_X(x)$ 的连续随机变量 X，若 $Y = g(X)$ 也是连续随机变量，则推导 Y 的概率密度函数稍微复杂一些。

- 若 $g(x)$ 在 X 的支撑集上是严格单调函数，并且具有连续可导的反函数 $h(y) = g^{-1}(y)$，则 Y 的概率密度函数为

$$f_Y(y) = \begin{cases} |h'(y)| \cdot f_X(h(y)), & \text{在 } h(y) \text{ 有定义处} \\ 0, & \text{否则} \end{cases}$$

连续到连续

证明：考虑 Y 的累积分布函数：$F_Y(y) = P(Y \le y) = P(g(X) \le y)$。

若 $g(x)$ 是严格递增函数，则 $h'(y) > 0$ 且：

$$F_Y(y) = P(X \le g^{-1}(y)) = F_X(h(y)) \Rightarrow f_Y(y) = F_Y'(y) = h'(y) f_X(h(y)).$$

若 $g(x)$ 是严格递减函数，则 $h'(y) < 0$ 且：

$$F_Y(y) = P(X \ge g^{-1}(y)) = 1 - F_X(h(y)) \Rightarrow f_Y(y) = F_Y'(y) = -h'(y) f_X(h(y)).$$

将这两种情况合并，Y 的概率密度函数即为上面红色公式给出的形式。

2.4 随机变量的变换

例 2.15

- 考虑通过网络传输文件所需的时间取决于网络速度 X，由于流量和其他条件的变化，X 是变化的，且 $X \sim$ 均匀$[2, 4]$（单位：Mbps）。 - 令 Y 表示传输一个 100Mb 文件所需的时间，请推导 Y 的概率密度函数。

解

- 根据所述场景，我们有：

$$f_X(x)=\begin{cases}1/2, & 若\ 2\le x\le 4\\ 0, & 否则\end{cases}$$

且

$$Y=g(X)=\frac{100}{X}.$$

- $g(x)=100/x$ 在 $[2,4]$ 上是严格递减函数，其反函数为 $h(y)=100/y$。所以

$$|h'(y)|\cdot f_X(h(y))=\frac{100}{y^2}\cdot\frac{1}{2}=\frac{50}{y^2}.$$

- Y 的概率密度函数为

$$f_Y(y)=\begin{cases}50/y^2, & 若\ 25\le y\le 50\\ 0, & 否则\end{cases}$$

2.4 随机变量的变换

- 随机变量变换的一个著名应用基于以下结果：- 若连续随机变量 X 的累积分布函数为 $F(x)$ 且其反函数 $F^{-1}(x)$ 存在。定义随机变量 $Y=F(X)$，则 $Y\sim$ 均匀$[0,1]$。 - 另一方面，若 $F(x)$ 是某个随机变量的累积分布函数且其反函数 $F^{-1}(x)$ 存在，令 $U\sim$ 均匀$[0,1]$，则对于 $X=F^{-1}(U)$，我们有 $X\sim F(x)$，即 X 的累积分布函数为 $F(x)$。

证明：考虑 Y 的累积分布函数：$F_Y(y)=P(Y\le y)=P(F(X)\le y)$。由于 $F(x)$ 是非递减函数且 $F^{-1}(x)$ 存在，则对任意 $y\in[0,1]$：

$$F_Y(y)=P(X\le F^{-1}(y))=F(F^{-1}(y))=y\implies f_Y(y)=F'_Y(y)=1.$$

这表明 $Y\sim$ 均匀$[0,1]$。第二个结果的证明类似，此处省略。

- 第二个结果可用于随机数抽样，称为**逆变换采样**。 - 这是一种从复杂分布生成随机样本的广泛应用技术。

2.4 随机变量的变换

- 当 $g(x)$ 在 X 的支撑集上不是严格单调函数时，如何推导 $Y=g(X)$ 的概率密度函数？

例 2.16

- 假设随机变量 $X \sim N(0,1)$，$Y = X^2$ 的概率密度函数是什么？

解

- 由于 $X \sim N(0,1)$，其概率密度函数为 $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, -\infty < x < \infty$。

- 虽然 $g(x) = x^2$ 不是单调函数，但它在 $(0,\infty)$ 上严格递增，在 $(-\infty,0)$ 上严格递减。对任意 $y \in (0,\infty)$，有

$$F_Y(y) = P(Y \le y) = P(X^2 \le y) = P(-\sqrt{y} \le X \le \sqrt{y}) = \Phi(\sqrt{y}) - \Phi(-\sqrt{y})$$

$$\Rightarrow F'_Y(y) = \phi(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} - \phi(-\sqrt{y}) \cdot \left(-\frac{1}{2\sqrt{y}}\right) = \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}} \Rightarrow f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}}, & y > 0 \\ 0, & \text{否则} \end{cases}$$

2.4 随机变量的变换

例 2.15 （续）

- 比较概率 $P(25 \le Y \le 30)$ 和 $P(45 \le Y \le 50)$。 - 计算 $E(Y)$，即传输 100Mb 文件所需时间的期望值。

解

- 基于 Y 的概率密度函数，不难得到

$$P(25 \le Y \le 30) = \int_{25}^{30} \frac{50}{y^2} dy = -\frac{50}{y}\bigg|_{25}^{30} = \frac{50}{25} - \frac{50}{30} \approx 0.333, \quad P(45 \le Y \le 50) = \frac{50}{45} - \frac{50}{50} \approx 0.111.$$

- 根据期望的定义，我们有

$$E(Y) = \int_{25}^{50} y \cdot \frac{50}{y^2} dy = 50 \ln y \bigg|_{25}^{50} = 50 \ln 50 - 50 \ln 25 \approx 34.657.$$

- $E\left(\frac{100}{X}\right)$ 与 $\frac{100}{E(X)}$ 之间有什么关系？ - 一般来说，我们是否有规则描述 $E(g(X))$ 与 $g(E(X))$ 之间的关系？

2.4 随机变量的变换

实际上，要计算 $Y = g(X)$ 的期望，无需先推导 Y 的概率质量函数/概率密度函数，我们可以直接使用 X 的概率质量函数/概率密度函数：

- 若 X 是离散随机变量，其概率质量函数为 $P(X = x_k) = p_k, k = 1, 2, \ldots$，且给定 $\sum_{k=1}^{\infty} |g(x_k)|p_k < \infty$，则

$$E(Y) = E(g(X)) = \sum_{k=1}^{\infty} g(x_k)p_k.$$

- 若 X 是连续随机变量，其概率密度函数为 $f(x)$，且给定 $\int_{-\infty}^{\infty} |g(x)|f(x)dx < \infty$，则

$$E(Y) = E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx.$$

例 2.15 （续）的解

- Y 的期望也可以直接使用 $f_X(x)$ 得到：

$$E(Y) = \int_2^4 \frac{100}{x} \cdot \frac{1}{2}dx = 50\ln x\Big|_2^4 = 50\ln 4 - 50\ln 2 \approx 34.657.$$

2.4 随机变量的变换

- 现在我们回到这个问题：一般来说，我们是否有规则描述 $E(g(X))$ 与 $g(E(X))$ 之间的关系？ - 对于某些类型的函数，我们确实有这样的规则。 - 例如，若 $g(x) = ax + b$，则我们总有 $E(g(X)) = g(E(X))$。 - 除了线性函数，还有其他情况吗？

凸函数与凹函数

- 函数 $g : S \to \mathbb{R}$ 被称为**凸函数**，如果对任意 $t \in [0, 1]$ 和 $x, y \in S$，有

$$g(tx + (1 - t)y) \leq tg(x) + (1 - t)g(y).$$

- 函数 $g : S \to \mathbb{R}$ 被称为**凹函数**，如果对任意 $t \in [0, 1]$ 和 $x, y \in S$，有

$$g(tx + (1 - t)y) \geq tg(x) + (1 - t)g(y).$$

$g(\cdot)$ 凸

$tg(x) + (1 - t)g(y)$

$$g(tx+(1-t)y)$$

$$x \quad y$$

$$tx+(1-t)y$$

$$g(\cdot) \text{ 凹}$$

$$x \quad y$$

2.4 随机变量的变换

琴生不等式

设 X 为一个随机变量，则 - 对任意**凸**函数 g，

$$E(g(X)) \geq g(E(X)).$$

- 对任意**凹**函数 g，

$$E(g(X)) \leq g(E(X)).$$

凸

$$x$$

$$y$$

$$x$$

$$y$$

凹

$$x$$

$$y$$

$$x$$

$$y$$

根据琴生不等式，我们有以下结果： - $E(|X|) \geq |E(X)|$ $(g(x)=|x|)$; - $E(X^2) \geq (E(X))^2$ $(g(x)=x^2)$; - 对 $p \geq 1$，有 $E(|X|^p) \geq |E(X)|^p$ $(g(x)=|x|^p,\, p \geq 1)$; - $E(e^{cX}) \geq e^{cE(X)}$ $(g(x)=e^{cX})$.

琴生不等式在信息论、机器学习和优化等领域有许多应用。

2.4 随机变量的变换

例 2.17

- 琴生不等式的一个应用与**KL 散度**有关。 - KL 散度在决策树上下文中称为**信息增益**，也称为**相对熵**。 - 您可能在机器学习或信息论的课程作业中了解到这个概念。这个概念实际上相当简单。 - 简而言之，如果您有两个概率分布 $p(x)$ 和 $q(x)$，KL 散度衡量它们之间的差异/距离：

$$D_{KL}(p||q) = \int p(x) \log \left(\frac{p(x)}{q(x)} \right) dx.$$

- KL 散度具有性质：$D_{KL}(p||q) \geq 0$ 且 $D_{KL}(p||q) = 0$ 当且仅当几乎处处有 $q(x) = p(x)$。 - 您能使用琴生不等式证明 $D_{KL}(p||q) \geq 0$ 吗？

2.4 随机变量的变换

解

- 令 X 为具有概率密度函数 $p(x)$ 的随机变量，定义另一个随机变量 $Y = q(X)/p(X)$。 - 令函数 $g(x) = -\log(x)$，则 $g(x)$ 是凸函数，因此由琴生不等式：

$$E(g(Y)) \geq g(E(Y)).$$

- 由于

$$E(g(Y)) = E\left(g\left(\frac{q(X)}{p(X)}\right)\right) = -\int p(x) \log \left(\frac{q(x)}{p(x)}\right) dx = D_{KL}(p||q),$$

$$E(Y) = E\left(\frac{q(X)}{p(X)}\right) = \int p(x)\frac{q(x)}{p(x)} dx = \int q(x) dx = 1 \Rightarrow g(E(Y)) = -\log(1) = 0.$$

- 综合起来，我们有

$$D_{KL}(p||q) \geq 0.$$

第2章结束

第2章