第二章　一维随机变量及其分布

2.1　随机变量

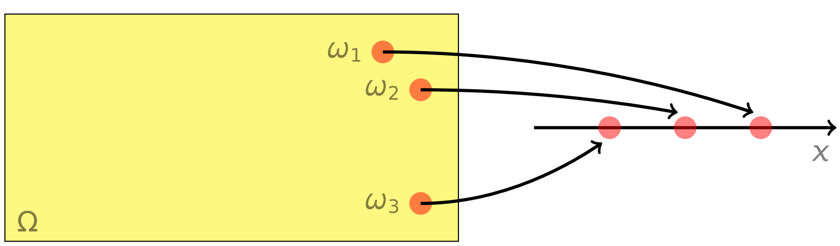
随机试验的结果通常可以用数量来表示：

- 扔一个硬币所得的结果；
- 掷一颗骰子所得的点数；
- 抽查样品时的废品个数；
- 广州每日的平均气温；
- 某电子管的使用寿命；

将试验结果数值化, 就产生了随机变量的概念.

定义 **1.** 设 $X = X(\omega)$ 是定义在样本空间 Ω 上的实值函数, 称 $X = X(\omega)$ 为随机变量.

下图给出样本点 ω 与实数 $X = X(\omega)$ 对应的示意图.



样本点 ω	HH	HT	TH	TT
X	2	1	1	0

随机变量一般用大写英文字母 X、Y、Z 或小写希腊字母 ξ、η、γ 来表示.

对实验结果 ω 本身就是一个数的随机试验, 令

$$X = X(\omega) = \omega,$$

则 X 就是一个随机变量.

1. 电视机的寿命 T.
2. 掷一颗骰子, 出现的点数 X.
3. 每天进入某超市的顾客数 Y.
4. $\cdots$

对于样本点本身不是数的随机试验, 这时可根据需要设计随机变量.

例 **1.** 检查一个产品, 只考察其合格与否, 则其样本空间为 $\Omega = \{$合格产品, 不合格产品$\}$. 这时可以设计一个随机变量 X 如下

$$X = \begin{cases} 1, & \omega = \text{合格产品}; \\ 0, & \omega = \text{不合格产品}. \end{cases}$$

例 **2.** 将一枚硬币抛掷两次, 感兴趣的是投掷中出现 H 的总次数, 而对出现 H,T 出现的顺序不关心. 例如, 我们只关心出现 H 的总次数是 1, 而不在乎出现的是“HT”还是“TH”. 以 X 表示两次投掷出现 H 的总次数, 对于样本空间 $\Omega\{\omega\} = \{HH, HT, TH, TT\}$ 中的每一个样本点, X 都有一个值与之对应, 即有 这样设计出来的 X 也是一个随机变量.

随机变量的取值随试验的结果而定, 而试验的各个结果出现有一定的概率, 因而随机变量的取值有一定的概率.

例如, 在例 2 中, X 的取值为 1, 记为 $\{X = 1\}$, 对应的样本点的集合为 $A = \{HT, TH\}$, 这是一个事件, 事件 A 发生当且仅当 $\{X = 1\}$ 发生. 我们称概率 $P(A) = P\{HT, TH\}$ 为 $\{X = 1\}$

X	x_1	$x_2,$	$\cdots$	x_n	$\cdots$
p_k	p_1	$p_2,$	$\cdots$	p_n	$\cdots$

的概率, 即

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

一般地, 若 I 是一个实数集, $\{X\in I\}$ 记为事件 B, 即

$$B=\{\omega|X(\omega)\in I\},$$

则

$$P\{X\in I\}=P(B)=P\{\omega|X(\omega)\in I\}.$$

按照随机变量可能取值的情况, 可以把它们分为两类：离散型随机变量和非离散型随机变量, 而非离散型随机变量中最重要的是连续型随机变量. 本章主要研究离散型及连续型随机变量.

2.2 离散型随机变量

定义 **1.** 如果随机变量的全部可能取的值只有有限个或可列无限多个, 则称这种随机变量为离散型随机变量.

一般地, 设离散型随机变量 X 所有可能取的值为 x_k $(k=1,2,...)$, X 取各个可能值的概率, 即事件 $\{X=x_k\}$ 的概率为

$$P\{X=x_k\}=p_k,\ k=1,2,...,\tag{1}$$

称 (1) 式为离散型随机变量 X 的分布律或概率分布.

分布律也可以用下面的表格来表示：

由概率的定义, 式 p_k 应满足以下条件：

1. $p_k\geq 0, k=1,2...;$
2. $\sum\limits_{k=1}^{\infty}p_k=1.$

例 **1.** 某系统有两台机器相互独立地运转. 设第一台与第二台机器发生故障的概率分别为 0.1, 0.2, 以 X 表示系统中发生故障的机器数, 求 X 的分布律.

分析: 求分布律需要求事件 $\{X=x_k\}$ 的概率

$$P\{X=X_k\}=p_k,\ k=1,2,...,$$

解. 设 A_i 表示事件"第 i 台机器发生故障", $i=1,2.$ 则

$$P\{X=0\}=P\left(\overline{A}_1\overline{A}_2\right)=0.9\times0.8=0.72$$

$$P\{X=1\}=P\left(A_1\overline{A}_2\right)+P\left(\overline{A}_1A_2\right)$$

$$=0.1\times0.8+0.9\times0.2=0.26$$

$$P\{X=2\}=P(A_1A_2)=0.1\times0.2=0.02$$

故所求概率分布为:

X	0	1	2
p_k	0.72	0.26	0.02

2.2.1 $(0-1)$ 分布

设随机变量 X 只可能取 0 与 1 两个值, 它的分布律是

$$P\{X=k\}=p^kq^{1-k},k=0,1,p+q=1(0<p<1)$$

则称 X 服从$(0-1)$ 分布或两点分布. $(0-1)$ 分布的分布律也可写成

X	0	1
p_k	q	p

例子. 抛一枚硬币，观察出现正面 H 还是反面 T，正面 $X=0$, 反面 $X=1$

对于一个随机试验, 如果它的样本空间只包含两个元素, 即 ω_1,ω_2 我们总能在 Ω 上定义一个服从 $(0-1)$ 分布的随机变量

$$X=\begin{cases}0, & \omega=\omega_1;\\ 1, & \omega=\omega_2.\end{cases}$$

来描述这个随机试验的结果.

检查产品的质量是否合格, 对新生婴儿的性别进行登记, 检验种子是否发芽以及前面多次讨论过的“抛硬币”试验都可以用$(0-1)$ 分布的随机变量来描述.

2.2.2 伯努力试验与二项分布

设试验 E 只有两个可能结果：A 及 $\overline{A}$, 则称为伯努利 (Bernoulli) 试验**.** 设 $P(A)=p$, $0<p<1$, 则 $P(\overline{A})=1-p$.

将伯努力试验独立重复地进行 n 次, 则称这一串重复的独立试验为 n 重伯努利试验.

独立: 各次的的试验结果互不影响. 重复: 每次试验中 $P(A)=p$ 保持不变.

定理 (伯努利定理)**.** 设一次试验中事件 A 发生的概率为 $p(0<p<1)$, 则 n 重伯努利试验中, 事件 A 恰好发生 $k(0\leq k\leq n)$ 次的概率为

$$C_n^k p^k(1-p)^{n-k}.$$

证明. 记 n 重伯努力试验中事件 A 正好出现 k 次这一事件为 B_k, 以 A_i 表示第 i 次试验中出现事件 A, 以 $\overline{A}_i$ 表示第 i 次试验中出现 $\overline{A}$, 则

$$B_k=A_1A_2\cdots A_k\overline{A}_{k+1}\cdots\overline{A}_n\cup\cdots\cup\overline{A}_1\overline{A}_2\cdots\overline{A}_{n-k}A_{n-k+1}\cdots A_n$$

右边的每一项表示某 k 次试验出现事件 A, 另外 $n-k$ 次试验出现 $\overline{A}$, 这种项共有 C_n^k 个, 而且两两互不相容.

由试验的独立性得

$$\begin{aligned}&P\left(A_1A_2\cdots A_k\overline{A}_{k+1}\cdots\overline{A}_n\right)\\=&P(A_1)P(A_2)\cdots P(A_k)P\left(\overline{A}_{k+1}\right)\cdots P\left(\overline{A}_n\right)\\=&p^k(1-p)^{n-k}\end{aligned}$$

同理可得, 右边各项所对应的概率均为 $p^k(1-p)^{n-k}$, 利用概率的加法定理知

$$P(B_k)=C_n^k p^k(1-p)^{n-k}.$$

k	$P\{X=k\}$	k	$P\{X=k\}$
0	0.012	6	0.1
1	0.058	7	0.055
2	0.137	8	0.022
3	0.205	9	0.007
4	0.218	10	0.002
5	0.175	≥ 11	< 0.001

在 n 重伯努力试验中, 若以 X 表示 n 重伯努力试验中事件 A 出现的次数, 显然有

$$P\{X=k\}=P(B_k)=C_n^k p^k(1-p)^{n-k}.$$

定义. 如果随机变量 X 的分布律为

$$P\{X=k\}=C_n^k p^k(1-p)^{n-k},$$

其中 $0<p<1$, $0\leq k\leq n$, 则称 X 服从参数为 n, p 的二项分布, 记为 $X\sim B(n,p)$.

小注: 非负性显然, 规范性由二项式定理可得, 即 $\sum_{k=0}^{n}P\{X=k\}=\sum_{k=0}^{n}C_n^k p^k(1-p)^{n-k}=(p+1-p)^n=1$

例 **2.** 已知某类产品的次品率为 0.2, 现从一大批这类产品中随机地抽查 20 件, 问恰好有 $k(k=0,1,2,,20)$ 件次品的概率是多少?

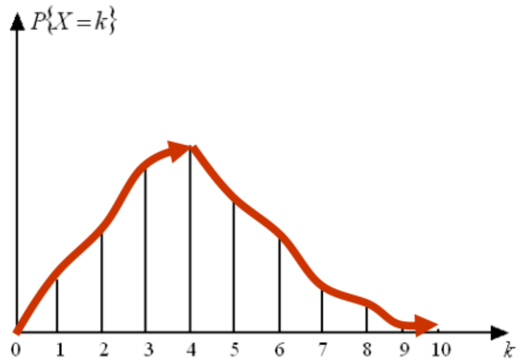
小注: 在物品总数很大, 而抽取数目较小时, 不放回抽样可以看成放回抽样, 从而可按伯努利试验来处理.

解. 我们将检查一件产品是否为次品看成是一次试验, 检查 20 件产品相当于做 20 重伯努利试验. 以 X 记抽出的 20 件产品中次品的件数, 那么 X 是一个随机变量, 且 $X\sim b(20,0.2)$, 则所求的概率为

$$P\{X=k\}=C_{20}^k(0.2)^k(0.8)^{20-k}, k=0,1,\cdots,20.$$

作出上表的图形, 如下图所示

从上图可以看出, 概率 $P\{X=k\}$ 先是随着 k 增加而增加, 直至达到最大值 $(k=4)$, 随后单调减少. 一般地, 对于固定的 n 及 p, 二项分布 $b(n,p)$ 都有类似的结果



例 **3.** 设某种鸭在正常情况下感染某种传染病的概率为 20%. 现新发明两种疫苗, 疫苗 A 注射到 9 只健康鸭后无一只感染传染病, 疫苗 B 注射到 25 只鸭后仅有一只感染, 试问应如何评价这两种疫苗, 能否初步估计哪种疫苗较为有效?

解**.** 若疫苗 A 完全无效, 则注射后鸭受感染的概率仍为 0.2, 故 9 只鸭中无一只感染的概率为

$$0.8^9 = 0.1342.$$

同理, 若疫苗 B 完全无效, 则 25 只鸭中至多有一只感染的概率为

$$0.8^{25} + C_{25}^1(0.2)^1(0.8)^{24} = 0.0274.$$

若 B 完全无效, 则 25 只健康鸭至多有一只感染的概率只有 0.0274, 由实际推断原理, 小概率事件在一次试验中实际上几乎是不发生的, 但现在却发生了, 有理由怀疑假设的正确性. 因此可以初步认为疫苗 B 是有效的, 且比 A 有效 (因为 0.0274 比 0.1342 小的多).

2.2.3 泊松分布

设随机变量 X 所有可能取值为 $0,1,2,...,$ 而取各个值的概率为

$$P\{X=k\} = \frac{\lambda^k}{k!}e^{-\lambda}, \ k=0,1,2\cdots$$

其中 $\lambda > 0$ 为常数, 则称 X 服从参数为 λ 的泊松分布, 记为 $X \sim \pi(\lambda)$.

显然, $P\{X=k\} \geq 0$, $k=0,1,2\cdots$, 且

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

即 $P\{X=k\}$ 满足分布律的两个条件.

泊松分布常与单位时间（或单位面积、单位产品等）上的计数过程相联系：

- 某地区每天发生火灾的次数.
- 某地区每年发生暴雨的次数.
- 某种玻璃每平方米内的气泡数.
- 某医院每天前来就诊的人数.
- 某份杂志各期的错别字数目.

例 **4.** 商店的历史销售记录表明, 某种商品每月的销售量服从参数为 $\lambda=10$ 的泊松分布. 为了以 95% 以上的概率保证该商品不脱销, 问商店在月底至少应进该商品多少件?

解**.** 设商店每月销售这种商品 X 件, 月底的进货量为 n 件, 按题意要求为

$$P\{X\le n\}\ge 0.95,$$

X 服从 $\lambda=10$ 的泊松分布, 则有 $\sum_{k=0}^{n}\frac{10^k}{k!}e^{-10}\ge 0.95$. 由附录的泊松分布表知

$$\sum_{k=0}^{14}\frac{10^k}{k!}\mathrm{e}^{-10}=0.917<0.95$$

$$\sum_{k=0}^{15}\frac{10^k}{k!}\mathrm{e}^{-10}=0.951>0.95$$

只要在月底进货 15 件 (假定上个月没有存货), 就可以 95% 的概率保证这种商品在下个月内不会脱销.

定理 (泊松定理)**.** 在 n 重伯努利试验中, 事件 A 在每次试验中发生概率为 p_n （注意这与实验的次数 n 有关）, 如果 $n\to\infty$ 时, $np_n\to\lambda$ (λ 为常数), 则对任意给定的非负整数 k, 有

$$\lim_{n\to\infty}C_n^k p_n^k(1-p_n)^{n-k}=\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}.$$

证明. 记 $np_n=\lambda_n$, 即 $p_n=\lambda_n/n$, 于是

$$\begin{aligned}C_n^k p_n^k(1-p_n)^{n-k}&=\frac{n(n-1)\cdots(n-k+1)}{k!}\left(\frac{\lambda_n}{n}\right)^k\left(1-\frac{\lambda_n}{n}\right)^{n-k}\\&=\frac{\lambda_n^k}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\left(1-\frac{\lambda_n}{n}\right)^{n-k}.\end{aligned}$$

对于固定的 k,

$$\lim_{n\to\infty}\lambda_n=\lambda,\lim_{n\to\infty}\left(1-\frac{\lambda_n}{n}\right)^{n-k}=e^{-\lambda},\lim_{n\to\infty}\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)=1$$

从而

$$\lim_{n\to\infty}C_n^k p_n^k(1-p_n)^{n-k}=\frac{\lambda^k}{k!}e^{-\lambda}.$$

对于任意的非负整数 k 成立.

由于泊松定理是在 $n\to\infty$ 条件下获得的, 故在计算二项分布 $b(n,p)$ 时, 当 n 很大, p 很小, 而乘积 $\lambda=np$ 大小适中时, 可以用泊松分布作近似, 即

$$C_n^k p^k(1-p)^{n-k}\approx\frac{(np)^k}{k!}e^{-np},k=0,1,2,\cdots$$

例 **5.** 为保证设备正常工作, 需要配备一些维修工. 如果各台设备发生故障是相互独立的, 且每台设备发生故障的概率都是 *0.01*. 试求在以下情况下, 求设备发生故障而不能及时修理的概率.

(1) 一名维修工负责 *20* 台设备.

(2) *3* 名维修工负责 *90* 台设备.

(3) *10* 名维修工负责 *500* 台设备.

解**.** *(1)* 以 X_1 表示 *20* 台设备中同时发生故障的台数, 则 $X_1\sim b(20,0.01)$. 用参数为 $\lambda=np=20\times0.01=0.2$ 的泊松分布作近似计算, 得所求概率为

$$P\{X_1>1\}\approx1-\sum_{k=0}^{1}\frac{0.2^k}{k!}e^{-0.2}=1-0.982=0.018.$$

(2) 以 X_2 表示 *90* 台设备中同时发生故障的台数, 则 $X_2\sim b(90,0.01)$. 用参数为 $\lambda=np=90\times0.01=0.9$ 的泊松分布作近似计算, 得所求概率为

$$P\{X_2>3\}\approx1-\sum_{k=0}^{3}\frac{0.9^k}{k!}e^{-0.9}=1-0.982=0.013.$$

注意: 此种情况下, 不但所求概率比 (1) 中有所降低, 而且 3 名维修工负责 90 台设备相当于每个维修工负责 30 台设备, 工作效率是 (1) 中的 1.5 倍. (3) 以 X_3 表示 500 台设备中同时发生故障的台数, 则

$X_3 \sim b(500, 0.01)$. 用参数为 $\lambda = np = 500 \times 0.01 = 5$ 的泊松分布作近似计算, 得所求概率为

$$P\{X_3 > 10\} \approx 1 - \sum_{k=0}^{10} \frac{5^k}{k!} e^{-5} = 1 - 0.982 = 0.014.$$

注意: 此种情况下所求概率与 (2) 中基本上一样, 而 10 名维修工负责 500 台设备相当于每个维修工负责 50 台设备, 工作效率是 (2) 的 1.67 倍, 是 (1) 中的 2.5 倍.

小注: 若干维修工共同负责大量设备的维修, 将提高工作的效率.

2.3　随机变量的分布函数

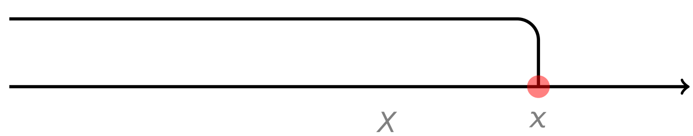
定义**.** 设 X 是一个随机变量, x 是任意实数, 函数

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

称为 X 的分布函数.

分布函数是一个普通的函数, 其定义域是整个实数轴**.**

在几何上, 它表示随机变量 X 的取值落在实数 x 左边的概率.

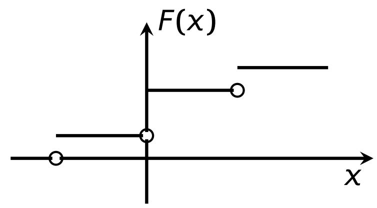


分布函数具有以下基本性质:

1. $0 \le F(x) \le 1$.
2. $F(x)$ 是 x 的不减函数.
3. $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$, $F(+\infty) = \lim_{x \to +\infty} F(x) = 1$.
4. $F(x+0) = F(x)$, 即 $F(x)$ 是右连续的.

例 **1.** 设随机变量 X 的分布律为 求 X 的分布函数, 并求 $P(0 \le X \le 1)$.

X	-1	0	1
p_k	1/4	1/2	1/4



解: (1) 由概率的有限可加性分布函数为:

$$F(x)=\begin{cases}0 & x<-1\\ \frac{1}{4} & 1\leq x<0\\ \frac{3}{4} & 0\leq x<1\\ 1 & x\geq 1\end{cases}$$

概率分布函数 $F(x)$ 的图像为

$$\begin{aligned}P\{0\leqslant X\leqslant 1\}&=P\{0<X\leqslant 1\}+P\{X=0\}\\&=F(1)-F(0)+P\{X=0\}=1-\frac{3}{4}+\frac{1}{2}=\frac{3}{4}.\end{aligned}$$

一般地, 设离散型随机变量 X 的概率分布为

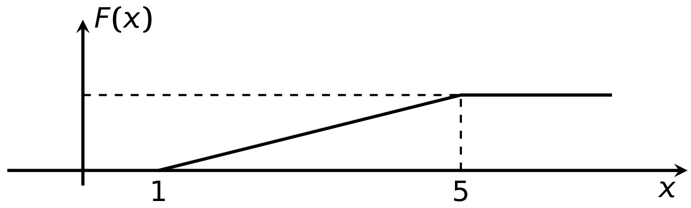
$$P\{X=x_k\}=p_k,\qquad k=1,2,\cdots.$$

则由概率的可列可加性可得 X 的分布函数为

$$F(x)=P\{X\leq x\}=\sum_{x_k\leq x}p_k,$$

这里的和式是对所有满足 $x_k\leq x$ 的 p_k 求和. 分布函数 $F(x)$ 在 $x=x_k(k=1,2,...)$ 处有跳跃值, 其跳跃值为 $p_k=P\{X=x_k\}$.

例 **2.** 在区间 $[1,5]$ 上任意掷一个质点, 用 X 表示这个质点与原点的距离, 则 X 是一个随机变量. 如果这个质点落在 $[1,5]$ 上任一子区间内的概率与这个区间的长度成正比, 求 X 的分布函数



解. 由题意知 $\{1 \le X \le 5\}$ 是一个必然事件, 即

$$P\{1 \le X \le 5\} = 1$$

若 $x < 1$, 则 $\{X \le x\}$ 是不可能事件,

$$F(x) = P\{X \le x\} = 0$$

若 $1 \le x \le 5$, 则

$$P\{1 \le X \le x\} = k(x-1)$$

特别取 $x = 5$ 由 $P\{1 \le X \le 5\} = 1$ 可得 $k = 1/4$, 从而

$$F(x) = P\{X \le x\} = P\{x < 1\} + P\{1 \le X \le x\} = \frac{1}{4}(x-1).$$

若 $x > 5$, 则 $\{X \le x\}$ 是必然事件, $F(x) = 1$. 综上, X 的分布函数为

$$F(x) = \begin{cases} 0, & x < 1 \\ \frac{1}{4}(x-1), & 1 \le x < 5 \\ 1, & x \ge 5 \end{cases}$$

$F(x)$ 的图形如下图, 它是一个定义在 $-\infty, +\infty$ 上的一个连续函, 在整个数轴上没有一个跳跃点.

2.4 连续型随机变量及其概率密度函数

定义. 对于随机变量 X 的分布函数 $F(x)$ 如果存在非负函数 $f(x)$, 使对于任意实数有

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

则 X 称为连续型随机变量, 其中函数 $f(x)$ 称为 x 的概率密度函数, 简称概率密度.

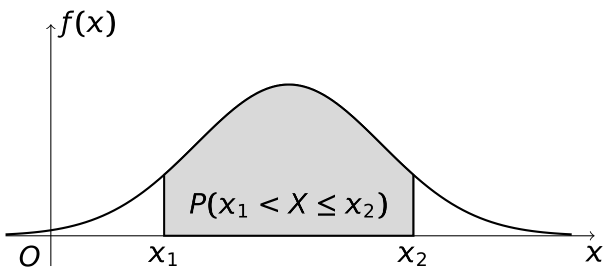
由定义知道, 概率密度 $f(x)$ 具有以下性质

1. $f(x) \geq 0$,
2. $\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1$.
3. 对于任意实数 $x_1, x_2, x_1 \leq x_2$, 有

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

4. 若 $f(x)$ 在点 x 连续, 则 $F'(x) = f(x)$.

小注: 性质 1 和性质 2 是概率密度函数的基本性质. 由性质 3 可知, X 落在区间 $(x_1, x_2]$ 上的概率 $P\{x_1 < X \leq x_2\}$ 等于区间 $(x_1, x_2]$ 上曲线 $f(x)$ 之下曲边梯形的面积.



由性质 4, 对于 $f(x)$ 的连续点 x, 有

$$f(x) = \lim_{\Delta x \to 0^+} \frac{F(x+\Delta x) - F(x)}{\Delta x} = \lim_{\Delta x \to 0^+} \frac{P\{x < X \leqslant x + \Delta x\}}{\Delta x}$$

这表明概率密度函数 $f(x)$ 不是随机变量 X 取 x 的概率, 而是 X 在点 x 的概率的密集程度, $f(x)$ 的大小能反映出 X 取 x 附近的值的概率大小 (因此对于连续型随机变量，用概率密度描述它的分布比分布函数更直观).

例 **1.** 设连续型随机变量 X 具有概率密度

$$f(x) = \begin{cases} kx+1, & 0 \leq x \leq 2 \\ 0, & \text{其它} \end{cases}$$

(1) 确定常数 k.

(2) 求 X 的分布函数 $F(x)$.

(3) 求 $P\{3/2 < X \le 5/2\}$.

解. *(1)* 由条件易知

$$\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1 \Longrightarrow \int_0^2 (kx+1)\mathrm{d}x = 1,$$

解得 $k = -1/2$.

(2) X 的分布函数为

$$F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t = \begin{cases} 0, & x < 0 \\ -\frac{1}{4}x^2 + x, & 0 \le x \le 2 \\ 1, & x \ge 2 \end{cases}$$

(3) 由 *(2)* 易知

$$\begin{aligned} P\{3/2 < x \le 5/2\} &= F\left(\frac{5}{2}\right) - F\left(\frac{3}{2}\right) \\ &= 1 - 0.9375 = 0.0625 \end{aligned}$$

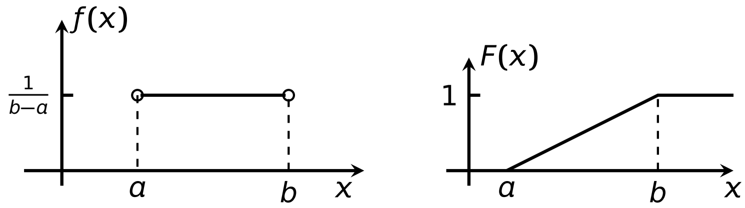
2.4.1 均匀分布

设连续型随机变量 X 具有概率密度

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

则称 X 在区间 (a,b) 上服从均匀分布, 记为 $X \sim U(a,b)$.

易知 $f(x) \ge 0$ 且 $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$.



满足概率密度函数的两个基本性质.

由均匀分布的概率密度函数容易求得其分布函数为

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

其概率密度函数和分布函数的图像为

注记**.** X 落在 (a,b) 子区间的概率只依赖于子区间的长度，与子区间的位置无关.

均匀分布有如下这些例子：

1. 四舍五入时产生的误差.
2. 查看当前时间时的分钟值.

例 **2.** 设随机变量 X 在 $[2,5]$ 上服从均匀分布, 现对 X 进行三次独立观测. 试求至少有两次测值大于 3 的概率.

解**.** 依题意得：X 的密度函数为

$$f(x)=\begin{cases} \frac{1}{3} & ,2<x<5 \\ 0 & ,\text{其它} \end{cases}$$

设 Y 表示三次独立观测其观测值大于 3 的次数

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

$$P(Y\geq 2)=C_3^2\left(\frac{2}{3}\right)^2\cdot\frac{1}{3}+C_3^3\left(\frac{2}{3}\right)^3=\frac{20}{27}.$$

2.4.2　指数分布

定义. 如果随机变量 X 有以下概率密度

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & otherwise \end{cases},$$

其中 $\lambda > 0$, 则称 X 服从参数为 λ 的指数分布, 记为

$$X \sim E(\lambda).$$

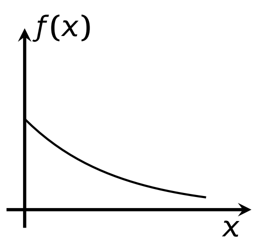
易知 $f(x) \geq 0$ 且 $\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{+\infty} \lambda e^{-\lambda x}\, dx = 1$,

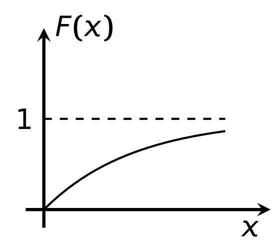
满足概率密度函数的两个基本性质.

赘述分布的分布函数为

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}.$$

指数分布的概率密度及分布函数分别如图所示:





指数分布经常作为时间间隔或等待时间的分布：

- 婴儿出生的时间间隔
- 客户来电的时间间隔
- 商品销售的时间间隔
- 网站访问的时间间隔

- 取号排队的等待时间
- 电子产品的寿命长度

例 **3.** 已知某种电子元件寿命 *(*单位：*h)* 服从参数 $\lambda = 1/1000$ 的指数分布*,* 求 *3* 个这样的元件使用 *1000* 小时至少有一个已损坏的概率*.*

解**.** X 的概率密度为

$$f(x) = \begin{cases} \dfrac{1}{1000}\mathrm{e}^{-\frac{x}{1000}}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

于是

$$P\{X > 1000\} = \int_{1000}^{+\infty} f(x)\mathrm{d}x = \mathrm{e}^{-1}$$

各元件的寿命是否超过 *1000* 小时是独立的*,* 因此 *3* 个元件使用 *1000* 小时都未损坏的概率为 e^{-3}*,* 从而至少有一个已损坏的概率为 $1 - \mathrm{e}^{-3}$*.*

2.4.3 正态分布

定义**.** 如果随机变量 X 有以下概率密度

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

其中 μ, σ 为常数且 $\sigma > 0$*,* 则称 X 服从参数为 μ, σ 的正态分布*,* 简记为

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

显然 $f(x) \ge 0$, 下证 $\int_{-\infty}^{+\infty} f(x) = 1$.

令 $\frac{x-\mu}{\sigma} = t$*,* 得

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\mathrm{d}x = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t.$$

又

$$\int_0^{+\infty}\mathrm{e}^{-x^2}\mathrm{d}x = \frac{\sqrt{\pi}}{2} \Longrightarrow \int_{-\infty}^{+\infty}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t = \sqrt{2\pi},$$

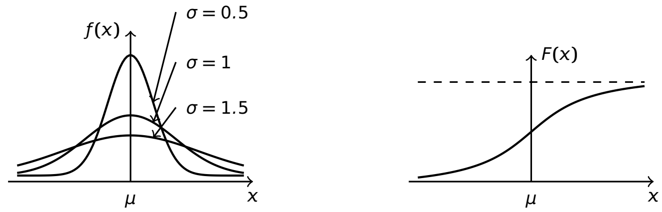
于是

$$\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\,\mathrm{d}x = 1.$$

正态分布的分布函数为

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x} \mathrm{e}^{-\frac{(t-\mu)^2}{2\sigma^2}}\,\mathrm{d}t$$

正态分布的概率密度函数和分布函数如图所示



(a) 密度函数 $f(x)$　　(b) 分布函数 $F(x)$

显然, 函数 $f(x)$ 的图形关于直线 $x = \mu$ 对称, $f(x)$ 在 $x = \mu$ 处达到最大. 当 μ 固定时, σ 的值越小, $f(x)$ 的图形越尖. 反之, σ 的值越大, $f(x)$ 的图形就越平.

当 $\mu = 0, \sigma = 1$ 时, 称 X 服从标准正态分布, 记为 $X \sim N(0,1)$. 其概率密度和分布函数分别用为 $\varphi(x)$ 和 $\Phi(x)$ 表示, 即有

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

标准正态分布的分布函数的性质

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

引理 **1.** 若 $X \sim N(\mu, \sigma^2)$, 则 $Y = \frac{X-u}{\sigma} \sim N(0,1)$.

例 **4.** 已知 $N \sim (8, 4^2)$ 求 $P\{X \le 16\}, P\{X \le 0\}$ 及 $P\{12 < X \le 20\}$

解**.** 由引理及 X 的分布函数, 查表得

$$P\{x \le 16\} = P\left\{\frac{x-8}{4} \le \frac{16-8}{4}\right\} = \Phi(2) = 0.9773,$$

$$P\{x \le 0\} = P\left\{\frac{x-8}{4} \le \frac{-8}{4}\right\} = \Phi(-2) = 1 - \Phi(2) = 0.0227,$$

$$
\begin{aligned}
P\{12 < x \le 20\} &= P\left\{\frac{12-8}{4} < \frac{x-8}{4} \le \frac{20-8}{4}\right\} \\
&= \Phi(3) - \Phi(1) \\
&= 0.9987 - 0.8413 = 0.1574.
\end{aligned}
$$

例 **5.** 设 $X \sim N(\mu, \sigma^2)$ 求 X 落在区间 $(\mu - k\sigma, \mu + k\sigma)$ 内的概率, $k = 1, 2, \ldots$.

解**.** 由引理可得

$$
\begin{aligned}
P\{|X-\mu| < k\sigma\} &= P\{\mu - k\sigma < X < \mu + k\sigma\} \\
&= \Phi(k) - \Phi(-k) = 2\Phi(k) - 1
\end{aligned}
$$

于是

$$
\begin{aligned}
P\{|X-\mu| < \sigma\} &= 2\Phi(1) - 1 = 0.6826 \\
P\{|X-\mu| < 2\sigma\} &= 2\Phi(2) - 1 = 0.9544 \\
P\{|X-\mu| < 3\sigma\} &= 2\Phi(3) - 1 = 0.9974
\end{aligned}
$$

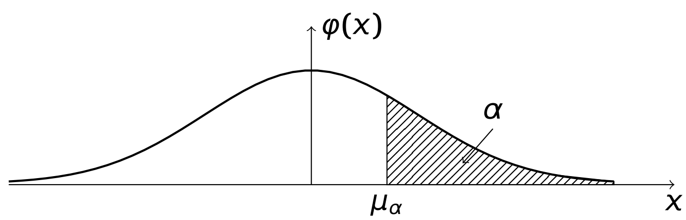
则 $P\{|X-\mu| \geqslant 3\sigma\} = 1 - P\{|X-\mu| < 3\sigma\} = 0.0026 < 0.003.$

X 落在 $(\mu - 3\sigma, \mu + 3\sigma)$ 以外的概率小于 *0.003,* 在实际问题中常认为它不会发生.

设随机变量 $X \sim N(0,1)$, 对给定的 $\alpha \in (0,1)$, 称满足条件

$$
P\{X > u_\alpha\} = \alpha
$$

的点 u_α 为标准正态分布的上 α 分位点**.**



易知, $\Phi(u_\alpha) = 1 - \alpha.$

2.5　随机变量的函数的分布

在实际问题中, 有时我们关心的随机变量 Y 不容易直接测量, 而是要测量另外一个随机变量 X,

把 Y 表示为 X 的函数 $Y=g(X)$. 由此引出的问题是：已知 X 的分布, 如何求 Y 的分布？例如：已知圆球直径 D 的分布, 求圆球体积 $V=\frac{\pi D^3}{6}$ 的分布.

例 **1.** 设随机变量 X 有如下概率分布：

X	-1	0	1	2
p_k	0.3	0.2	0.1	0.4

求 *(1)* $Y=2X$, *(2)* $Z=(X-1)^2$ 的分布律.

解**.** *(1)* Y 的所有可能取值为 $-2,0,2,4$. 由

$$P\{Y=2k\}=P\{X=k\}=p_k$$

得 Y 的分布律为

Y	-2	0	2	4
p_k	0.3	0.2	0.1	0.4

(2) Z 的所有可能取值为 $0,1,4$, 故

$$\begin{aligned}P\{z=0\}&=P\left\{(X-1)^2=0\right\}=P\{X=1\}=0.1,\\ P\{Z=1\}&=P\left\{(X-1)^2=1\right\}\\ &=P\{X=0\}+P\{X=2\}=0.6,\\ P\{Z=4\}&=P\left\{(X-1)^2=4\right\}=P\{X=-1\}=0.3.\end{aligned}$$

故 Z 的分布律为

Z	0	1	4
p_k	0.1	0.6	0.3

例 **2.** 设随机变量 $X\sim N(\mu,\sigma^2)$, 试证明 X 的线性函数 $Y=aX+b$ 也服从正态分布.

解**.** 分别记 X,Y 的概率密度函数为 $f_X(x),f_Y(x)$, 分布函数为 $F_X(x),F_Y(y)$. 不妨设 $a>0$, 则

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

对 $F_Y(y)$ 关于 y 求导, 得 $Y=aX+b$ 的概率密度函数为

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

而 X 的概率密度函数为

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

所以

$$f_Y(y)=\frac{1}{\sqrt{2\pi}(a\sigma)}e^{-\frac{[y-(a\mu+b)]^2}{2(a\sigma)^2}},-\infty<x<+\infty.$$

若 $a<0$, 以同样的方法可以求得

$$f_Y(y)=\frac{1}{\sqrt{2\pi}(|a|\sigma)}e^{-\frac{[y-(a\mu+b)]^2}{2(a\sigma)^2}},-\infty<x<+\infty.$$

故 $\boxed{Y\sim N\left(a\mu+b,(a\sigma)^2\right)}$.

小注: 取 $a=1/\sigma$, $b=-\mu/\sigma$, 得

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

例 **3.** 设随机变量 X 具有概率密度函数 $f_X(x),-\infty<x<+\infty$, 求 $Y=X^2$ 的概率密度函数.

解**.** 分别记 X,Y 的概率密度函数为 $f_X(x),f_Y(x)$, 分布函数为 $F_X(x),F_Y(y)$. 先求 Y 的分布函数 $F_Y(y)$. 由于 $Y=X^2\geq 0$, 故当 $y\leq 0$ 时, $F_Y(y)=0$. 当 $y>0$ 时有

$$\begin{aligned}F_Y(y)&=P\{Y\leqslant y\}=P\left\{X^2\leqslant y\right\}\\&=P\{-\sqrt{y}\leqslant X\leqslant\sqrt{y}\}\\&=F_X(\sqrt{y})-F_X(-\sqrt{y})\end{aligned}$$

对 $F_Y(y)$ 关于 y 求导, 得 Y 的概率密度函数为

$$f_Y(y)=\begin{cases}\frac{1}{2\sqrt{y}}\left[f_X(\sqrt{y})+f_X(-\sqrt{y})\right], & y>0\\ 0, & y\leqslant 0\end{cases}$$

特别地, 若 $X \sim N(0,1)$, 其概率密度函数为

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

则 $Y = X^2$ 的概率密度函数是

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}}, & y > 0 \\ 0, & y \leqslant 0 \end{cases}$$

此时称 Y 服从自由度为 1 的 χ^2 分布

定理 **1.** 设随机变量 X 具有概率密度 $f_X(x), -\infty < x < +\infty$; 函数 $g(x)$ 处处可导且恒有 $g'(x) > 0$(或恒有 $g'(x) < 0$) 则 $Y = g(X)$ 是连续型随机变量, 其概率密度为

$$f_Y(y) = \begin{cases} f_X[h(y)] \cdot |h'(y)|, & y \in (\alpha, \beta) \\ 0, & otherwise. \end{cases}$$

其中 $h(y)$ 是 $g(x)$ 的反函数, $\alpha = \min(g(-\infty), g(+\infty))$, $\beta = \max(g(-\infty), g(+\infty))$.

证明. 先考虑 $g'(x) > 0$ 的情况. 此时 $g(x)$ 在 $(-\infty, +\infty)$ 严格单调递增, 它的反函数 $h(y)$ 存在, 且在 (α, β) 严格单调递增. 分别记 X, Y 的概率密度函数为 $f_X(x), f_Y(x)$, 分布函数为 $F_X(x), F_Y(y)$. 先求 Y 的分布函数 $F_Y(y)$. 由于 $Y = g(x)$ 在 (α, β) 取值, 故当 $y \leq \alpha$ 时, $F_Y(y) = 0$; 当 $Y \geq \beta$ 时, $F_Y(y) = 1$; 当 $\alpha < y < \beta$ 时

$$F_Y(y) = P\{Y \leq y\} = P\{g(X) \leq y\}$$
$$= P\{X \leq h(y)\} = F_X[h(y)]$$

对 $F_Y(y)$ 关于 y 求导, 即得 Y 的概率密度函数为

$$f_Y(y) = \begin{cases} f_X[h(y)] \cdot h'(y), & y \in (\alpha, \beta) \\ 0, & \text{otherwise.} \end{cases}$$

再考虑 $g'(x) < 0$ 的情况, 同样的有

$$f_Y(y) = \begin{cases} f_X[h(y)] \cdot [-h'(y)], & y \in (\alpha, \beta) \\ 0, & \text{otherwise.} \end{cases}$$

综合以上两种情况, 命题得证.

注记**.** 若 $f_X(x)$ 在有限区间 $[a,b]$ 以外等于零, 则只需假设在 $[a,b]$ 区间上恒有 $g'(x)>0$(或恒有 $g'(x)<0$), 此时 $\alpha=\min(g(a),g(b))$, $\beta=\max(g(a),g(b))$.

例 **4.** 设随机变量 X 在 $(-\pi/2,\pi/2)$ 内服从均匀分布 $Y=\sin X$, 试求随机变量 Y 的概率密度函数.

解**.** $Y=\sin X$ 对应的函数 $y=g(x)=\sin x$ 在 $(-\pi/2,\pi/2)$ 上恒有 $g'(x)=\cos x>0$, 且有反函数

$$x=h(y)=\arcsin y, h'(y)=\frac{1}{\sqrt{1-y^2}}.$$

又 X 的概率密度为

$$f_X(x)=\begin{cases}\frac{1}{\pi}, & -\frac{\pi}{2}<x<\frac{\pi}{2}\\ 0, & \text{其它}\end{cases}$$

由前面结论得 $Y=\sin X$ 的概率密度函数为

$$f_Y(y)=\begin{cases}\frac{1}{\pi}\cdot\frac{1}{\sqrt{1-y^2}}, & -1<y<1\\ 0 & \text{其它}\end{cases}$$