第 4 章　连续型随机变量

4.1 分布函数

离散型随机变量利用概率分布列将随机变量的取值和对应的概率全部罗列出来. 然而一些随机现象的试验结果可能不止可列个取值, 此时不能一一列举出来, 例如候车的等待时间、一个地区的降雨量、一盏电灯的寿命等. 特别地, 对于连续性随机变量, 它在任意一个特定值的概率为 0 (将在 4.2 节介绍), 此时用分布列来描述这一类型的随机变量就根本行不通.

对于一些非离散型随机变量, 我们可能更关心在某个区间内的概率, 而不是它在某个特定点值的概率. 例如, 对于一盏电灯而言, 我们关心其寿命大于 1000 个小时的概率, 而不是恰好 1005 个小时的概率. 针对这些随机现象, 我们关注于随机变量 X 在一个区间 $[x_1, x_2]$ 上的概率 $P(x_1 \leqslant X \leqslant x_2)$. 为此引入分布函数的概念:

定义 4.1　给定随机变量 X, 对任意实数 $x \in (-\infty, +\infty)$, 函数

$$F(x) = P(X \leqslant x)$$

称为随机变量 X 的 **分布函数** (cumulative distribution function).

分布函数 $F(x)$ 是定义在 $(-\infty, +\infty)$ 的普通函数, 将普通函数与随机事件的概率关联起来, 有利于利用数学分析的知识来研究随机变量. 分布函数不限制随机变量的类型, 无论时离散型随机变量还是非离散型随机变量, 都有各自的分布函数.

分布函数的本质是概率, 考虑随机事件 $\{X \in (-\infty, x]\}$ 的概率. 对任意实数 $x_1 < x_2$ 有

$$P(x_1 < X \leqslant x_2) = P(X \leqslant x_2) - P(X \leqslant x_1) = F(x_2) - F(x_1).$$

若已知随机变量 X 的分布函数 $F(x)$, 则可以知道 X 落入任意区间 $(x_1, x_2]$ 上的概率, 因此分布函数完整地了刻画随机变量的统计规律性. 分布函数具有良好的分析性质:

定理 4.1　分布函数 $F(x)$ 具有如下性质:

- 单调性: 若 $x_1 < x_2$, 则 $F(x_1) \leqslant F(x_2)$;
- 规范性: $F(x) \in [0, 1]$, 且 $F(-\infty) = \lim_{x\to-\infty} F(x) = 0$, $F(+\infty) = \lim_{x\to+\infty} F(x) = 1$;
- 右连续性: $F(x + 0) = \lim_{\Delta x\to 0^+} F(x + \Delta x) = F(x)$.

证明　根据概率的非负性, 对任意 $x_1 < x_2$ 有

$$F(x_2) - F(x_1) = P(x_1 < X \leqslant x_2) \geqslant 0\ .$$

根据规范性有

$$
\begin{aligned}
1 = P(-\infty < X < +\infty) &= \sum_{n=-\infty}^{+\infty} P(n < X \leqslant n+1) = \sum_{n=-\infty}^{+\infty} F(n+1) - F(n) \\
&= \lim_{n\to-\infty} F(n) - \lim_{m\to+\infty} F(m) \ .
\end{aligned}
$$

根据 $F(x)$ 的单调性有 $F(-\infty) = \lim_{x\to-\infty} F(x) = \lim_{n\to-\infty} F(n)$ 和 $F(+\infty) = \lim_{x\to+\infty} F(x) = \lim_{n\to+\infty} F(n)$, 以及结合 $F(-\infty), F(+\infty) \in [0,1]$ 和 $F(+\infty) - F(-\infty) = 1$ 可得

$$
F(-\infty) = 0 \quad \text{和} \quad F(+\infty) = 1 \ .
$$

针对右连续性, 设 $\{x_n\}_{n=1}^{\infty}$ 是一个单调下降的数列且 $x_n \to x$, 则有

$$
F(x_1) - F(x) = P(x < X \leqslant x_1) = \sum_{n=1}^{+\infty} F(x_n) - F(x_{n+1}) = F(x_1) - \lim_{n\to+\infty} F(x_n) \ .
$$

于是得到 $\lim_{n\to+\infty} F(x_n) = F(x)$, 再结合函数 $F(x)$ 的单调性有

$$
F(x+0) = \lim_{n\to+\infty} F(x_n) = F(x) \ ,
$$

由此完成证明.

通过上面的证明发现, 分布函数的三条基本性质, 分别对应于概率的三条公理. 因此, 任何分布函数都满足三条基本性质, 而满足上面三条基本性质的函数必是某随机变量的分布函数.

有了分布函数, 就很容易计算随机变量 X 在很多区间上的概率, 例如

$$
\begin{aligned}
P(X > a) &= 1 - F(a) \\
P(X < a) &= F(a-0) = \lim_{x\to a^-} F(x) \\
P(X = a) &= F(a) - F(a-0) \\
P(X \geqslant a) &= 1 - F(a-0) \\
P(a \leqslant X \leqslant b) &= F(b) - F(a-0).
\end{aligned}
$$

针对离散型的随机变量 X, 设其分布列为 $p_k = P(X = x_k)$ $(k = 1, 2, \cdots)$, 根据概率的可列可加性可得 X 的分布函数为

$$
F(x) = P(X \leqslant x) = \sum_{k:\ x_k \leqslant x} p_k \ . \tag{4.1}
$$

例 4.1 随机变量 X 的分布列为 $P(X = -1) = P(X = 3) = 1/4$ 和 $P(X = 2) = 1/2$, 求 X 的分布函数.

解 当 $x < -1$ 时, 根据 (4.1) 有

$$F(x) = P(X \leqslant x) = P(\emptyset) = 0;$$

当 $-1 \leqslant x < 2$ 时, 根据 (4.1) 有

$$F(x) = P(X \leqslant x) = P(X = -1) = \frac{1}{4};$$

当 $2 \leqslant x < 3$ 时, 根据 (4.1) 有

$$F(x) = P(X \leqslant x) = P(X = -1) + P(X = 2) = \frac{3}{4};$$

当 $x \geqslant 3$ 时有 $F(x) = 1$. 如图 4.1(a) 所示, 分布函数 $F(x)$ 是一条阶梯形的曲线, 在 $x = -1, 2, 3$ 处有跳跃点.

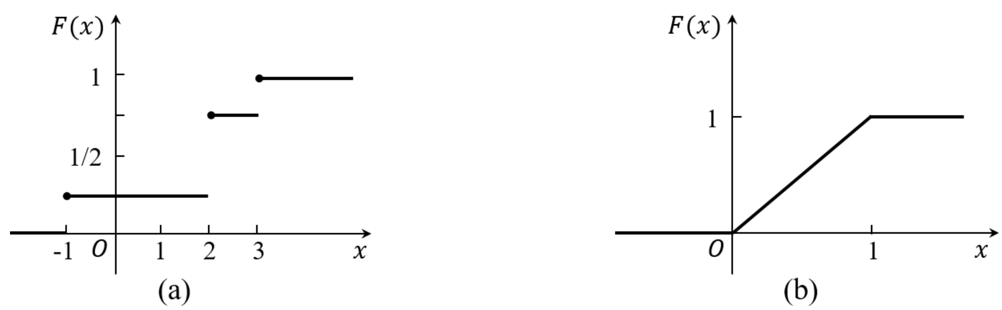


图 4.1 图 (a) 和 (b) 分别给出了例 4.1 和 4.2 的分布函数

例 4.2 在 $[0, 1]$ 区间随机抛一个点, 用 X 表示落点的坐标, 假设 X 落入 $[0, 1]$ 区间内任一子区间的概率与区间长度成正比, 求 X 的分布函数.

解 设随机变量 X 的分布函数为 $F(x)$, 其中 $x \in [0, 1]$, 当 $x < 0$ 时有 $F(x) = 0$; 当 $x > 1$ 时有 $F(x) = 1$. 当 $x \in [0, 1]$ 时有

$$F(x) = P(X \leqslant x) = kx.$$

根据 $F(1) = 1$ 求解可得 $k = 1$. 从而得到 X 的分布函数为

$$F(x) = \begin{cases} 0 & x < 0, \\ x & 0 \leqslant x \leqslant 1, \\ 1 & x > 1. \end{cases}$$

如图 4.1(b) 所示, 分布函数 $F(x)$ 是一条连续的折线.

例 4.3 随机变量 X 的分布函数 $F(x) = A + B \arctan x$, $x \in (-\infty, +\infty)$, 求 $P(X \leqslant 1)$.

解 由分布函数的性质有

$$0 = F(-\infty) = \lim_{x\to-\infty} A + B\arctan x = A - \pi B/2,$$
$$1 = F(+\infty) = \lim_{x\to+\infty} A + B\arctan x = A + \pi B/2,$$

求解可得 $A = 1/2$ 和 $B = 1/\pi$, 从而得到 $P(X \leqslant 1) = 3/4$.

4.2 概率密度函数

离散型随机变量的取值是有限个或可列个离散的单点, 本节研究连续型随机变量, 即随机变量的取值充满整个区间 $[a, b]$ 或 $(a, +\infty)$, 例如火车的到站时间、或一盏灯泡的寿命等. 离散型和连续型随机变量是实际应用中常遇到的两种随机变量.

定义 4.2 设随机变量 X 的分布函数为 $F(x)$, 如果存在可积函数 $f(x)$, 使得对任意实数 x 有

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

则称 X 为 **连续型随机变量**, 函数 $f(x)$ 为随机变量 X 的 **概率密度函数** (probability density function), 简称 **密度函数**.

下面给出概率密度函数的一系列性质:

引理 4.1 概率密度函数 $f(x)$ 满足非负性 $f(x) \geqslant 0$ 和规范性 $\int_{-\infty}^{+\infty} f(t)dt = 1$.

任意概率密度函数必然满足非负性和规范性; 而对满足非负性和规范性的任意函数 $f(x)$, 其必为某个随机变量的密度函数, 并有分布函数为 $F(x) = \int_{-\infty}^{x} f(t)dt$, 密度函数完整地刻画了随机变量的统计规律. 分布函数和密度函数都能刻画连续随机变量的统计规律, 但密度函数在图形上对各种分布特征的显示要优越得多, 比分布函数更常用.

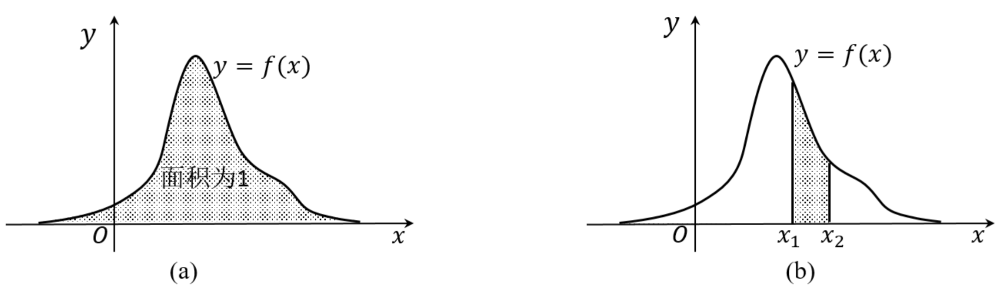


图 4.2 概率密度函数的几何解释

根据规范性可知曲线 $y = f(x)$ 与 x 轴所围成的面积为 1 (如图 4.2(a) 所示). 对任意 $x_1 < x_2$, 有

$$P(x_1 < X \leqslant x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(t)dt \ .$$

由此给出概率密度的几何解释: 随机变量 X 落入区间 $(x_1, x_2]$ 的概率等于由 x 轴, $x = x_1$, $x = x_2$ 和 $y = f(x)$ 所围成的曲边梯形的面积, 如图 4.2(b) 所示.

引理 4.2 对连续随机变量 X, 分布函数 $F(x)$ 在整个实数域上连续; 若密度函数 $f(x)$ 在 x 点连续, 则分布函数 $F(x)$ 在 x 点可导, 且有 $F^{'}(x) = f(x)$.

证明 该引理根据函数的积分性质直接可得: 若函数 $f(x)$ 在实数域上可积, 则积分函数

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

在实数域上连续; 若函数 $f(x)$ 在实数域上连续, 则 $F(x) = \int_{-\infty}^{x} f(t)dt$ 在实数域上可导, 且有 $F'(x) = f(x)$ 成立.

引理 4.3 对任意常数 c 和连续型随机变量 X, 有 $P(X = c) = 0$.

证明 对任意 $\Delta x > 0$ 有事件 $\{X = x\} \subset \{X \in (x - \Delta x, x]\}$, 根据积分中值定理有

$$P(X = x) \leqslant \lim_{\Delta x \to 0} P(x - \Delta x \leqslant X \leqslant x) = \lim_{\Delta x \to 0} \int_{x-\Delta x}^{x} f(t)dt \leqslant \lim_{\Delta x \to 0} f(\xi)\Delta x = 0,$$

其中 $\xi = \arg\max_{x \in (x-\Delta x, x]} f(x)$, 根据概率的非负性完成证明.

根据上面的引理, 一个事件的概率为 0, 不能推出该事件是不可能事件; 一个事件的概率为 1, 也不能推出该事件是必然事件. 此外, 连续随机变量的概率无需强调端点, 因为

$$P(a \leqslant X \leqslant b) = P(a < X < b) = P(a \leqslant X < b) = P(a < X \leqslant b).$$

因为 $f(x) \neq 0 = P(X = x)$, 由此说明概率密度函数不是概率.

若 $f(x)$ 在点 x 连续, 由连续性定义有

$$\lim_{\Delta x \to 0} \frac{P(x - \Delta x \leqslant X \leqslant x + \Delta x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{\int_{x-\Delta x}^{x+\Delta x} f(t)dt}{\Delta x} = \lim_{\Delta x \to 0} \frac{2\Delta x \cdot f(\xi)}{\Delta x} = 2f(x),$$

其中 $\xi \in (x - \Delta x, x + \Delta x)$. 由此可得

$$P(x - \Delta x \leqslant X \leqslant x + \Delta x) \approx 2f(x)\Delta x,$$

若概率密度 $f(x)$ 越大, 则 X 在 x 附近取值的概率越大.

例 4.4 设随机变量 X 的密度函数

$$f(x)=\begin{cases} x & 0<x\leqslant 1 \\ a-x & 1<x<2 \\ 0 & \text{其它}, \end{cases}$$

求其分布函数 $F(x)$.

解 根据概率密度的规范性有

$$1=\int_{-\infty}^{+\infty} f(t)dt=\int_0^1 tdt+\int_1^2 (a-t)dt=a-1,$$

从而求解出 $a=2$, 于是得到具体的密度函数 $f(x)$. 当 $x\leqslant 0$ 时有 $F(x)=0$; 当 $0<x\leqslant 1$ 时, 有

$$F(x)=\int_0^x f(t)dt=x^2/2;$$

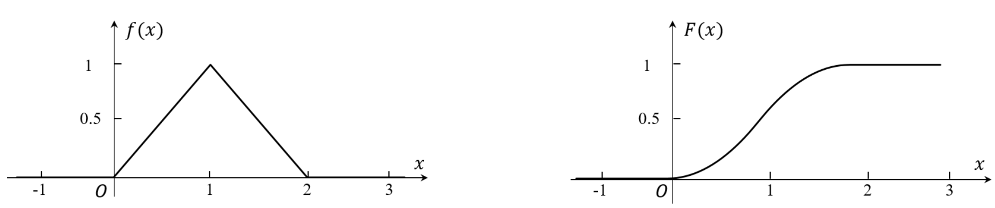
当 $1<x\leqslant 2$ 时, 有

$$F(x)=\int_0^1 f(t)dt+\int_1^x f(t)dt=1/2+\int_1^x (2-t)dt=-x^2/2+2x-1;$$

当 $x\geqslant 2$ 时有 $F(x)=1$. 综合可得

$$F(x)=\begin{cases} 0 & x\leqslant 0, \\ x^2/2 & 0<x\leqslant 1, \\ -x^2/2+2x-1 & 1<x\leqslant 2, \\ 1 & x\geqslant 2. \end{cases}$$

随机变量 X 的密度函数和分布函数如图 4.3 所示.



(a) 概率密度函数 (b) 分布函数

图 4.3 例 4.4 中随机变量 X 的密度函数和分布函数图

例 4.5 对连续随机变量 X, 当 $x \in (0,3)$ 时密度函数 $f(x) = cx^2$, 在其它点的密度函数 $f(x) = 0$. 设随机变量

$$Y = \begin{cases} 2, & X \leqslant 1 \\ X, & X \in (1,2) \\ 1, & X \geqslant 2 \end{cases}$$

求随机变量 Y 的分布函数, 以及计算概率 $P(Y \geqslant X)$.

解 根据概率密度函数的规范性有 $1 = \int_{-\infty}^{+\infty} f(t)dt = 9c$, 由此可得 $c = 1/9$.

用 $F_Y(y)$ 表示随机变量 Y 的分布函数. 当 $y < 1$ 时, 有 $F_Y(y) = P(Y \leqslant y) = 0$; 当 $y \geqslant 2$ 时, 有 $F_Y(y) = P(Y \leqslant y) = 1$; 当 $1 \leqslant y < 2$ 时有

$$\begin{aligned} F_Y(y) &= P(Y \leqslant y) = P(Y = 1) + P(1 < Y \leqslant y) \\ &= P(X \geqslant 2) + P(1 < X \leqslant y) = \int_2^3 t^2/9dt + \int_1^y t^2/9dt = (18 + y^3)/27\ . \end{aligned}$$

由此可得随机变量 Y 的分布函数为

$$F_Y(y) = \begin{cases} 0, & y < 1, \\ (18 + y^3)/27, & y \in [1,2), \\ 1, & y \geqslant 2. \end{cases}$$

可以观察发现随机变量 Y 不是连续型随机变量, 也不是离散型随机变量. 最后计算概率

$$P(X \leqslant Y) = P(X < 2) = \int_0^2 t^2/9dt = 8/27.$$

例 4.6 已知一个靶半径为 2 米的圆盘, 击中靶上任一同心圆盘上的点的概率与该圆盘的面积成正比. 假设射击都能击中靶, 用 X 表示击中点与圆心的距离, 求 X 的概率密度函数.

解 根据题意分析随机变量 X 的分布函数 $F(x)$. 当 $x < 0$ 时有 $F(x) = 0$; 当 $0 \leqslant x \leqslant 2$ 时有

$$F(x) = P(X \leqslant x) = P(0 \leqslant X \leqslant x) = kx^2.$$

根据分布函数的性质有 $F(2) = 1 = 4k$, 求解可得 $k = 1/4$, 进一步得到 X 的概率密度

$$f(x) = \begin{cases} x/2 & 0 \leqslant x \leqslant 2 \\ 0 & \text{其它}. \end{cases}$$

4.3 连续型随机变量的期望和方差

本节研究连续型随机变量的期望和方差, 有利于了解这些变量的整体数字特征.

定义 4.3 设连续随机变量 X 的概率密度函数为 $f(x)$, 若积分 $\int_{-\infty}^{+\infty}|x|f(x)dx$ 收敛, 称 $\int_{-\infty}^{+\infty}xf(x)dx$ 为随机变量 X 的 **期望**, 记为 $E(X)$, 即

$$E(X)=\int_{-\infty}^{+\infty}tf(t)dt.$$

类似于离散型随机变量, 连续型随机变量的期望具有以下一些性质:

引理 4.4 (线性关系) 对任意任意常数 a,b 和连续随机变量 X, 有 $E(aX+b)=aE(X)+b$.

引理 4.5 (Jensen 不等式) 对连续随机变量 X 和函数 $g(x)$,

- 若 $g(x)$ 是凸函数, 则有 $g(E(X))\leqslant E[g(X)]$;
- 若 $g(x)$ 是凹函数, 则有 $g(E(X))\geqslant E[g(X)]$.

对于非负的连续型随机变量, 也可以利用 $P(X>t)$ 来直接计算期望:

引理 4.6 若连续型随机变量 $X\geqslant 0$, 则有

$$E[X]=\int_0^{\infty}P(X>t)dt.$$

该定理对随机变量函数 $Y=g(X)\geqslant 0$ 也成立, 即 $E[g(X)]=\int_0^{\infty}P(g(X)>t)dt$.

证明 设随机变量 X 的概率密度为 $f(x)$, 首先观察得到

$$X=\int_0^X 1dt=\int_0^{+\infty}\mathbb{I}[t<X]dt=\int_0^{+\infty}\mathbb{I}[X>t]dt,$$

这里 $\mathbb{I}[\cdot]$ 表示指示函数, 如果论断为真, 其值为 1, 否则为 0. 两边同时取期望有

$$\begin{aligned}
E[X] &= E\left[\int_0^{+\infty}\mathbb{I}[X>t]dt\right]\\
&= \int_0^{+\infty}\left[\int_0^{+\infty}\mathbb{I}[x>t]f(x)dt\right]dx \qquad \text{(积分换序)}\\
&= \int_0^{+\infty}\left[\int_0^{+\infty}\mathbb{I}[x>t]f(x)dx\right]dt\\
&= \int_0^{+\infty}\left[\int_0^{t}\mathbb{I}[x>t]f(x)dx+\int_t^{+\infty}\mathbb{I}[x>t]f(x)dx\right]dt\\
&= \int_0^{+\infty}\left[\int_t^{+\infty}f(x)dx\right]dt=\int_0^{+\infty}P(X>t)dt.
\end{aligned}$$

对于连续随机变量函数的期望有

定理 4.2 设随机变量 X 的密度函数为 $f(x)$, 且 $\int_{-\infty}^{+\infty} g(t)f(t)dt$ 绝对可积, 则

$$E(g(X)) = \int_{-\infty}^{+\infty} g(t)f(t)dt.$$

该定理表明, 若已知随机变量 X 的密度函数为 $f(x)$, 以及随机变量函数 $Y = g(X)$, 可以直接利用随机变量 X 的密度函数来计算 Y 的期望, 而不需要知道随机变量 Y 的密度函数.

证明 该定理对一般的可积函数 $g(x)$ 均成立, 但证明过程却非常复杂, 这里仅给出非负随机变量函数 $g(x) \geqslant 0$ 的证明. 根据引理 4.6 有

$$\begin{aligned}
E[g(X)] &= \int_0^{+\infty} P(g(X) \geqslant t)dt = \int_0^{+\infty} \int_{x:\ g(x)\geqslant t} f(x)dxdt \\
&= \int_{x:\ g(x)\geqslant 0} \int_0^{g(x)} f(x)dtdx = \int_{x:\ g(x)\geqslant 0} g(x)f(x)dx = \int_{-\infty}^{+\infty} g(x)f(x)dx\ ,
\end{aligned}$$

由此完成证明.

下面介绍物理学中用到的柯西分布 (Cauchy distribution), 它的期望不存在.

例 4.7 设连续随机变量 X 的密度函数为 $f(x) = 1/\pi(1+x^2)$ $(x \in \mathbb{R})$, 求期望 $E(X)$.

因为积分

$$\int_{-\infty}^{+\infty} \frac{|x|}{\pi(1+x^2)}dx = 2\int_0^{+\infty} \frac{x}{\pi(1+x^2)}dx = \frac{1}{\pi}\left[\ln(1+x^2)\right]_0^{+\infty} = +\infty,$$

由此可知期望 $E(X)$ 不存在.

例 4.8 古人运送粮草, 如果早到每天需要的存储费用 c 元, 如果晚到每天需要的延期费用为 C 元. 粮草在运送过程中存在天气、路况等不确定因素, 因此运送需要的天数是随机的, 其概率密度函数为 $f(x)$, 问什么时候出发才能使费用的期望值最小?

解 用随机变量 X 表示实际的运送天数, 分布函数为 $F(x)$. 不妨假设提前了 t 天出发, 那么所需费用为

$$\ell_t(X) = \begin{cases} c(t-X) & X \leqslant t, \\ C(X-t) & X > t. \end{cases}$$

因此可得

$$\begin{aligned}
E[\ell_t(X)] &= \int_0^{+\infty} \ell_t(x)f(x)dx = \int_0^t c(t-x)f(x)dx + \int_t^{+\infty} C(x-t)f(x)dx \\
&= ctF(t) - c\int_0^t xf(x)dx + C\int_t^{+\infty} xf(x)dx - Ct(1-F(t))\ .
\end{aligned}$$

对上式中的 t 求导、并令导数为零可得

$$\frac{d}{dt}E[\ell_t(X)] = cF(t) - C(1 - F(t)) = (c + C)F(t) - C \ .$$

求解可得期望最小的天数 t^* 满足

$$F(t^*) = C/(C + c) \ .$$

定义 4.4 设连续随机变量 X 的概率密度为 $f(x)$, 若 $\int_{-\infty}^{+\infty}(t - E(X))^2 f(t)dt$ 收敛, 称为随机变量 X 的 **方差**, 记为 $\mathrm{Var}(X)$, 即

$$\mathrm{Var}(X) = E(X - E(X))^2 = \int_{-\infty}^{+\infty}(t - E(X))^2 f(t)dt \ .$$

其等价性定义为

$$\mathrm{Var}(X) = E(X - E(X))^2 = E(X^2) - (E(X))^2 = \int_{-\infty}^{+\infty} t^2 f(t)dt - \left(\int_{-\infty}^{+\infty} tf(t)dt\right)^2 .$$

类似于离散型随机变量, 连续型随机变量的方差具有如下性质:

- 对连续型随机变量 X 和常数 a, b, 有 $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$;
- 对连续型随机变量 X 和常数 a, 有 $\mathrm{Var}(X) = E(X - E(X))^2 \leqslant E(X - a)^2$;
- 对连续型随机变量 $X \in [a, b]$, 有 $\mathrm{Var}(X) = (b - E(X))(E(X) - 1) \leqslant (b - a)^2/4$.

4.4 常用连续型随机变量

下面介绍几种常用的连续型随机变量.

4.4.1 均匀分布(uniform distribution)

定义 4.5 若随机变量 X 的概率密度函数为

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

称 X 在区间 $[a, b]$ 上服从 **均匀分布**, 记 $X \sim U(a, b)$.

根据上面的定义很容易发现服从均匀分布的随机变量落入区间任何一点的概率相同. 对任意实数 $x \in \mathbb{R}$ 有 $f(x) \geqslant 0$ 且

$$\int_{-\infty}^{+\infty} f(t)dt = \int_{-\infty}^{a} f(t)dt + \int_{a}^{b} f(t)dt + \int_{b}^{+\infty} f(t)dt = \int_{a}^{b} \frac{1}{b - a}dt = 1 \ .$$

若随机变量 $X \sim U(a,b)$, 则 X 落入内任一子区间 $[x, x+\Delta]$ 的概率

$$P(x \leqslant X \leqslant x+\Delta) = \int_x^{x+\Delta} \frac{1}{b-a} dt = \frac{\Delta}{b-a} \ .$$

该概率与子区间的具体位置 x 无关, 而与子区间长度 Δ 成正比, 由此给出了均匀分布的几何解释: 若随机变量 $X \ \sim U(a,b)$, 则 X 落入 $[a,b]$ 内任一子区间的概率与该区间的长度成正比, 与位置无关.

根据分布函数的定义可知 $X \sim U(a,b)$ 的分布函数为

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

随机变量 $X \sim U(a,b)$ 的密度函数和分布函数的示意图如下:

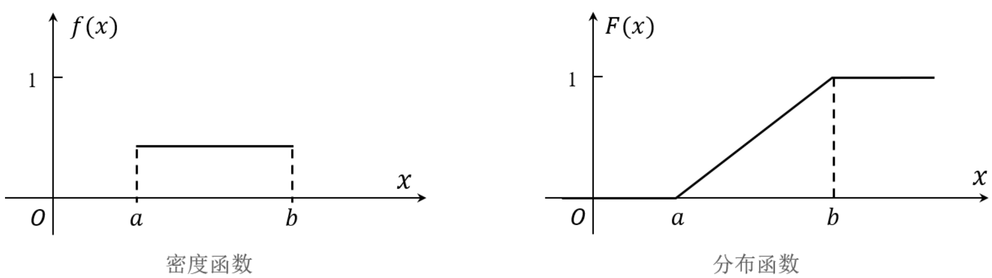


图 4.4 随机变量 $X \sim U(a,b)$ 的密度函数和分布函数

定理 4.3 若随机变量 $X \sim U(a,b)$, 则有

$$E(X) = (a+b)/2 \qquad \text{和} \qquad \mathrm{Var}(X) = (b-a)^2/12 \ .$$

证明 根据期望的定义有

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} t f(t) dt = \frac{1}{b-a} \int_a^b t dt = \frac{a+b}{2} \ , \\ E(X^2) &= \int_{-\infty}^{+\infty} t^2 f(t) dt = \frac{1}{b-a} \int_a^b t^2 dt = \frac{a^2+ab+b^2}{3} \ , \end{aligned}$$

从而得到方差

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

例 4.9 已知随机变量 $X \sim U(a,b)$, 对 $a<c<d<b$, 求 $P(X \leqslant c|X \leqslant d]$.

解 根据条件概率的定义有

$$P(X \leqslant c|X \leqslant d] = \frac{P(\{X \leqslant d\} \cap \{X \leqslant c\})}{P(X \leqslant d)} = \frac{P(X \leqslant c)}{P(X \leqslant d)} = \frac{c-a}{d-a} \ ,$$

即在 $X \leqslant d$ 的条件下, 随机变量 X 服从 $U(a,d)$.

例 4.10 设随机变量 $\xi \sim U(-3,6)$, 试求方程 $4x^2+4\xi x+(\xi+2)=0$ 有实根的概率.

解 易知随机变量 ξ 的概率密度函数

$$f(t) = \begin{cases} 1/9 & x \in [-3,6] \\ 0 & \text{其它.} \end{cases}$$

设事件 A 表示方程有实根, 于是有

$$\begin{aligned} P(A) &= P((4\xi)^2 - 4\times 4\times(\xi+2) \geqslant 0) = P((\xi+1)(\xi-2) \geqslant 0) \\ &= P\left(\{\xi \geqslant -1\} \cap \{\xi \geqslant 2\} \geqslant 0\right) + P\left(\{\xi \leqslant -1\} \cap \{\xi \leqslant 2\} \geqslant 0\right) \\ &= P(\xi \leqslant -1) + P(\xi \geqslant 2) = \int_{-3}^{-1} \frac{1}{9} dt + \int_{2}^{6} \frac{1}{9} dt = \frac{2}{3}. \end{aligned}$$

4.4.2 指数分布

指数分布常用于电话的通话时间和银行的服务等待时间, 也可以用于描述动物和电子元件的寿命, 在可靠性理论和排队论中具有广泛的应用.

定义 4.6 给定常数 $\lambda>0$, 若随机变量 X 的密度函数

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

称 X 服从 **参数为 λ 的指数分布**, 记 $X \sim e(\lambda)$.

对任意实数 x 有密度函数 $f(x) \geqslant 0$, 以及

$$\int_{-\infty}^{+\infty} f(t)dt = \int_{0}^{+\infty} \lambda e^{-\lambda t} dt = \left[-e^{-\lambda t}\right]_0^{+\infty} = 1.$$

根据指数分布的密度函数很容易得到分布函数, 即当 $x \leqslant 0$ 时, 分布函数 $F(x)=0$; 当 $x>0$ 时, 分布函数

$$F(x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}.$$

指数分布的密度函数和分布函数如图 4.5 所示.

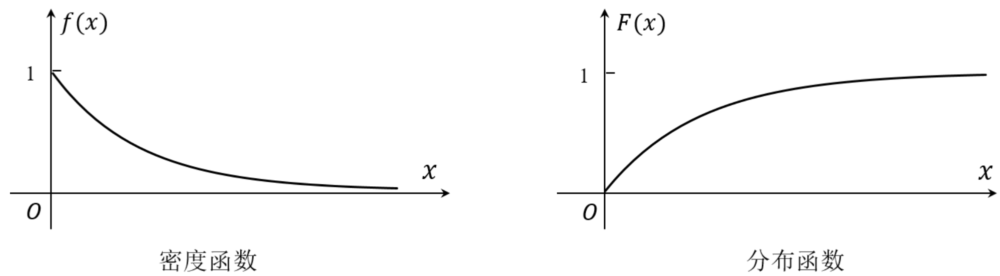


图 4.5 指数分布的密度函数和分布函数

引理 4.7 若随机变量 $X \sim e(\lambda)$, 则 $E(X) = 1/\lambda$ 和 $\mathrm{Var}(X) = 1/\lambda^2$.

证明 根据连续函数的定义有

$$E(X) = \int_0^{+\infty} t\lambda e^{-\lambda t} dt = \left[-te^{-\lambda t}\right]_0^{+\infty} + \int_0^{+\infty} e^{-\lambda t} dt = -\frac{1}{\lambda}\left[e^{-\lambda t}\right]_0^{+\infty} = \frac{1}{\lambda} ,$$

对非负的随机变量 $X \geqslant 0$, 可以利用引理 4.6 和分布函数 $F(x) = P(X \leqslant x) = 1 - e^{-\lambda x}$ 有

$$E[X] = \int_0^{+\infty} P(X > t) dt = \int_0^{+\infty} 1 - F(t) dt = \int_0^{+\infty} e^{-\lambda t} dt = \frac{1}{\lambda} .$$

对于方差, 首先计算

$$E(X^2) = \lambda \int_0^{+\infty} t^2 e^{-\lambda t} dt = \left[-t^2 e^{-\lambda t}\right]_0^{+\infty} + \int_0^{+\infty} 2te^{-\lambda t} dt = \frac{2}{\lambda} E(X) = \frac{2}{\lambda^2} ,$$

于是得到 $\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = 1/\lambda^2$.

下面研究指数分布的一个重要性质: 指数分布的无记忆性.

定理 4.4 若随机变量 $X \sim e(\lambda)$, 则对任意 $s > 0, t > 0$, 有

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

证明 对任意 $x > 0$, 根据分布函数有 $P(X > x) = 1 - F(x) = e^{-\lambda x}$, 从而得到

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

定理得证.

指数分布是唯一具有无记忆性的连续型随机变量, 例如一盏灯泡的寿命 X 服从指数分布, 若已经使用了 s 个小时, 则再使用 t 个小时的概率与已使用过 s 个小时无关, 将这个经历给“忘记”了.

引理 4.8 若随机变量 $X_1, X_2, \cdots, X_n$ 是相互独立的、且分别服从参数为 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 的指数分布, 则有

$$X = \min\{X_1, X_2, \cdots, X_n\} \sim e(\lambda_1 + \lambda_2 + \cdots + \lambda_n) \ .$$

解 这里随机变量的相互独立性可以理解为随机变量取不同值的随机事件相互独立. 计算随机变量 X 的分布函数

$$\begin{aligned} F_X(x) &= P(X \leqslant x) = 1 - P(\min(X_1, X_2, \cdots, X_n) > x) \\ &= 1 - \prod_{i=1}^{n} P(X_i > x) = 1 - \prod_{i=1}^{n} \exp(-\lambda_i x) = 1 - \exp\left(-x \sum_{i=1}^{n} \lambda_i\right) \ , \end{aligned}$$

由此完成证明.

4.4.3 正态分布

正太分布是概率统计中最重要的一种分布, 最早由法国数学家棣莫弗 (De Moivre, 1667-1754) 在 1730s 提出, 用于近似抛硬币试验中随机事件的概率, 即中心极限定理的雏形. 德国数学家高斯 (Gauss, 1777-1855) 在 1800s 首次将正太分布应用于预测天文学中星体的位置, 由此才展示出正太分布的应用价值, 后来发现很多随机现象可以通过正太分布来描述, 正太分布因此被称为高斯分布.

正太分布在概率统计中的重要性主要体现在以下几方面:

- 现实生活中很多随机现象需要用正太分布进行描述, 如人的身高或体重, 某地区的降雨量等;
- 很多分布可以通过正太分布来进行近似计算, 如后面所学的中心极限定理;
- 数理统计中常用的统计分布是由正太分布导出的, 如后面所学的 χ^2 分布、t 分布和 F 分布.

定义 4.7 给定任何实数 u 和 $\sigma > 0$, 若随机变量 X 的概率密度为

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

称随机变量 X 服从 **参数为 (μ, σ^2) 的正态分布** (normal distribution), 又称为 **高斯分布** (Gaussian distribution), 记为 $X \sim \mathcal{N}(\mu, \sigma^2)$.

特别地, 当 $\mu = 0$ 和 $\sigma = 1$ 时的正太分布 $\mathcal{N}(0,1)$ 被称为 **标准正态分布**, 此时密度函数为

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad x \in (-\infty, +\infty) \ .$$

对任意 $x \in (-\infty, +\infty)$ 有 $f(x) \geqslant 0$, 利用极坐标变换 $(x = r\cos\theta, y = r\sin\theta)$ 有

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

$$= \int_0^{2\pi} d\theta \int_0^{+\infty} e^{-\frac{r^2}{2}} r dr = \int_0^{2\pi} d\theta \int_0^{+\infty} e^{-\frac{r^2}{2}} d\frac{r^2}{2} = 2\pi \ ,$$

由此验证了 $\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = 1$, 利用简单的变量替换可验证一般正太分布的密度函数.

关于标准正太分布和一般的正太分布, 有如下关系:

定理 4.5 若随机变量 $X \sim \mathcal{N}(\mu, \sigma^2)$, 则有 $Y = (X-\mu)/\sigma \sim \mathcal{N}(0,1)$; 若随机变量 $X \sim \mathcal{N}(0,1)$, 则有 $Y = \sigma X + \mu \sim \mathcal{N}(\mu, \sigma^2)$.

证明 若随机变量 $X \sim \mathcal{N}(\mu, \sigma^2)$, 则 $Y = (X-\mu)/\sigma$ 的分布函数

$$F_Y(y) = P[Y \leqslant y] = P[X - \mu \leqslant y\sigma] = P[X \leqslant y\sigma + \mu] = \int_{-\infty}^{\mu + y\sigma} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \ .$$

令 $x = (t-\mu)/\sigma$, 代入上面的分布函数有

$$F_Y(y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \ ,$$

由此可证 $Y \sim \mathcal{N}(0,1)$.

另一方面, 若随机变量 $X \sim \mathcal{N}(0,1)$, 则有 $Y = \sigma X + \mu$ 的分布函数

$$F_Y(y) = P(Y \leqslant y) = P(\sigma X + \mu \leqslant y) = P(X \leqslant (y-\mu)/\sigma) = \int_{-\infty}^{(y-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \ .$$

令 $t = (x-\mu)/\sigma$, 代入上面的分布函数有

$$F_Y(y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \ ,$$

由此可证 $Y \sim \mathcal{N}(\mu, \sigma^2)$.

关于正太分布的数字特征有

定理 4.6 若随机变量 $X \sim \mathcal{N}(\mu, \sigma^2)$, 则有 $E(X) = \mu$ 和 $\mathrm{Var}(X) = \sigma^2$; 特别地, 若随机变量 $X \sim \mathcal{N}(0,1)$, 则有 $E(X) = 0$ 和 $\mathrm{Var}(X) = 1$.

正太分布的两个参数分别表示正太分布的期望和方差.

证明 这里仅仅证明标准正太分布的期望为 0 和方差为 1, 结合定理 4.5 直接可得 $X \sim \mathcal{N}(\mu, \sigma^2)$ 的期望和方差. 若随机变量 $X \sim \mathcal{N}(0,1)$, 根据奇函数在对称的区间上积分为 0 有

$$E(X) = \int_{-\infty}^{+\infty} \frac{t}{\sqrt{2\pi}} e^{-t^2/2} dt = 0 \ .$$

根据方差的定义和分部积分有

$$\mathrm{Var}(X) = \int_{-\infty}^{+\infty} \frac{t^2}{\sqrt{2\pi}} e^{-t^2/2} dt = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t d e^{-t^2/2} = \left[\frac{te^{-t^2/2}}{-\sqrt{2\pi}}\right]_{t=-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-t^2/2} dt = 1 \ .$$

由此完成证明.

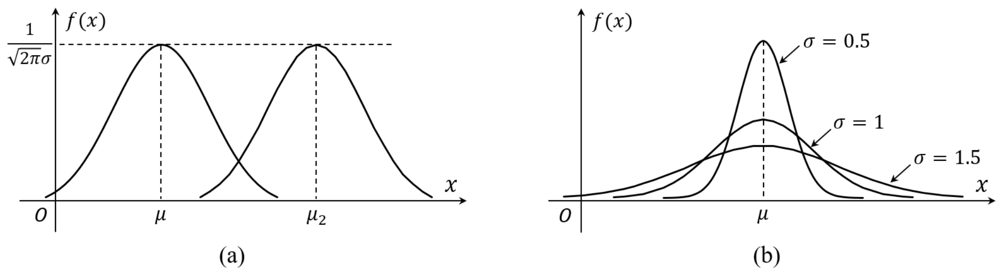


图 4.6 正太分布的密度函数

正太分布的密度函数如图 4.7(a) 所示, 具有以下一些特点:

1) 曲线 $f(x)$ 关于 $x = \mu$ 对称, 先单调递增, 之后单调递减, 在 $x = \mu$ 处取最大值 $1/\sqrt{2\pi}\sigma$. 说明随机变量 X 的取值主要集中在 $x = \mu = E(X)$ 附近, 离 $x = \mu$ 越远的区间概率越小.

2) 根据 $\lim_{x\to\pm\infty} f(x) = 0$ 可知曲线 $f(x)$ 以 x 轴为渐近线; 根据 $f''(x) = 0$ 可知曲线 $f(x)$ 的拐点为 $x = \mu \pm \sigma$.

3) 固定标准差 σ 而改变期望 μ 的值, 曲线 $f(x)$ 形状不变, 仅沿 x 轴左右平行移动, 如图 4.7(a).

4) 固定期望 μ 而改变标准差 σ 的值, 曲线 $f(x)$ 的对称点不变, 但最大值 $1/\sqrt{2\pi}\sigma$ 和拐点 $x = \mu \pm \sigma$ 发生了改变. 如图 4.7(b) 所示: 当 σ 越小, 曲线顶峰越高, 曲线越陡峭, 分布越集中, 方差越小; 反之 σ 越大, 曲线顶峰越地, 曲线越平坦, 分布越分散, 方差越大.

关于正太分布的概率估计, 有下面的不等式:

定理 4.7 若 $X \sim \mathcal{N}(0,1)$, 对任意 $\epsilon > 0$ 有

$$\begin{aligned} P(X \geqslant \epsilon) &\leqslant \frac{1}{2} e^{-\epsilon^2/2} \\ P(|X| \geqslant \epsilon) &\leqslant \min\left\{1, \sqrt{\frac{2}{\pi}} \frac{1}{\epsilon} e^{-\epsilon^2/2}\right\}. \end{aligned}$$

在上面的定理中, 第一个不等式具有广泛的应用, 在 $\epsilon \in (0,1)$ 时对真实的概率有更好的估计; 第二个不等式被称为 Mill 不等式, 在 $\epsilon \in (1,+\infty)$ 时对真实的概率有更好的估计.

证明 针对第一个不等式, 我们有

$$
\begin{aligned}
P(X \geqslant \epsilon) &= \int_{\epsilon}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = \int_{0}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-(x+\epsilon)^2/2} dx \\
&\leqslant e^{-\epsilon^2/2} \int_{0}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \frac{1}{2} e^{-\epsilon^2/2} \ .
\end{aligned}
$$

对于 Mill 不等式, 根据 $\mathcal{N}(0,1)$ 的概率密度 $f(x) = e^{-x^2/2}/\sqrt{2\pi}$ 有 $f'(x) = -xf(x)$, 进一步可得

$$
\begin{aligned}
P(|X| \geqslant \epsilon) &= 2\int_{\epsilon}^{+\infty} f(t) dt = 2\int_{\epsilon}^{+\infty} \frac{tf(t)}{t} dt \\
&\leqslant 2\int_{\epsilon}^{+\infty} \frac{tf(t)}{\epsilon} dt = -2\int_{\epsilon}^{+\infty} \frac{f'(t)}{\epsilon} dt = -\frac{2}{\epsilon} \left[f(t)\right]_{\epsilon}^{+\infty} = \frac{2}{\sqrt{2\pi}\epsilon} e^{-\epsilon^2/2} \ .
\end{aligned}
$$

由此完成证明.

若随机变量 $X \sim \mathcal{N}(\mu, \sigma^2)$, 则有分布函数

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

该分布函数没有显示的表达式, 只能求数值解, 函数如图 4.7(a) 所示. 为便于研究正太分布的分布函数, 利用定理 4.5 可将其它正太分布都转化为标准正太分布 $\mathcal{N}(0,1)$, 设其分布函数为

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

表 4.1 给出了标准正太分布 $\Phi(x)$ 的函数表, 在计算具体的概率时可供查询. 下面给出关于分布函数 $\Phi(x)$ 的一些性质:

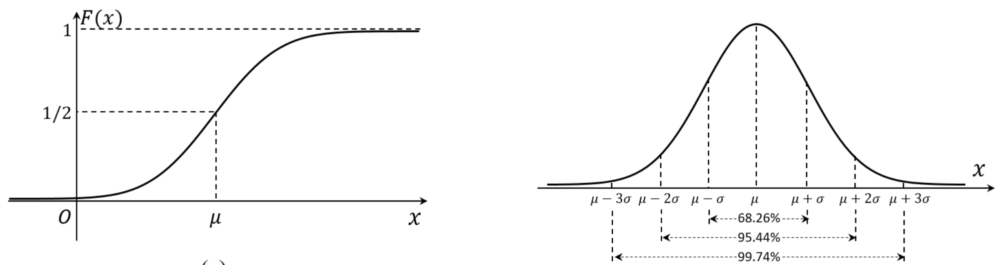
1) 根据对称性有 $\Phi(x) + \Phi(-x) = 1$.

2) 若随机变量 $X \sim \mathcal{N}(\mu, \sigma^2)$, 则对任意实数 $a < b$ 有

$$
\begin{aligned}
P(X < a) &= P\left(\frac{X-\mu}{\sigma} \leqslant \frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{a-\mu}{\sigma}\right) \ , \\
P(X > b) &= 1 - P\left(\frac{X-\mu}{\sigma} \leqslant \frac{b-\mu}{\sigma}\right) = 1 - \Phi\left(\frac{b-\mu}{\sigma}\right) \ , \\
P(a \leqslant X \leqslant b) &= P\left(\frac{a-\mu}{\sigma} \leqslant \frac{X-\mu}{\sigma} \leqslant \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \ .
\end{aligned}
$$

3) 若随机变量 $X \sim \mathcal{N}(\mu, \sigma^2)$, 则对任意实数 $k > 0$ 有

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



(a) (b)

图 4.7 正太分布函数和 3σ 原则

特别的, 当 $k = 1, 2, 3$ 时通过表 4.1 有

$$P(|x-\mu| < \sigma) = 0.6826, \quad P(|x-\mu| < 2\sigma) = 0.9544, \quad P(|x-\mu| < 3\sigma) = 0.9974 \ .$$

如图 4.7(b) 所示, 尽管随机变量 $X \sim \mathcal{N}(\mu, \sigma^2)$ 的取值范围为整个实数域 $\mathbb{R}$, 但其取值落在 $[\mu - 3\sigma, \mu + 3\sigma]$ 之外的概率不超过千分之三, 也就是 $X \sim \mathcal{N}(\mu, \sigma^2)$ 的取值几乎总在 $[\mu - 3\sigma, \mu + 3\sigma]$ 之内, 这就是人们所说的“3σ 原则”, 在实际的统计推断, 特别是产品质量检测中具有重要的应用.

4) 若随机变量 $X \sim \mathcal{N}(\mu, \sigma^2)$, 且已知 $P(X < c) = p$, 则有

$$p = P(X < c) = P\left(\frac{X-\mu}{\sigma} < \frac{c-\mu}{\sigma}\right) = \Phi\left(\frac{c-\mu}{\sigma}\right) \ ,$$

由此可反解出 $c = \mu + \sigma\Phi^{-1}(p)$. 这里 $\Phi^{-1}(x)$ 表示标准正太分布函数 $\Phi(x)$ 的反函数, 可根据表 4.1 由里向外查得, 例如 $\Phi^{-1}(0.5871) = 0.22$.

例 4.11 已知某公司员工每个月的工资服从正太分布 $\mathcal{N}(6000, \sigma^2)$, 问题:

i) 若已知标准差 $\sigma = 500$, 求工资在 5000 与 7000 之间的员工在公司中占比多少?

ii) 当标准差 σ 为何值时, 工资在 5000 与 7000 之间的员工在公司中占比为 0.803?

解 用随机变量 X 表示公司员工每个月的工资, 则 $X \sim \mathcal{N}(6000, \sigma^2)$. 针对问题 i), 当 $\sigma = 500$ 时通过查询表 4.1 有

$$P(5000 \leqslant X \leqslant 7000) = P\left(-2 \leqslant \frac{X - 6000}{500} \leqslant 2\right) = \Phi(2) - \Phi(-2) = 2\Phi(2) - 1 = 0.9544.$$

针对问题 ii) 有

$$P(5000 \leqslant X \leqslant 7000) = P\left(-\frac{1000}{\sigma} \leqslant \frac{x - 6000}{\sigma} \leqslant \frac{1000}{\sigma}\right) = 2\Phi(\frac{1000}{\sigma}) - 1 = 0.803 \ ,$$

于是得到 $\Phi(1000/\sigma) = 0.9015$, 通过由内自外查表 4.1 有 $\sigma \approx 775.2$.

表 4.1 标准正态分布表 $\Phi(x) = \int_{-\infty}^{x} e^{-t^2/2}/\sqrt{2\pi}dt.$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

4.5 连续随机变量函数的分布

当知道一个随机变量的概率分布后, 经常会考虑它的一些函数的分布, 例如已知一个圆的直径 X 服从均匀分布 $U(a,b)$, 则可以考虑圆的面积 $Y=\pi(X/2)^2$ 的分布. 一般地, 若已知随机变量 X 的分布函数, 以及 $g(x)$ 是定义在随机变量 X 所有可能取值的函数, 则称 $Y=g(X)$ 为随机变量 X 的函数, 很显然 Y 也是随机变量. 研究的问题可以归纳为: 若已知随机变量 X 的概率分布和函数 $g(x)$, 如何求解随机变量 $Y=g(X)$ 的概率分布.

设离散型随机变量 X 的分布列为 $p_k = P(X = x_k)$ $(k = 1, 2, \cdots)$, 求解随机变量 $Y = g(X)$ 的分布列较为简单, 将相等的项 $g(x_i) = g(x_j)$ 合并, 相应的概率相加即可. 例如

例 4.12 若随机变量 X 的概率分布列为 $P(X = k) = 1/2^k$ $(k = 1, 2, \cdots)$, 求随机变量 $Y = \cos(\pi X/2)$ 的分布列.

解 当 $n = 1, 2, \cdots$ 时有

$$\cos(k\pi/2) = \begin{cases} 1 & k = 4n \\ 0 & k = 2n-1 \\ -1 & k = 4n-2 \ , \end{cases}$$

所以 $\cos(\pi X/2)$ 的取值为 $\{-1, 0, +1\}$, 其概率分别为

$$\begin{aligned} P(Y = 1) &= \sum_{n=1}^{\infty} P(X = 4n) = \sum_{n=1}^{\infty} \frac{1}{2^{4n}} = \frac{1}{15} \ , \\ P(Y = 0) &= \sum_{n=1}^{\infty} P(X = 2n-1) = \sum_{n=1}^{\infty} \frac{1}{2^{2n-1}} = \frac{2}{3} \ , \\ P(Y = -1) &= \sum_{n=1}^{\infty} P(X = 4n-2) = \sum_{n=1}^{\infty} \frac{1}{2^{4n-2}} = \frac{4}{15} \ , \end{aligned}$$

由此给出随机变量 Y 的分布列.

针对连续型随机变量, 一般采用概率密度函数来刻画其概率分布. 若已知函数 $g(x)$ 和随机变量 X 的概率密度函数为 $f_X(x)$, 求解随机变量 $Y = g(X)$ 的概率密度 $f_Y(y)$, 通常分为以下两步:

1) 求解随机变量 Y 的分布函数 $F_Y(y) = P(Y \leqslant y) = P(g(X) \leqslant y) = \int_{g(x)\leqslant y} f_X(x)dx$;
2) 求解随机变量 Y 的密度函数 $f_Y(y) = F'_Y(y)$.

求解此类问题常用到的数学工具是积分求导公式: 设函数 $F(y) = \int_{\psi(y)}^{\varphi(y)} f(x)dx$, 则有

$$F'(y) = f(\varphi(y))\varphi'(y) - f(\psi(y))\psi'(y) \ . \tag{4.2}$$

下面来看两个例子.

例 4.13 设随机变量 X 的概率密度为 $f_X(x)$, 求 $Y = X^2$ 的概率密度.

解 根据分布函数的定义可知

$$F_Y(y) = P(Y \leqslant y) = P(X^2 \leqslant y) \ .$$

当 $y \leqslant 0$ 时有 $F_Y(y) = 0$; 当 $y > 0$ 时有

$$F_Y(y) = P(X^2 \leqslant y) = P(-\sqrt{y} \leqslant x \leqslant \sqrt{y}) = \int_{-\sqrt{y}}^{\sqrt{y}} f_X(x)dx \ ,$$

根据 (4.2) 得到密度函数

$$f_Y(y) = F'_Y(y) = f_X(\sqrt{y})/2\sqrt{y} + f_X(-\sqrt{y})/2\sqrt{y} \ .$$

综上所述有

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

例 4.14 设随机变量 X 的密度函数为

$$f(x) = \begin{cases} x/8 & 0 < x < 4 \\ 0 & \text{其它}, \end{cases}$$

求 $Y = 2X + 8$ 的密度函数.

解 首先求解分布函数

$$F_Y(y) = P(Y \leqslant y) = P(2X + 8 \leqslant y) = P(X \leqslant (y-8)/2) = F_X((y-8)/2) \ ,$$

求导可得密度函数

$$f_Y(y) = f_X((y-8)/2)/2 = \begin{cases} (y-8)/32 & (y-8)/2 \in [0,4] \\ 0 & \text{其它} \end{cases} = \begin{cases} (y-8)/32 & y \in [8,16] \\ 0 & \text{其它} \ . \end{cases}$$

针对一般情况有如下定理:

定理 4.8 设随机变量 X 的概率密度是定义在实数域上的函数 $f_X(x)$, 函数 $y = g(x)$ 处处可导且严格单调 (即 $g'(x) > 0$ 或 $g'(x) < 0$), 令其反函数 $x = g^{-1}(y) = h(y)$, 则 $Y = g(X)$ 的概率密度为

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

其中 $\alpha = \min\{g(-\infty), g(+\infty)\}$ 和 $\beta = \max\{g(-\infty), g(+\infty)\}$.

可将上述定理推广至区间函数 $x \in [a,b]$, 上述定理依旧成立, 此时有 $\alpha = \min\{g(a), g(b)\}$ 和 $\beta = \max\{g(a), g(b)\}$.

证明 证明思路与前面的例题类似, 这里不妨假设 $g'(x) > 0$, 同理可以考虑 $g'(x) < 0$) 的情况. 根据 $g'(x) > 0$ 可知其反函数 $x = h(y)$ 也严格单调, 且 $g(x) \in [\alpha, \beta]$. 因此, 当 $y \leqslant \alpha$ 时, 有 $F_Y(y) = 0$; 当 $y \geqslant \beta$ 时有 $F_Y(y) = 1$; 当 $\alpha < y < \beta$ 时,

$$F_Y(y) = P(g(X) < y) = P(X \leqslant h(y)) = F(h(y)).$$

于是有随机变量 Y 的概率密度

$$f_Y(y) = F'(h(y)) \cdot h'(y) = f_X(h(y)) \cdot h'(y).$$

根据$x = h(y)$ 严格单调可知 $h'(y) > 0$.

定理 4.9 设 $X \sim \mathcal{N}(\mu, \sigma^2)$, 则 $Y = aX + b$ $(a > 0)$ 服从正太分布 $\mathcal{N}(a\mu + b, a^2\sigma^2)$.

证明 设函数 $g(x) = ax + b$, 可得 $\alpha = -\infty$, $\beta = +\infty$, 以及 $y = g(x)$ 的反函数为

$$x = h(y) = (y - b)/a,$$

且有 $h'(y) = 1/a$. 根据定理 4.8 可知

$$f_Y(y) = \frac{1}{a} f_X(\frac{y-b}{a}) = \frac{1}{a} \frac{1}{\sqrt{2\pi}\sigma} e^{-(\frac{y-b}{a} - \mu)^2)^2/2\sigma^2} = \frac{1}{\sqrt{2\pi}a\sigma} e^{-(y-b-a\mu)^2/2a^2\sigma^2},$$

由此证明了 $Y = aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

例 4.15 设随机变量 X 的分布函数 $F_X(x)$ 是严格单调的连续函数, 则 $Y = F_X(X) \sim U(0, 1)$.

证明 令 $Y = F_X(X)$ 的分布函数为 $G(y)$, 则

$$G(y) = P(Y \leqslant y) = P(F_X(X) \leqslant y) \ .$$

因为分布函数 $F_X(x) \in [0, 1]$, 当 $y < 0$ 时有 $G(y) = 0$; 当 $y \geqslant 1$ 时有 $G(y) = 1$; 当 $y \in [0, 1]$ 时, 由于 $F_X(X)$ 严格单调, 所以 $F_X^{-1}(y)$ 存在且严格单调, 于是有

$$G(y) = P(F_X(X) \leqslant y) = P\left(X \leqslant F_X^{-1}(y)\right) = F_X\left(F_X^{-1}(y)\right) = y \ .$$

综上所述有密度函数

$$f_Y(y) = \begin{cases} 1 & y \in [0, 1] \\ 0 & \text{其它} \ . \end{cases}$$

4.6 常用分布的随机数*

随机数在计算机仿真学和密码学等领域具有广泛的应用, 通过产生大量的随机数据来实现对真实世界的模拟. 可以利用物理随机过程来产生真实的随机数, 例如反复抛掷硬币、骰子、抽签、摇号等, 这些方法可以得到质量很高的随机数, 但其数量和类型通常较少、难以满足实际的需求.

现在的主流方法是使用计算机产生伪随机数, 通过计算机的确定算法来生成“类似”真随机数的伪随机数. 由于算法给出的结果总是确定的, 所以伪随机数并不是真正的随机数, 但是好的伪随机数序列与真实随机数序列表现几乎相同, 很难进行区分.

本节首先介绍如何生成在 $(0,1)$ 上均匀分布的随机数, 然后根据此随机数构造其它常用分布的随机数. 这里仅给出具体的构造方法, 关于其中的原理, 有兴趣的读者可以参考相关书籍.

4.6.1 区间 $(0,1)$ 上均匀分布的随机数

目前有很多方法生成 $(0,1)$ 上均匀分布的随机数, 这里介绍最常用的线性同余法. 通过下面的迭代方式产生一系列随机数 $x_1, x_2, \cdots, x_k$,

$$x_i = ax_{i-1} + c \qquad \bmod \quad m\ ,$$

其中 $1 \leqslant i \leqslant k \leqslant m$, 常数 x_0 为初始给定的种子. 为了获得较好的随机性, 通常 m 的取值应足够大, 如 $m = 2^k$ $(k = 31, 63)$, 常数 c 与 m 互质, 常数 $a-1$ 被 m 的因子整除, 例如一种可行的选择是

$$x_i = 31415926x_{i-1} + 453806245 \qquad \bmod \quad 2^{31}.$$

最后得到在区间 $(0,1)$ 上均匀分布的随机数 $x_1/m, x_2/m, \cdots, x_k/m$. 很多时候会根据实际情况选择不同的初始种子 x_0. 有了在 $(0,1)$ 上均匀分布的随机数, 则可以构造一些服从常用分布的随机数.

4.6.2 常用离散型分布的随机数

定理 4.10 若随机变量 $X \sim U(0,1)$, 以及 $F(y)$ 是某个离散型随机变量的分布函数, 其可能的取值为 $\{y_1, y_2, \cdots\}$, 不妨假设 $y_1 < y_2 < \cdots$. 设随机变量

$$Y = \begin{cases} y_1 & X \leqslant F(y_1) \\ y_i & X \in (F(y_{i-1}), F(y_i)]\ (i \geqslant 2)\ , \end{cases}$$

则 Y 的分布函数 $F_Y(y) = F(y)$ (如图 4.8 所示).

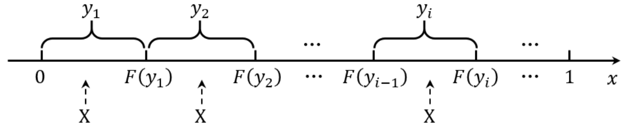


图 4.8 离散分布的随机数生成

证明　当 $y < y_1$ 时有

$$F_Y(y) = P(Y \leqslant y) \leqslant P(Y < y_1) = 0 = F(y)\ .$$

当 $y_1 \leqslant y < y_2$ 时根据 $X \sim U(0,1)$ 有

$$F_Y(y) = P(Y \leqslant y) = P(Y = y_1) = P(X \leqslant F(y_1)) = F(y_1) = F(y)\ .$$

当 $y \geqslant y_2$ 时再次根据 $X \sim U(0,1)$ 有

$$F_Y(y) = P(Y \leqslant y) = \sum_{y_i\colon y_i \leqslant y} P(Y = y_i) = \sum_{y_i\colon y_i \leqslant y} P\left(X \in (F(y_{i-1}), F(y_i)]\right) = \max_{y_i\colon y_i \leqslant y} F(y_i) = F(y)\ .$$

定理 4.10 说明可以利用均匀分布的随机数来构造其它离散型随机变量的随机数, 以下以伯努利分布为例, 可类似生成其它常用分布的随机数.

例 4.16　已知在区间 $(0,1)$ 上均匀分布的随机数 $x_1, x_2, \cdots, x_k\ (k \geqslant 1)$, 如何生成参数为 p 的伯努利分布的随机数.

解　设随机变量 Y 服从参数为 p 的伯努利分布, 则 Y 的分布函数 $F(0) = 1 - p$ 和 $F(1) = 1$. 根据定理 4.10, 当 $1 \leqslant i \leqslant k$ 时构造随机数

$$y_i = \begin{cases} 0 & x_i \leqslant F(0) = 1 - p \\ 1 & x_i \in (F(0), F(1)] = (1-p, 1]\ , \end{cases}$$

其服从参数 p 的伯努利分布.

4.6.3 常用连续型分布的随机数

定理 4.11　若随机变量 $X \sim U(0,1)$, 且 $F(y)$ 是某一个连续的分布函数, 很显然反函数 $F^{-1}(y)$ 存在, 则随机变量

$$Y = F^{-1}(X)$$

的分布函数为 $F_Y(y) = F(y)$.

证明　根据分布函数的定义和 $F(x)$ 的单调性有

$$F_Y(y) = P(Y \leqslant y) = P(F^{-1}(X) \leqslant y) = P(X \leqslant F(y))\ .$$

再根据分布函数 $F(y)$ 的非负性和随机变量 $X \sim U(0,1)$ 有

$$F_Y(y) = F(y)\ ,$$

由此完成证明.

例 4.17 若已知在区间 $(0,1)$ 上均匀分布的随机数 $x_1, x_2, \cdots, x_k$ $(k \geqslant 1)$, 如何生成参数为 $\lambda = 4$ 的指数分布的随机数.

解 若随机变量 Y 服从参数为 $\lambda = 4$ 的指数分布, 则有连续的分布函数 $F_Y(y) = 1 - e^{-4\lambda y}$, 以及反函数 $F_Y^{-1}(y) = -\ln(1-y)/4$. 根据定理 4.11 可构造随机数

$$y_i = -\ln(1 - x_i)/4 \qquad (i \in [k])$$

服从参数为 $\lambda = 4$ 的指数分布.

由于正太分布的分布函数不存在显示表达式, 所以它的反函数也不存在显示表达式, 因此不能利用定理 4.11 来直接生成服从正太分布的随机数. 这里简要介绍一种生成标准正太分布的随机数的方法、以及一些相关的结论, 详细的证明可参考相关书籍.

设 X 和 Y 是相互独立的标准正太分布随机变量, 则它们的极坐标

$$\begin{cases} R = \sqrt{X^2 + Y^2} \\ \theta = \arctan(Y/X) \end{cases}$$

相互独立, 且 R^2 服从参数为 $1/2$ 的指数分布, 而 θ 服从在 $(0, 2\pi)$ 上的均匀分布. 设相互独立的随机变量 X_1 和 X_2 在区间 $(0,1)$ 上服从均匀分布, 则有

$$R = (-2\ln X_1)^{1/2} \qquad \text{和} \qquad \theta = 2\pi X_2\ .$$

由此可得相互独立的标准正太分布随机变量

$$X = R\cos(\theta) = (-2\ln X_1)^{1/2}\cos(2\pi X_2) \quad \text{和} \quad Y = R\sin(\theta) = (-2\ln X_1)^{1/2}\sin(2\pi X_2)\ ,$$

这种产生标准正太分布随机变量的方法被称为 **Box-Muller 方法**.

习题

4.1 若随机变量 X 的取值区间为 $[1,c]$, 且落入 $[1,c]$ 任意小区间的概率与小区间的长度成正比, 求 X 的分布函数.

4.2 用随机变量 X 表示某银行从下午开始营业起到第一个顾客到达的等待时间 (分), 设 X 的分布函数为

$$F(x)=\begin{cases}1-c\exp(-x/8) & x\geqslant 0\\ 0 & x<0\end{cases}$$

求这些事件的概率: 1) $P(X\leqslant 4)$; 2) $P(X\geqslant 8)$; 3) $P(4\leqslant X\leqslant 8)$; 4) $P(X\leqslant 4$ 或$X\geqslant 8)$; 5) $P(X=6)$.

4.3 若随机变量 X 的分布函数

$$F(x)=\begin{cases}0 & x<1\\ \ln x & 1\leqslant x<e\\ 1 & x\geqslant e,\end{cases}$$

求随机变量 X 的密度函数.

4.4 若随机变量 X 的密度函数

$$f(x)=\begin{cases}x & x\in[0,1)\\ c-x & x\in[1,2)\\ 0 & \text{其它},\end{cases}$$

求随机变量 X 的分布函数, 并画出分布函数和密度函数.

4.5 已知长方形的宽服从均匀分布 $U(0,2)$ (单位: 米), 以及长方形的面积为 10 (单位: 平方米), 求长方形的周长的期望与方差.

4.6 设随机变量 X 的概率密度为

$$f(x)=\begin{cases}Ae^{-x} & x>0\\ 0 & x\leqslant 0.\end{cases}$$

求 $Y=e^{-2X}$ 的期望.

4.7 已知随机变量 $X\sim U(0,1)$, 对任意 $\lambda>0$ 求 $E[\lambda^{\max(X,1-X)}]$.

4.8 电池的故障是电动汽车的核心问题, 设相继两次事故之间的时间 T 服从参数为 1/40 的指数分布, 求概率 $P(X > 45)$, 以及求最小的 τ 使得 $P(X > t) \geqslant 60\%$.

4.9 设乘客在一公交车站等待公交车的时间服从参数为 1/6 的指数分布, 某乘客若等待时间超过 10 分钟则换乘出租车离开. 该乘客一个月内有 10 天乘公交站 (每天是否乘出租车相互独立), 用 Y 表示该乘客因未等到公交车而换乘出租车的次数, 求 Y 的分布函数.

4.10 设随机变量 X 服从瑞利分布, 其概率密度为

$$f(x) = \begin{cases} \frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)} & x > 0 \\ 0 & x \leqslant 0 \ , \end{cases}$$

求期望 $E(X)$ 和方差 $\mathrm{Var}(X)$.

4.11 设随机变量 $X \sim \mathcal{N}(4, 49)$, 问题: i) 求概率 $P(3 < X \leqslant 7)$, $P(|X| \leqslant 4)$ 和 $P > 6$; ii) 求常数 c 使得 $P(X > c) = P(X \leqslant c)$; iii) 求常数 a 至多有多大时满足 $P(X > d) \geqslant 0.9$.

4.12 设一批产品的寿命 $X \sim \mathcal{N}(180, \sigma^2)$ $(\sigma > 0)$, 要保证有 $P(140 < X \leqslant 220) \geqslant 0.9$ 成立, 则 σ 最大值是多少.

4.13 证明

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

4.14 若 $X \sim N(0, 1)$,对任意实数 $\epsilon > 0$, 求证

$$P(X \geqslant \epsilon) \geqslant \frac{1}{3} e^{-\frac{(\epsilon+1)^2}{2}}$$

4.15 已知随机变量 X 的概率密度函数 $f_X(x)$, 求随机变量 $Y = |X|$ 的概率密度 $f_Y(y)$.

4.16 若随机变量 $X \sim \mathcal{N}(\mu, \sigma^2)$, 求随机变量 $Y = e^X$ 的密度函数.

4.17 若随机变量 $X \sim \mathcal{N}(0, 1)$, 求随机变量 $Y = X^2$ 和 $Z = 2X^2 + 1$ 的密度函数.

4.18 若随机变量 $X \sim e(\lambda)$, 求随机变量 $Y = aX$ $(a > 0)$ 的概率分布.

4.19 若随机变量 $X \sim U(0, 1)$, 求随机变量 $Y = -\ln X$ 和 $Z = e^X$ 的概率分布.

4.20 设随机变量 X 的密度函数为

$$f(x) = \begin{cases} 2x/\pi & x \in (0, \pi) \\ 0 & \text{其它} \ . \end{cases}$$

求 $Y = \sin X$ 的密度函数.

4.21 **大作业:** 编程生成 100000 在 $(0,1)$ 区间上均匀分布的随机数, 并以此生成伯努利分布、二项分布、泊松分布、几何分布、指数分布、正太分布的随机数, 需要将各种分布的参数作为输入变量, 最后查资料验证方法的正确性.