第 4 章　连续型随机变量

4.1 概念与性质

4.1.1 分布函数

定义 4.1　给定任意随机变量 X 和 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**, 分布函数的本质是概率.

对任意实数 $x_1 < x_2$, 有

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

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: $F(x + 0) = \lim_{\Delta x \to 0^+} F(x + \Delta x) = F(x)$.

任一分布函数必满足上述三性质, 而满足上述三性质的函数必是某随机变量的分布函数, 因此分布函数可由上述三性质完全刻画.

可利用分布函数 $F(x)$ 表示随机事件的概率, 例如

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

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, 有

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

当 $-1 \leqslant x < 2$ 时, 有

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

当 $2 \leqslant x < 3$ 时, 有

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

当 $x \geqslant 3$ 时有 $F(x) = 1$.

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

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

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

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

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

4.1.2 概率密度函数

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

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

成立, 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**, 简称 **概率密度**.

根据分布函数的性质可得到概率密度函数的性质:

性质 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, 而对任意满足非负性和规范性的函数函数 $f(x)$, 可引入新的随机变量 X, 其分布函数为 $G(x) = \int_{-\infty}^{x} f(t)dt$. 由此说明概率密度函数可由函数的非负性和规范性完全刻画.

对任意 $x_1 \leqslant x_2$, 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 x 轴, $x = x_1$, $x = x_2$ 和 $y = f(x)$ 所围成的曲边梯形的面积.

定理 4.1 对连续随机变量 X, 其分布函数 $F(x)$ 在整个实数域上连续; 若 $f(x)$ 在 x 点连续, 则 $F(x)$ 在 x 点可导, 且 $F'(x) = f(x)$.

证明 根据函数的积分性质: 若 $f(x)$ 在 $[a, b]$ 上可积, 则 $\phi(x) = \int_a^x f(t)dt$ 在 $[a, b]$ 上连续. 若 $f(x)$ 在 $[a, b]$ 上连续, 则 $\phi(x) = \int_a^x f(t)dt$ 在 $[a, b]$ 上可导, 且 $\phi'(x) = f(x)$.

性质 4.2 对连续型随机变量 X 和常数 $x \in (-\infty, +\infty)$, 有 $P(X = x) = 0$.

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.

由此可知, 连续随机变量无需考虑端点, 即

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

同时说明了概率密度函数不是概率, 即 $P(X = x) = 0 \neq f(x)$.

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 设连续随机变量 X 的密度函数

$$f(x)=\begin{cases}c(4x-2x^2) & 0<x<2\\ 0 & \text{其它},\end{cases}$$

求概率 $P(X>1)$.

解 根据概率密度函数的规范性有

$$1=\int_{-\infty}^{+\infty}f(t)dt=\int_0^2 c(4t-2t^2)dt=\frac{8}{3}c,$$

得到 $c=3/8$, 所以

$$P(X>1)=\int_1^{\infty}f(t)dt=\int_1^2 f(t)dt=\int_1^2\frac{3}{8}(4t-2t^2)dt=\frac{1}{2}.$$

例 4.5 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

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

求解可得 $a=2$, 于是有

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

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

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

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

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

4.1.3 连续随机变量的期望和方差

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 $\int_{-\infty}^{+\infty} xf(x)dx$ 绝对收敛, 称 $\int_{-\infty}^{+\infty} xf(x)dx$ 为随机变量 X 的 **期望**, 记为 $E(X)$, 即

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

与离散性随机变量类似, 连续随机变量的期望具有如下一些列性质:

性质 4.3 (线性关系) 对任意任意常数 a, b 和随机变量 X, 有

$$E(aX + b) = aE(X) + b;$$

对常数 $c_1, \dots, c_n$ 和连续函数 $g_1(x), \dots, g_n(x)$, 有

$$E\left(\sum_{i=1}^{n} c_i g_i(X)\right) = \sum_{i=1}^{n} c_i E(g_i(X)).$$

性质 4.4 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

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

性质 4.5 (Jensen 不等式) 对连续随机变量 X 和凸函数 $f(x)$ 有

$$f(E(X)) \leqslant E[f(X)];$$

对连续随机变量 X 和凹函数 $f(x)$ 有

$$f(E(X)) \geqslant E[f(X)].$$

例 4.7 设随机变量 X 的密度函数为 $f(x) = \begin{cases} cx & x \in [0,1] \\ 0 & \text{其它} \end{cases}$, 求 $E(X^m)$ (m 为正整数).

解 根据概率密度函数的规范性 $\int_{-\infty}^{+\infty} f(t)dt = 1$, 求解可得 $c = 2$, 进一步有

$$E(X^m) = \int_0^1 t^m \cdot 2tdt = 2\int_0^1 t^{m+1}dt = \frac{2}{m+2}.$$

在很多实际问题中可能不知道随机变量的分布, 因此不能直接计算期望, 但可以估计概率 $P(X > t)$ 来计算期望:

定理 4.2 对非负随机变量 X, 有

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

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

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

例 4.8 利用此定理计算例 4.7 中随机变量 X 的期望 $E(X) = 2/3$.