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

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

其等价性定义为

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

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 $\mathrm{Var}(X)\leqslant(b-E(X))(E(X)-a)\leqslant(b-a)^2/4$.

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

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

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 $[x,x+\Delta]\subseteq[a,b]$ 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子区间的长度成正比, 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:

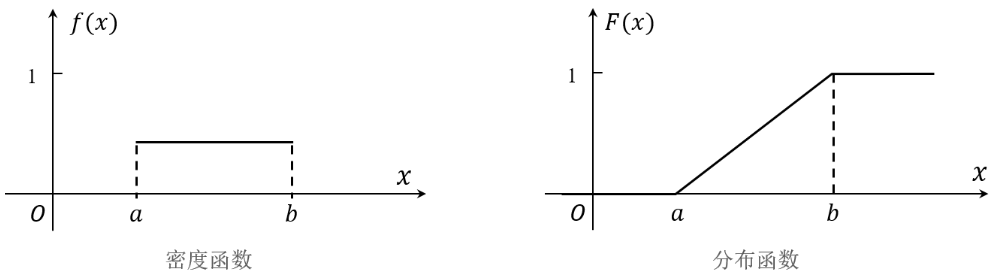


图 4.4 随机变量 $X \sim U(a,b)$ 的密度函数和分布函数

性质 4.8 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

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

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

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

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 $X\sim U(-3,6)$, 试求方程 $4t^2+4Xt+(X+2)=0$ 有实根的概率.

解 易知随机变量 X 的概率密度函数

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

方程 $4t^2+4Xt+(X+2)=0$ 有实根等价于 $16X^2-16(X+2)\geqslant 0$, 因此所求概率为

$$\begin{aligned}&P(16X^2-16(X+2)\geqslant 0)=P((X+1)(X-2)\geqslant 0)\\ =\;& P\left(\{X\geqslant -1\}\cap\{X\geqslant 2\}\right)+P\left(\{X\leqslant -1\}\cap\{X\leqslant 2\}\right)\\ =\;& P(X\leqslant -1)+P(X\geqslant 2)=\int_{-3}^{-1}\frac{1}{9}dt+\int_{2}^{6}\frac{1}{9}dt=\frac{2}{3}.\end{aligned}$$

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 $X\sim e(\lambda)$.

首先有密度函数 $f(x)\geqslant 0$, 以及

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

进一步得到指数分布随机变量的分布函数 $F(x)$: 当 $x\leqslant 0$ 时 $F(x)=0$; 当 $x>0$ 时

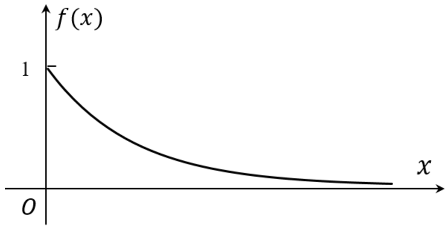
$$F(x)=\int_0^x\lambda e^{-\lambda t}dt=1-e^{-\lambda x}.$$

指数分布的密度函数和分布函数如图 4.5 所示.

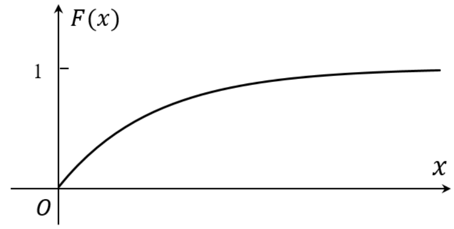
性质 4.9 若随机变量 $X\sim e(\lambda)$, 则 $E(X)=1/\lambda$ 和 $\mathrm{Var}(X)=1/\lambda^2$.

证明 根据连续型随机变量期望的定义有

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



密度函数



分布函数

图 4.5 指数分布的密度函数和分布函数

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

由此可得 $\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = 1/\lambda^2$.

下面研究指数分布的一个重要性质: 无记忆性.

定理 4.2 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 给定任意常数 $x > 0$, 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, 将这个记忆给“忘记”了.

例 4.11 若随机变量 X 服从参数为 λ 的指数分布 $(\lambda > 0)$, 而且 X 落入区间 $(2, 3)$ 的概率达到最大值, 求概率 $P(X > 5|X > 3)$.

解 根据指数分布的定义有

$$P(2 < X < 3) = \int_2^3 \lambda e^{-\lambda t}dt = e^{-2\lambda} - e^{-3\lambda}.$$

设 $h(\lambda) = e^{-2\lambda} - e^{-3\lambda}$, 求导可得

$$h'(\lambda) = 3e^{-3\lambda} - 2e^{-2\lambda} = e^{-2\lambda}(3e^{-\lambda} - 2).$$

求解 $h'(\lambda)=0$ 可得 $\lambda=\ln(3/2)$. 容易得到在 $(-\infty,\ln(3/2))$ 上导数 $h'(\lambda)>0$, 而在 $(\ln(3/2),+\infty)$ 上导数 $h'(\lambda)<0$, 可知 $h(\lambda)$ 在 $\lambda=\ln(3/2)$ 取得最大值. 再根据指数分布的无记忆性有

$$P(X>5|X>3)=P(X>2)=e^{-2\lambda}=4/9.$$

4.4.3 正态分布

正态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 首次将正态分布应用于预测天文学中星体的位置, 由此才展示出正态分布的应用价值, 后来发现利用正态分布可以描述很多的随机现象, 正态分布也被称为高斯分布.

正态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态分布进行描述, 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态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态分布导出的, 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

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

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 $\sigma=1$ 的正态分布 $\mathcal{N}(0,1)$ 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

$$\begin{aligned}\left(\int_{-\infty}^{+\infty}e^{-\frac{x^2}{2}}dx\right)^2 &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}e^{-\frac{x^2+y^2}{2}}dxdy\\ &= \int_0^{2\pi}d\theta\int_0^{+\infty}e^{-\frac{r^2}{2}}rdr=\int_0^{2\pi}d\theta\int_0^{+\infty}e^{-\frac{r^2}{2}}d\frac{r^2}{2}=2\pi,\end{aligned}$$

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

关于标准正态分布和一般的正态分布, 有如下关系:

定理 4.3 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 $Y=\sigma X+\mu\sim\mathcal{N}(\mu,\sigma^2)$.