Ch 4.3 连续型随机变量的期望和方差

回顾前一次课

分布函数：$F(x) = P(X \le x)$

- 单调性: 若$x_1 < x_2$, 则$F(x_1) \le F(x_2)$
- 规范性: $F(x) \in [0,1]$且$F(-\infty) = 0, F(+\infty) = 1$
- 右连续性: $F(x+0) = F(x)$

连续随机变量、概率密度函数：$F(x) = \int_{-\infty}^{x} f(t)dt$

- 非负性：$f(x) \ge 0$
- 规范性：$\int_{-\infty}^{+\infty} f(t)dt = 1$

连续随机变量分布函数$F(x)$与$f(x)$的连续、可导关系

连续型随机变量$P(X = x) = 0$

连续函数的期望

设连续随机变量X的密度函数为$f(x)$, 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

$$E(X)=\int_{-\infty}^{+\infty}xf(x)dx$$

连续函数期望的性质

- 对任意常数a, b和连续随机变量X, 有 $E(aX+b)=aE(X)+b$
- 对常数$c_1, \ldots, c_n$和连续函数$g_1(x), \ldots, g_n(x)$, 有

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

- 对连续随机变量X和凸函数$f(x)$有$f\big(E(X)\big) \leq E\big(f(X)\big)$
- 对连续随机变量X和凹函数$f(x)$有$f\big(E(X)\big) \geq E\big(f(X)\big)$

非负随机变量期望的等价定义

定理： 对非负随机变量X, 有

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

推论: 对非负随机变量$g(X)$, 有$E[\mathrm{g}(X)] = \int_0^{+\infty} P(g(X) > t)dt$

随机变量函数的期望

定理: 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, 则随机变量$Y = g(X)$的期望

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

例题

物理学中用到的柯西分布 (Cauchy distribution)

设随机变量X的密度函数为

$$f(x) = \frac{1}{\pi(1+x^2)} \quad (x \in \mathbb{R})$$

求期望$E(X)$

例题

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，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?

方差

设连续随机变量X的概率密度函数为$f(x)$,称

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

为随机变量X的方差

等价定义

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

方差的性质

- 对常数a, b和连续随机变量X, 有$\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$

- 对连续型随机变量X和常数a,

$$\mathrm{Var}(X) = E\big(X - E(X)\big)^2 \le (X-a)^2$$

- 对连续型随机变量$X \in [a, b]$, 有

$$\mathrm{Var}(X) = \big(b - E(X)\big)(E(X) - a) \le (b-a)^2/4$$

Ch 4.4 常用连续型随机变量

均匀分布

若随机变量X的概率密度为

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

称$\boldsymbol{X}$服从区间$[\boldsymbol{a},\boldsymbol{b}]$上的**均匀分布**, 记$\boldsymbol{X}\sim\boldsymbol{U}(\boldsymbol{a},\boldsymbol{b})$

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

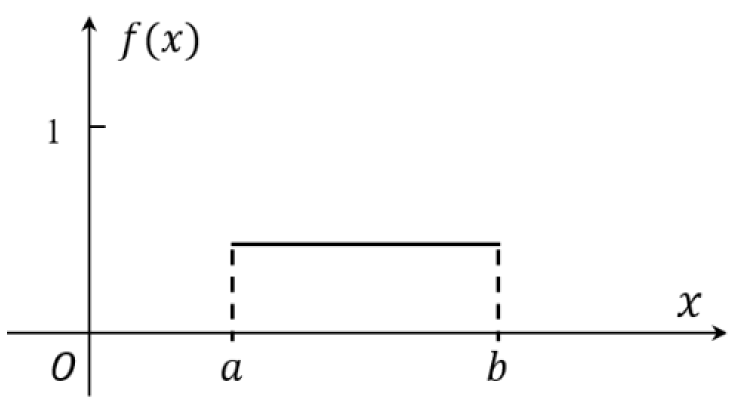
$$P(x\le X\le x+\Delta)=\int_x^{x+\Delta}\frac{1}{b-a}dt=\frac{\Delta}{b-a}$$

几何解释: 若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

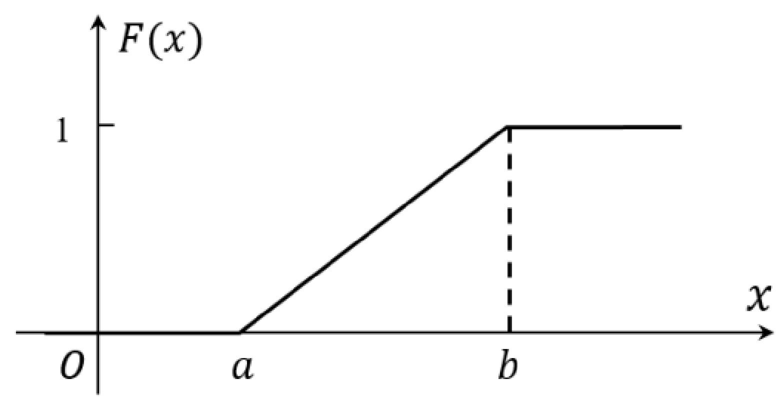
均匀分布

$X \sim U(a, b)$ 的分布函数为

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



密度函数



分布函数

均匀分布的期望与方差

若$X \sim U(a,b)$，则

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

例题

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

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

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

给定常数$\lambda > 0$, 若随机变量X的密度函数

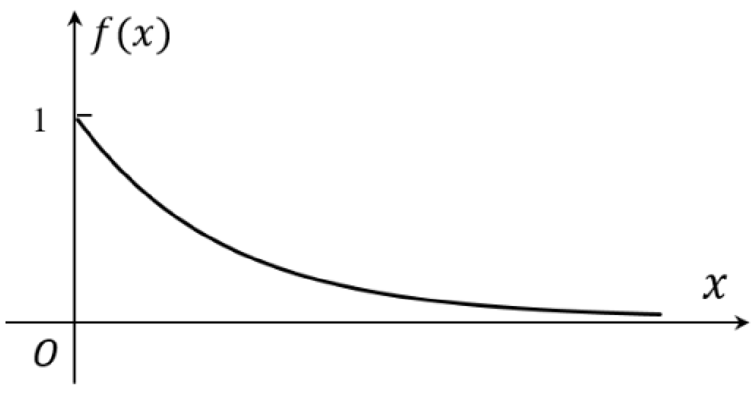
$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{其它} \end{cases}$$

称X服从参数为λ的指数分布, 记 $X \sim e(\lambda)$

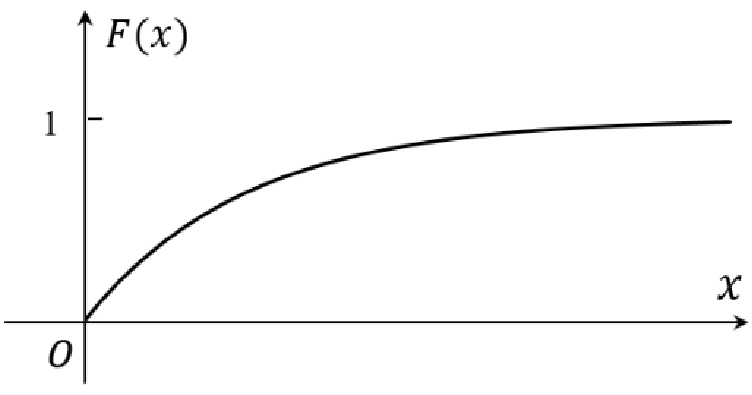
指数分布的分布函数

分布函数：当$x \le 0$时有$F(x) = 0$，当$x > 0$时有

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



指数分布的密度函数



指数分布的分布函数

指数分布的期望与方差

若随机变量$X \sim e(\lambda)$, 则

$$E(X) = \frac{1}{\lambda} \qquad \text{和} \qquad \mathrm{Var}(X) = \frac{1}{\lambda^2}$$

指数分布的无记忆性

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

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

指数分布是唯一具有无记忆性的连续型随机变量

例题

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

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