第一章随机变量及其信息度量

第二节凸函数

陈兴同

中国矿业大学 数学学院

2021 年 8 月

内容提要

内容提要

内容提要

凸函数定义 1.2.1

一元凸函数是指在某区间上图像向下或向上弯曲的一元函数。

设 $y = f(x), x \in D = (a,b) \subset R$ 是一元函数，

（1） 如果 $\forall x_1, x_2 \in D, \lambda \in [0,1]$ 总有

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2),$$

则称 $f(x)$ 为区间 D 上的凸函数；如果 $-f(x)$ 为凸函数，则称 $f(x)$ 为区间 D 上的凹函数。

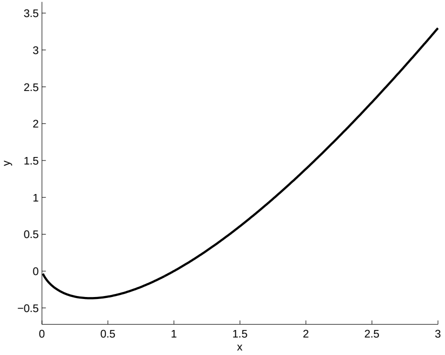
（2） 如果 $\forall x_1, x_2 \in D, x_1 \neq x_2, \lambda \in (0,1)$ 总有

$$f(\lambda x_1 + (1-\lambda)x_2) < \lambda f(x_1) + (1-\lambda) f(x_2),$$

则称 $f(x)$ 为区间 D 上的严格凸函数；如果 $-f(x)$ 为严格凸函数，则称 $f(x)$ 为区间 D 上的严格凹函数。

凸函数图形

常用凸函数有：$f(x)=x^2, f(x)=e^x, f(x)=x\log x(x>0), f(x)=\log x(x>0), f(x)=\sqrt{x}(x\geq 0)$.



凸函数判定

(定理)1.2.1

（1） 设函数 $f(x)$ 在区间 $[a,b]$ 上二阶连续可微，如果对任何 $x\in[a,b]$ 有 $f''(x)\geq 0$，则 f(x) 是区间 $[a,b]$ 上的凸函数。

（2） 设函数 $f(x)$ 在区间 $[a,b]$ 上二阶连续可微，如果对任何 $x\in[a,b]$ 有 $f''(x)>0$，则 f(x) 是区间 $[a,b]$ 上的严格凸函数。

（3） 设函数 $f(x)$ 在区间 $[a,b]$ 上二阶连续可微，如果对任何 $x\in[a,b]$ 有 $f''(x)\leq 0$，则 f(x) 是区间 $[a,b]$ 上的凹函数。

（4） 设函数 $f(x)$ 在区间 $[a,b]$ 上二阶连续可微，如果对任何 $x\in[a,b]$ 有 $f''(x)<0$，则 f(x) 是区间 $[a,b]$ 上的严格凹函数。

凸集合

(定义)1.2.2

*设 $\mathcal{G}$ 为 n 维欧氏空间 R^n 中的一个非空集合，如果对于 $\mathcal{G}$ 中任意两个点 $x^{(1)}, x^{(2)}$，连接它们的线段仍属于 $\mathcal{G}$，即 $\forall\lambda \in [0,1]$ 有 $\lambda x^{(1)} + (1-\lambda)x^{(2)} \in \mathcal{G}$，则称 $\mathcal{G}$ 为 R^n 中的一个**凸集合**。如果它又是闭集，则称为**闭凸集**。*

常见凸集合

- 超平面：$H=\{x\in R^n|p^Tx=\alpha\},p\in R^n,\alpha\in R$。
- 半空间：$H^-=\{x\in R^n|p^Tx\leq\alpha\},p\in R^n,\alpha\in R$。
- 射线：$L=\{x\in R^n|x=x^{(0)}+\lambda d,\lambda\geq 0\},x^{(0)},d\in R^n$。

凸函数定义1.2.3

设 $\mathcal{G}$ 为 R^n 中的非空凸集合，$f(x)=f(x_1,\cdots,x_n)$ 是定义在 $\mathcal{G}$ 上的 n 元实函数。

（1） 若对 $\forall x^{(1)},x^{(2)}\in\mathcal{G},\forall\lambda\in[0,1]$ 有 $f(\lambda x^{(1)}+(1-\lambda)x^{(2)})\leq\lambda f(x^{(1)})+(1-\lambda)f(x^{(2)})$，则称 $f(x)$ 为 $\mathcal{G}$ 上的 n 元凸函数。

（2） 若 $-f(x)$ 为 $\mathcal{G}$ 上的 n 元凸函数，则称 $f(x)$ 为 $\mathcal{G}$ 上的 n 元凹函数。

（3） 若对 $\forall x^{(1)},x^{(2)}\in\mathcal{G},x^{(1)}\neq x^{(2)},\forall\lambda\in(0,1)$ 有 $f(\lambda x^{(1)}+(1-\lambda)x^{(2)})<\lambda f(x^{(1)})+(1-\lambda)f(x^{(2)})$，则称 $f(x)$ 为 $\mathcal{G}$ 上的 n 元严格凸函数。

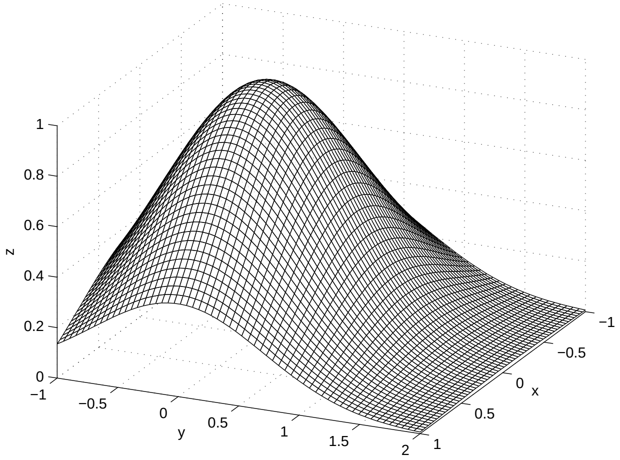
（4） 若 $-f(x)$ 为 $\mathcal{G}$ 上的 n 元严格凸函数，则称 $f(x)$ 为 $\mathcal{G}$ 上的 n 元严格凹函数。

又凹又凸函数

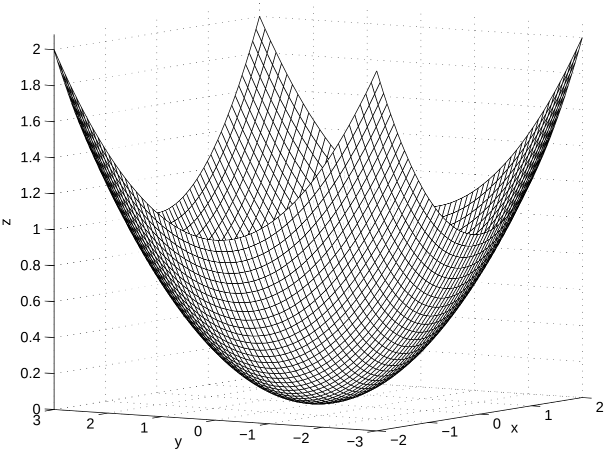
注意 ***超平面函数*** $f(x) = a_1x_1 + a_2x_2 + \cdots + a_nx_n = \mathbf{a}^T x$ ***是不凸不凹函数，可根据需要规定它的凹凸性。***

二元凸函数图形

凹



二元凸函数图形



定理 1.2.2：Jensen 不等式

（1）设 $f(x)$ 是区间 $[a,b]$ 上的凸函数，$x_1,x_2,\cdots,x_n\in[a,b]$，并且 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 构成概率分布律，那么

$$f(\lambda_1x_1+\lambda_2x_2+\cdots+\lambda_nx_n)\le\lambda_1f(x_1)+\lambda_2f(x_2)+\cdots+\lambda_nf(x_n).$$

（2）设有限离散型随机变量 X 的取值空间是 $\mathcal{X}$，如果 $f(x)$ 是区间 $D\supseteq\mathcal{X}$ 上的一个凸函数，则 $f(E(X))\le E\left[f(X)\right]$。

（3）设 $f(x)$ 是区间 $[a,b]$ 上的严格凸（凹）函数，$x_1,x_2,\cdots,x_n\in[a,b]$，并且 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 构成概率分布律。若

$$f(\lambda_1x_1+\lambda_2x_2+\cdots+\lambda_nx_n)=\lambda_1f(x_1)+\lambda_2f(x_2)+\cdots+\lambda_nf(x_n),$$

则或者 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 构成退化分布，或者 $x_1=x_2=\cdots=x_n$。

练习：

证明上面定理 1.2.2（1）和（3）。

定理 1.2.3：概率分布不等式

如果向量 $(p_1, p_2, \cdots, p_n), (q_1, q_2, \cdots, q_n)$ 是两个概率分布，且 $q_i > 0, i = 1, 2, \cdots, n$，则有不等式

$$\sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} \geq 0 \text{ 或} \sum_{i=1}^{n} p_i \log p_i \geq \sum_{i=1}^{n} p_i \log q_i \qquad 1.12 \; \cancel{(3.1)}$$

并且等号成立当且仅当两个概率分布相同即 $p_i = q_i, i = 1, 2, \cdots, n$。

证明：

易证明函数 $f(x)=x\log x=x\ln x\log e, x\geq 0$ 是严格凸函数，如图 1-1a 。现在取

$$x_i=\frac{p_i}{q_i}, i=1,2,\cdots,n,$$

则由 Jensen 不等式得

$$\begin{aligned}\sum_{i=1}^{n} p_i\log\frac{p_i}{q_i}&=\sum_{i=1}^{n} q_i\frac{p_i}{q_i}\log\frac{p_i}{q_i}=\sum_{i=1}^{n} q_i f(x_i)\\ \geq\ & f\left(\sum_{i=1}^{n} q_i x_i\right)=f\left(\sum_{i=1}^{n} p_i\right)=f(1)=0.\end{aligned}$$

又因为 $(q_1,q_2,\cdots,q_n)$ 不是退化分布，根据定理 **??**(3)，等号成立当且仅当 $x_1=x_2=\cdots=x_n$，这就证明了 $p_i=q_i, i=1,2,\cdots,n$。

定理 1.2.4：对数和不等式

设 $(a_1, a_2, \cdots, a_n)$ 是非负向量，而 $(b_1, b_2, \cdots, b_n)$ 是正向量，则

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq (\sum_{i=1}^{n} a_i) \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i}, \qquad (\cancel{3.2}\ 1.13)$$

其中等号成立当且仅当这两个向量成比例。

证明:

由于函数 $f(x)=x\log x, x\geq 0$ 是严格凸函数，取

$$\lambda_i=\frac{b_i}{\sum_{j=1}^n b_i}, x_i=\frac{a_i}{b_i},$$

则 $\lambda_i>0, \sum_{i=1}^n \lambda_i=1, x_i\geq 0$，故由 Jensen 不等式

$$\sum_{i=1}^n \lambda_i f(x_i)\geq f(\sum_{i=1}^n \lambda_i x_i),$$

证明续：

得到

$$\sum_{i=1}^{n} \frac{b_i}{\sum_{j=1}^{n} b_j} \frac{a_i}{b_i} \log \frac{a_i}{b_i} \geq \sum_{i=1}^{n} \left(\frac{b_i}{\sum_{j=1}^{n} b_j} \frac{a_i}{b_i} \right) \log \sum_{i=1}^{n} \left(\frac{b_i}{\sum_{j=1}^{n} b_j} \frac{a_i}{b_i} \right),$$

$$\sum_{i=1}^{n} \frac{a_i}{\sum_{j=1}^{n} b_j} \log \frac{a_i}{b_i} \geq \sum_{i=1}^{n} \left(\frac{a_i}{\sum_{j=1}^{n} b_j} \right) \log \sum_{i=1}^{n} \left(\frac{a_i}{\sum_{j=1}^{n} b_j} \right),$$

两边同乘常量 $\sum_{j=1}^{n} b_j$ 得结果。由于函数 $f(x)$ 是严格凸函数，根据定理 1.2.2(3)，等号成立当且仅当所有 x_i 全相等，这就证明了等号成立的充要条件。

练习：

在此定理中可能会遇到 $0\log\frac{0}{q}, 0\log 0$，*故规定* $0\log\frac{0}{q} = 0\log 0 = 0$。

证明：$\lim_{x\to 0^+} x\log x = 0$