

0x00359141, 而单精度浮点数 3510593.0 的十六进制表示为 0x4A564504。推导出这个浮点表示, 并解释整数和浮点数表示的位之间的关系。

习题 2.49

- A. 对于一种具有 n 位小数的浮点格式, 给出不能准确描述的最小正整数的公式 (因为要想准确表示它需要 $n+1$ 位小数)。假设阶码字段长度 k 足够大, 可以表示的阶码范围不会限制这个问题。
- B. 对于单精度格式 ($n=23$), 这个整数的数字值是多少?

2.4.4 舍入

因为表示方法限制了浮点数的范围和精度, 所以浮点运算只能近似地表示实数运算。因此, 对于值 x , 我们一般想用一种系统的方法, 能够找到“最接近的”匹配值 x' , 它可以用期望的浮点形式表示出来。这就是舍入 (rounding) 运算的任务。一个关键问题是在两个可能值的中间确定舍入方向。例如, 如果我有 1.50 美元, 想把它舍入到最接近的美元数, 应该是 1 美元还是 2 美元呢? 一种可选择的方法是维持实际数字的下界和上界。例如, 我们可以确定可表示的值 x^- 和 x^+ , 使得 x 的值位于它们之间: $x^- \leq x \leq x^+$ 。IEEE 浮点格式定义了四种不同的舍入方式。默认的方法是找到最接近的匹配, 而其他三种可用于计算上界和下界。

图 2-37 举例说明了四种舍入方式, 将一个金额数舍入到最接近的整数美元数。向偶数舍入 (round-to-even), 也被称为向最接近的值舍入 (round-to-nearest), 是默认的方式, 试图找到一个最接近的匹配值。因此, 它将 1.40 美元舍入成 1 美元, 而将 1.60 美元舍入成 2 美元, 因为它们是最接近的整数美元值。唯一的设计决策是确定两个可能结果中间数值的舍入效果。向偶数舍入方式采用的方法是: 它将数字向上或者向下舍入, 使得结果的最低有效数字是偶数。因此, 这种方法将 1.5 美元和 2.5 美元都舍入成 2 美元。

方式	1.40	1.60	1.50	2.50	-1.50
向偶数舍入	1	2	2	2	-2
向零舍入	1	1	1	2	-1
向下舍入	1	1	1	2	-2
向上舍入	2	2	2	3	-1

图 2-37 以美元舍入为例说明舍入方式 (第一种方法是舍入到一个最接近的值, 而其他三种方法向上或向下限定结果, 单位为美元)

其他三种方式产生实际值的确界 (guaranteed bound)。这些方法在一些数字应用中是很有用的。向零舍入方式把正数向下舍入, 把负数向上舍入, 得到值 \hat{x} , 使得 $|\hat{x}| \leq |x|$ 。向下舍入方式把正数和负数都向下舍入, 得到值 x^- , 使得 $x^- \leq x$ 。向上舍入方式把正数和负数都向上舍入, 得到值 x^+ , 满足 $x \leq x^+$ 。

向偶数舍入初看上去好像是个相当随意的目标——有什么理由偏向取偶数呢? 为什么不始终把位于两个可表示的值中间的值都向上舍入呢? 使用这种方法的一个问题就是很容易假想到这样的情景: 这种方法舍入一组数值, 会在计算这些值的平均数中引入统计偏差。我们采用这种方式舍入得到的一组数的平均值将比这些数本身的平均值略高一些。相反, 如果我们总是把两个可表示值中间的数字向下舍入, 那么舍入后的一组数的平均值将比这些数本身的平均值略低一些。向偶数舍入在大多数现实情况中避免了这种统计偏差。