Representación de Números y Errores

Ing. Pablo Antuña

18 de julio de 2016

1. Representación de Números en una Computadora

Es un hecho conocido que los números naturales 1, 2, 3,... son infinitos. También es un hecho conocido que cualquier computadora, por más cara y poderosa que sea, tiene una capacidad limitada de almacenamiento. Una conclusión directa de ésto es que es imposible que una computadora pueda representar todos los números que existen. Querer representar todos los números en una computadora es como querer escribir todos los números en una hoja de papel: algo imposible, por más grande que sea la hoja. De aquí surgen las siguientes preguntas:

- ¿Cuáles son los números que son manejados por una computadora?
- ¿Cómo se representan estos números?

Vamos a responder estas preguntas de acuerdo al tipo de número.

1.1. Números Enteros

Un número entero es representado como un número binario, es decir, una secuencia de 0's y 1's. Cada dígito binario se llama *bit*. Dependiendo de la arquitectura de la computadora, se utilizan hasta 32 o 64 bits. Por lo tanto, se pueden representar 2^{32} o 2^{64} números enteros. Para la mayoría de los propósitos prácticos, esta cantidad es suficiente.

1.2. Números Reales

La representación de números reales es más complicada que la de los enteros. Esto es debido a que existen muchos más números reales que números enteros. Por ejemplo, existen infinitos números reales entre 0 y 1. Por lo tanto, en una computadora se utiliza solamente un subconjunto bien pequeño de los números reales.

La mayoría de las computadoras modernas basan su representación de números reales en el **Standard IEEE 754**, que es una normativa de la llamada aritmética en **punto flotante**. De acuerdo a éste standard, se representa un número real con 64 bits. El primer bit es un indicador de signo. Le sigue un exponente de 11 bits, denominado **característica**, y una fracción binaria de 52 bits, llamada **mantisa**. La base para el exponente es 2.

No es interesante para nosotros hacer un estudio exhaustivo de la aritmética en punto flotante. Sin embargo, los hechos que un ingeniero debe saber son los siguientes:

1. En una computadora son representables solamente una cantidad finita de números reales.
2. Un número real que no sea representable exactamente es aproximado por uno que sí sea representable. Ésto introduce un error, llamado **error de redondeo**.
3. Como solamente una cantidad finita de números son representables en una computadora, existe un número máximo y un número mínimo. Cuando en un cálculo aparece un número mayor que el máximo representable se produce un **overflow**, y la consecuencia es que se detienen los cálculos. Cuando en un cálculo aparece un número menor al mínimo representable, se produce un **underflow**, y generalmente este número pequeño se iguala a cero.
4. **Importante:** No es válido preguntar a la computadora si dos números reales son iguales. Es decir, nunca se debe preguntar si A=B cuando A y B son números reales. Ésto es debido a que la computadora usa solamente aproximaciones de A y B, no sus valores exactos, por lo tanto la respuesta que dé será inservible.
5. Por otro lado, sí es válido preguntar a la computadora si dos números enteros son iguales, ya que en este caso sí se cuenta con una representación exacta: una secuencia de ceros y unos que se puede comparar con otra.
6. Los números reales que son representables de forma exacta en una computadora están mayormente concentrados en las cercanías de 0. Por ejemplo, hay más números representables entre 0 y 1 que entre 15 y 16. Por lo tanto, se tendrá mayor precisión en los cálculos si se utilizan números pequeños.

2. Conceptos de Errores

Definición 1 *Si* p_0 *es una aproximación de* p*, el* ***error absoluto*** *es* $|p - p_0|$ *y el* ***error relativo*** *es* $\frac{|p-p_0|}{|p|}$*, siempre que* $p \neq 0$.

Como una medida de precisión, el error absoluto puede llevar a confusiones. Por otro lado, el error relativo es más significativo, porque toma en cuenta el tamaño del valor.

Definición 2 *El número* p_0 *aproxima a* p *con* t ***cifras significativas*** *si* t *es el mayor entero no negativo para el cual*

$$\frac{|p - p_0|}{|p|} < 5 \times 10^{-t}.$$

En otras palabras, una aproximación tiene t cifras significativas si el error relativo es menor 5×10^{-t}. Por ejemplo, una aproximación tiene dos cifras significativas si el error relativo es menor a 0,05.

3. Propagación de Errores

Además de la representación imprecisa de los números, la aritmética realizada en una computadora no es exacta. Ver ejemplos en clase.