IEEE floating-point in C#

Peter Sestoft, IT University of Copenhagen
sestoft@itu.dk • 2009-01-05

1 Results

C# distinguishes -0.0 from +0.0 internally but prints both as 0 by default. Moreover, there does not seem to be any standard way to force printing of the sign of -0.0, so to display the results below we detected this special case explicitly using `(r==0 && 1/r<0)`.

Results are as required by the IEEE 754-2008 standard, except for `Math.Pow(NaN, +/-0.0)` and `Math.Atan2(+/-0.0,+/-0.0)`. Source code is in cs/Numbers.cs.

1.1 Arithmetic operators

These all agree with the corresponding operators in Java.

+	-Inf	-2	-0	0	2	+Inf	NaN
-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	NaN	NaN
-2	-Inf	-4	-2	-2	0	+Inf	NaN
-0	-Inf	-2	-0	0	2	+Inf	NaN
0	-Inf	-2	0	0	2	+Inf	NaN
2	-Inf	0	2	2	4	+Inf	NaN
+Inf	NaN	+Inf	+Inf	+Inf	+Inf	+Inf	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

-	-Inf	-2	-0	0	2	+Inf	NaN
-Inf	NaN	-Inf	-Inf	-Inf	-Inf	-Inf	NaN
-2	+Inf	0	-2	-2	-4	-Inf	NaN
-0	+Inf	2	0	-0	-2	-Inf	NaN
0	+Inf	2	0	0	-2	-Inf	NaN
2	+Inf	4	2	2	0	-Inf	NaN
+Inf	+Inf	+Inf	+Inf	+Inf	+Inf	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

*	-Inf	-2	-0	0	2	+Inf	NaN
-Inf	+Inf	+Inf	NaN	NaN	-Inf	-Inf	NaN
-2	+Inf	4	0	-0	-4	-Inf	NaN
-0	NaN	0	0	-0	-0	NaN	NaN
0	NaN	-0	-0	0	0	NaN	NaN
2	-Inf	-4	-0	0	4	+Inf	NaN
+Inf	-Inf	-Inf	NaN	NaN	+Inf	+Inf	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

/	-Inf	-2	-0	0	2	+Inf	NaN
-Inf	NaN	+Inf	+Inf	-Inf	-Inf	NaN	NaN
-2	0	1	+Inf	-Inf	-1	-0	NaN
-0	0	0	NaN	NaN	-0	-0	NaN
0	-0	-0	NaN	NaN	0	0	NaN
2	-0	-1	-Inf	+Inf	1	0	NaN
+Inf	NaN	-Inf	-Inf	+Inf	+Inf	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

%	-Inf	-2	-0	0	2	+Inf	NaN
-Inf	NaN	NaN	NaN	NaN	NaN	NaN	NaN
-2	-2	-0	NaN	NaN	-0	-2	NaN
-0	-0	-0	NaN	NaN	-0	-0	NaN
0	0	0	NaN	NaN	0	0	NaN
2	2	0	NaN	NaN	0	2	NaN
+Inf	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

1.2 Comparison operators

==	-Inf	-2	-0	0	2	+Inf	NaN
-Inf	True	False	False	False	False	False	False
-2	False	True	False	False	False	False	False
-0	False	False	True	True	False	False	False
0	False	False	True	True	False	False	False
2	False	False	False	False	True	False	False
+Inf	False	False	False	False	False	True	False
NaN	False	False	False	False	False	False	False

!=	-Inf	-2	-0	0	2	+Inf	NaN
-Inf	False	True	True	True	True	True	True
-2	True	False	True	True	True	True	True
-0	True	True	False	False	True	True	True
0	True	True	False	False	True	True	True
2	True	True	True	True	False	True	True
+Inf	True	True	True	True	True	False	True
NaN	True	True	True	True	True	True	True

<	-Inf	-2	-0	0	2	+Inf	NaN
-Inf	False	True	True	True	True	True	False
-2	False	False	True	True	True	True	False
-0	False	False	False	False	True	True	False
0	False	False	False	False	True	True	False
2	False	False	False	False	False	True	False
+Inf	False	False	False	False	False	False	False
NaN	False	False	False	False	False	False	False

<=	-Inf	-2	-0	0	2	+Inf	NaN
-Inf	True	True	True	True	True	True	False
-2	False	True	True	True	True	True	False
-0	False	False	True	True	True	True	False
0	False	False	True	True	True	True	False
2	False	False	False	False	True	True	False
+Inf	False	False	False	False	False	True	False
NaN	False	False	False	False	False	False	False

>	-Inf	-2	-0	0	2	+Inf	NaN
-Inf	False	False	False	False	False	False	False
-2	True	False	False	False	False	False	False
-0	True	True	False	False	False	False	False
0	True	True	False	False	False	False	False
2	True	True	True	True	False	False	False
+Inf	True	True	True	True	True	False	False
NaN	False	False	False	False	False	False	False

>=	-Inf	-2	-0	0	2	+Inf	NaN
-Inf	True	False	False	False	False	False	False
-2	True	True	False	False	False	False	False
-0	True	True	True	True	False	False	False
0	True	True	True	True	False	False	False
2	True	True	True	True	True	False	False
+Inf	True	True	True	True	True	True	False
NaN	False	False	False	False	False	False	False

1.3 Two-argument mathematical functions

The `Atan2` function disagrees with IEEE 754-2008 and with Java on `Atan2(+/-0.0,+/-0.0)`.

Math.Atan2	-Inf	-2	-0	0	2	+Inf	NaN
-Inf	NaN	-1.571	-1.571	-1.571	-1.571	NaN	NaN
-2	-3.142	-2.356	-1.571	-1.571	-0.785	-0.000	NaN
-0	-3.142	-3.142	-3.142	-0.000	-0.000	-0.000	NaN
0	3.142	3.142	3.142	0.000	0.000	0.000	NaN
2	3.142	2.356	1.571	1.571	0.785	0.000	NaN
+Inf	NaN	1.571	1.571	1.571	1.571	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

The `IEEERemainder` function agrees with floating-point remainder `x%y` on the arguments shown here, but not in general; for instance, `IEEERemainder(7,2)` is -1 whereas `7%2` is $+1$.

Math.IEEERemainder	-Inf	-2	-0	0	2	+Inf	NaN
-Inf	NaN	NaN	NaN	NaN	NaN	NaN	NaN
-2	-2	-0	NaN	NaN	-0	-2	NaN
-0	-0	-0	NaN	NaN	-0	-0	NaN
0	0	0	NaN	NaN	0	0	NaN
2	2	0	NaN	NaN	0	2	NaN
+Inf	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Math.Max	-Inf	-2	-0	0	2	+Inf	NaN
-Inf	-Inf	-2	-0	0	2	+Inf	NaN
-2	-2	-2	-0	0	2	+Inf	NaN
-0	-0	-0	-0	0	2	+Inf	NaN
0	0	0	-0	0	2	+Inf	NaN
2	2	2	2	2	2	+Inf	NaN
+Inf	+Inf	+Inf	+Inf	+Inf	+Inf	+Inf	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Math.Min	-Inf	-2	-0	0	2	+Inf	NaN
-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf	NaN
-2	-Inf	-2	-2	-2	-2	-2	NaN
-0	-Inf	-2	-0	0	-0	-0	NaN
0	-Inf	-2	-0	0	0	0	NaN
2	-Inf	-2	-0	0	2	2	NaN
+Inf	-Inf	-2	-0	0	2	+Inf	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

The `Pow` function disagrees with IEEE 754-2008 and with Java on `Pow(NaN,+/-0.0)`.

Math.Pow	-Inf	-2	-0	0	2	+Inf	NaN
-Inf	0	0	1	1	+Inf	+Inf	NaN
-2	0	0.25	1	1	4	+Inf	NaN
-0	+Inf	+Inf	1	1	0	0	NaN
0	+Inf	+Inf	1	1	0	0	NaN
2	0	0.25	1	1	4	+Inf	NaN
+Inf	0	0	1	1	+Inf	+Inf	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

1.4 One-argument mathematical functions

These all agree with IEEE 754-2008 and with the corresponding functions in Java.

	-Inf	-2.000	-0.000	0.000	2.000	+Inf	NaN
Math.Abs	+Inf	2.000	0.000	0.000	2.000	+Inf	NaN
Math.Acos	NaN	NaN	1.571	1.571	NaN	NaN	NaN
Math.Asin	NaN	NaN	-0.000	0.000	NaN	NaN	NaN
Math.Atan	-1.571	-1.107	-0.000	0.000	1.107	1.571	NaN
Math.Ceiling	-Inf	-2.000	-0.000	0.000	2.000	+Inf	NaN
Math.Cos	NaN	-0.416	1.000	1.000	-0.416	NaN	NaN
Math.Exp	0.000	0.135	1.000	1.000	7.389	+Inf	NaN
Math.Floor	-Inf	-2.000	-0.000	0.000	2.000	+Inf	NaN
Math.Log	NaN	NaN	-Inf	-Inf	0.693	+Inf	NaN
Math.Log10	NaN	NaN	-Inf	-Inf	0.301	+Inf	NaN
Math.Round	-Inf	-2.000	-0.000	0.000	2.000	+Inf	NaN
Math.Sin	NaN	-0.909	-0.000	0.000	0.909	NaN	NaN
Math.Sqrt	NaN	NaN	-0.000	0.000	1.414	+Inf	NaN
Math.Tan	NaN	2.185	-0.000	0.000	-2.185	NaN	NaN