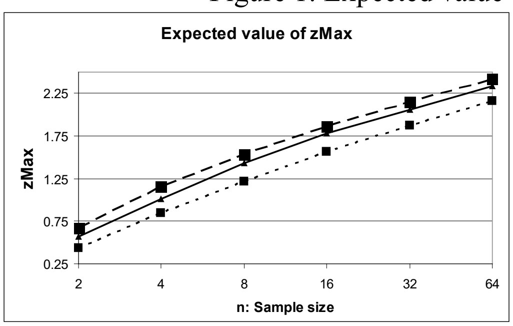
Milo Schield

Background: The mathematics of extreme value theory is daunting. See Wikipedia. This is true even when sampling from a standard normal distribution. Obtaining the results empirically using simulation is time consuming. But there is a short cut way of estimating the extreme values of any sample from any distribution.

SHORT-CUT: Assume the n values will arrange themselves with equidistant separation in probability space. There are two ways to do this. The first way assumes the end points (highest and lowest values) have the same distance to the ends in probability space as to their nearest neighbor. For a two-item sample, the minimum and maximum are at the 33 1/3 percentile and the 66 2/3rds percentile. A second way assumes that each data point has the same space on both sides and these spaces do not overlap. For a two-item sample, the minimum and maximum are at the 25^{th} and 75^{th} percentile. Each point has 25% of the probability space on either side and the spaces do not overlap.

Convert these percentiles to z-scores – assuming a normal distribution. In the first case, the 67^{th} percentile gives zMax = +0.431. In the second case, the 75^{th} percentile gives zMax = 0.674.
Figure 1 shows the expected value of zMax as a function of sample size along with the results for these two simple models.

Figure 1: Expected value of zMax as a function of sample size



Sample	Expected Value of zMax		
Size	= -NormSInv	Expected	= -NormSInv
n	(1/(n+1))	Theory	(1/(2n))
2	0.431	0.57	0.674
4	0.842	1.01	1.150
8	1.221	1.43	1.534
16	1.565	1.78	1.863
32	1.876	2.06	2.154
64	2.160	2.34	2.418

Source of theory: www.panix.com/~kts/Thesis/extreme/extreme2.html

In the case of the normal distribution, it turns out that these two simple models enclose the theoretical solution between them for n < 65. Going from area under the normal curve to z is not easy. A good fit for zMax as a function of n is given by [1 + LN(n)] / 3 for n < 65

SIMPLE MODEL FOR Zmax: If this relationship between these simple models and the exact theoretical models holds for n > 64, then a simple model is given by zMax = 1.8 plus half the number of zeroes in the sample size (when expressed as a power of 10). [*** Supply a proof of this ***] Range: 100 to 100 million. [***Why this upper limit?***] So for n = 10,000, zMax is expected to be 3.8: 1.8 + 2.

APPLICATION: Consider the tallest humans. Assume an average adult male height of 70 inches and a standard deviation of three inches. In a world of ten billion adults, zMax should be 6.8: 1.8 plus five. This gives a height of 91 inches (70 + 3*7) or seven feet seven inches. According to "the list of tallest people" in Wikipedia, heights exceeding eight feet have been confirmed by Guinness World Records. These heights in excess of those predicted indicate a non-random process such as a genetic defect.