An Introduction to Extreme Value Theory

by Robert A. Bear, CPCU, FCAS, MAAA

Robert A. Bear, CPCU, FCAS, MAAA, an ARIAS Certified Arbitrator, is a consultant whose firm, RAB Actuarial Solutions LLC, offers actuarial, reinsurance and statistical modeling services. He currently serves as chairperson of the Casualty Actuarial Society (CAS) Dynamic Risk Modeling Committee. As chairperson of the CAS Loss Simulation Model Working Party (LSMWP), he led the development of the Loss Simulation Model that can be used to test loss reserving methods and models. He may be contacted at rabsolutions@gmail.com, and his blog on actuarial and statistical modeling issues is available at rabactuarial.blogspot.com.

Introduction

A branch of statistics known as extreme value theory has attracted much interest after the natural and financial catastrophes of recent years. In this introduction to this subject, I hope to show that even with limited data that does not permit one to confidently fit a loss distribution, one can calculate reasonable ballpark estimates of the likelihood that more extreme events will occur in the future. I will use an example that should be of interest to international underwriters to illustrate some basic concepts and results.

Example

Table 1 summarizes the data for my example: The data represents estimated annual damages, adjusted for inflation and normalized for growth in personal wealth and population, arising from tropical cyclones making landfall in the United States from 1900 through 2006 (in millions of U.S. dollars). The formulas that I will apply assume that the observations are realizations from independent, identically distributed random variables. If this assumption is true, theory would lead one to expect 5.25 records between 1901 and 2006. (Note that 1901 is considered to be a record, as is any year since that exceeded all previous observations.)

Table 1
U.S. Tropical Cyclone Data from 1900 through 2006 (Millions of U.S. Dollars)

Estimated Annual Damages, Adjusted for Inflation and Normalized for Growth in Personal Wealth and Population.

Year	Normalized Loss	Year	Normalized Loss	Year	Normalized Loss	Year	Normalized Loss
1900	104,330	1927	-	1954	37,455	1981	180
1901	213	1928	35,298	1955	24,438	1982	45
1902	-	1929	390	1956	606	1983	7,843
1903	6,803	1930	-	1957	4,034	1984	304
1904	1,139	1931	-	1958	535	1985	11,622
1905		1932	6,210	1959	902	1986	53
1906	4,080	1933	14,006	1960	31,469	1987	20
1907	-	1934	932	1961	15,192	1988	182
1908	-	1935	9,150	1962	97	1989	17,609
1909	3,081	1936	838	1963	259	1990	133
1910	876	1937	-	1964	16,478	1991	3,196
1911	235	1938	41,140	1965	22,324	1992	60,547
1912	-	1939		1966	353	1993	133
1913	724	1940	1,224	1967	4,217	1994	2,036
1914	-	1941	2,530	1968	690	1995	7,877
1915	74,262	1942	2,475	1969	22,286	1996	6,864
1916	7,919	1943	3,746	1970	5,909	1997	172
1917	-	1944	54,760	1971	2,188	1998	6,323
1918	886	1945	14,676	1972	18,458	1999	8,692
1919	14,392	1946	4,953	1973	153	2000	38
1920	367	1947	20,071	1974	1,127	2001	7,319
1921	3,348	1948	4,249	1975	2,931	2002	1,566
1922	-	1949	16,147	1976	511	2003	4,423
1923	-	1950	5,806	1977	56	2004	51,587
1924	-	1951	376	1978	153	2005	121,296
1925	-	1952	120	1979	14,801	2006	500
1926	169,398	1953	59	1980	1,682		

Source of Data: The government publication known as NOAA Technical Memorandum NWS TPC-5 and titled *The Deadliest, Costliest, and Most Intense United States Tropical Cyclones From 1851 to 2006 (and Other Frequently Requested Hurricane Facts)*.

This National Weather Service publication was updated on April 15, 2007, for return period information by Eric S. Blake, Edward N. Rappaport and Christopher W. Landsea at the National Hurricane Center in Miami, Fla.

We actually observed four records during this period, which is close enough that our key assumption is not contradicted by the data. (If I had included 1900 in the test, which was an extreme year for hurricanes, we would have observed only two records through 2006 compared to 5.25 expected records. Such miscreant data points are referred to as "outliers" by statisticians and are not treated favorably.)

Gumbel's Method of Exceedances

Let us assume that we have sorted our data from largest to smallest for a period spanning n years. The largest value is called the first order statistic; the next largest value is the second order statistic, and so on. Gumbel's Method of Exceedances states that the expected number (mean or average) of exceedances of the kth order statistic among the next r observations is rk/(n+1). For our 106-year data set, the expected number of exceedances of the fifth largest value ($54,760 m in 1944) during the next 10 years is (10)(5)/107=0.47. The probability of a record loss next year is (1)(1)/107=.0093.

Table 2 provides a retrospective test of this method, where predictions for 1982–2006 are based on the history for 1901 through 1981.

Table 3 provides projections based upon this method for the 15 years subsequent to 2006 (2007–2021), based upon the history from 1901 through 2006.

Table 4 summarizes the next year exceedance probabilities for the first 12-order statistics, along with estimated return periods (reciprocals of the exceedance probabilities). The probabilities of exceeding each loss level of interest before the return period is calculated by subtracting from one the probability that the loss level is below the level of interest in all years before the return period. It can be shown that the probability of exceeding a low probability loss level before the return period is approximately 63.2 percent.

Table 4 shows that this approximation works reasonably well for return periods above 20. Thus, the time to the estimated return period encompasses approximately 63.2 percent of the time to exceedance probability distribution rather than 50 percent as many may assume. (The mean is well above the median, so one can't assume that one has a 50 percent chance of dodging the bullet before the estimated return period.)

Table 2

Order Statistic	Annual Loss	Year	Predicted Exceedances	Observed Exceedances
1	169,398	1926	0.30	0
2	74,262	1915	0.61	1
5	37,455	1954	1.52	3
10	22,286	1969	3.05	3
25	5,806	1950	7.62	11

Table 3

Order Statistic	Annual Loss	Year	Predicted Exceedances
1	169,398	1926	0.14
2	121,296	2005	0.28
5	54,760	1944	0.70
10	31,469	1960	1.40
25	9,150	1935	3.50

Table 4

Order Statistic	Normalized Hurricane Damage	Year	Probability of Exceedance Next Year (p)	Estimated Return Period (RP)	Probability of Exceedance Before RP
1	169,398	1926	0.93%	107.00	63.38%
2	121,296	2005	1.87%	53.50	63.21%
3	74,262	1915	2.80%	35.67	63.04%
4	60,547	1992	3.74%	26.75	62.86%
5	54,760	1944	4.67%	21.40	63.39%
6	51,587	2004	5.61%	17.83	62.51%
7	41,140	1938	6.54%	15.29	63.76%
8	37,455	1954	7.48%	13.38	63.59%
9	35,298	1928	8.41%	11.89	61.96%
10	31,469	1960	9.35%	10.70	62.51%
11	24,438	1955	10.28%	9.73	62.33%
12	22,324	1965	11.21%	8.92	61.39%

Conclusions

Extreme value theory is potentially a very useful tool to aid those insuring or managing risks from low probability loss events. As I have just scratched the surface of the subject, additional information may be found in Wikipedia, at www.en.wikipedia.org/wiki/Extreme_value_theory.

Anytime something bad happens, you invariably hear that "it could always be worse." Now, with the aid of extreme value theory, you can go forth and calculate the odds that it will be worse next time or that it will be just less awful! ■