TITLE: ESTIMATING PROBABLE MAXIMUM LOSS WITH ORDER STATISTICS

AUTHOR: Margaret E. Wilkinson

Miss Wilkinson is an Assistant Vice President of Tillinghast, Nelson & Warren, Inc. She has both a Bachelor of Arts and a Master of Arts in Mathematics from Vanderbilt University. She also has a Master of Science in Statistics from Stanford University.

Miss Wilkinson is an Associate of the Casualty Actuarial Society (1981) and a Member of the American Academy of Actuaries (1982).

Miss Wilkinson has been with TN&W since she graduated from Stanford in 1978. She has experience in reserving, ratemaking and Annual Statement preparation. She has been involved with projects for clients of all sizes, from the small self-insured hospital to the large reinsurance pool. Her experience includes all the major casualty and property lines, as well as reinsurance.

REVIEWER: Professor Albert J. Beer

Mr. Beer is Associate of Actuarial Science at The College of Insurance in New York City. In addition to presenting a number of seminars in the quantitative aspects of Risk Management throughout the country, Al also acts as an independent Risk Management and actuarial consultant. He received his B.S. from Manhattan College, his M.S. from the University of Colorado and became an F.C.A.S. in 1980.

INTRODUCTION

In the past there has been much discussion about the definition of probable maximum loss (PML), but little attention has been given to its quantification. This paper will introduce the concept of order statistics as a tool to use in estimating the PML. Two approaches will be used that will lead to six specific methods for estimating the PML. These six methods will then be illustrated with specific examples.

The term PML is usually used in connection with property insurance but it can also be applied to liability insurance. In fact, there is some controversy over whether the appropriate term, from a risk management viewpoint, is probable maximum loss, maximum possible loss, estimated maximum loss or one of many other similar phrases.

McGuinness offers two definitions:[1]

> "The probable maximum loss <u>for a property</u> is that proportion of

[1] McGuinness, John S., "Is 'Probable Maximum Loss' (PML) A Useful Concept?", <u>Proceedings of the Casualty Actuarial Society</u> Vol. LVI, 1969, p. 31.

> total value of the property which will equal or exceed, in a stated proportion of all cases, the amount of loss from a specified peril or group of perils.
>
> The probable maximum loss under a given insurance contract is that proportion of the limit of liability which will equal or exceed, in a stated proportion of all cases, the amount of any loss covered by a contract."

The first definition is pertinent to insureds and risk managers, while the second is pertinent to underwriters. These definitions were later combined by McGuinness into one generalized definition:[2]

> "The PML for a specified financial interest is that proportion of the total value of the interest which will equal or exceed, in a stated proportion of all cases, the amount of any financial loss to the interest from a specified event or group of events."

A guest reviewer of McGuinness' paper, who is an underwriter, offered the following observations:[3]

> "It is true that the definitions may vary between underwriters when put down

[2] McGuinness, John S., "Author's Review of Discussions in Volume LVI, Pages 40-48", Proceedings of the Casualty Actuarial Society Vol. LVII, 1970, p. 107.

[3] Black, Edward B., "Discussion by Edward B. Black", Proceedings of the Casualty Actuarial Society Vol. LVI, 1969, p. 46.

> in words, but I feel strongly that there is a universal understanding as to the end result which all underwriters expect PML to accomplish. ... PML, no matter how you define it, is simply Probable Maximum Loss. It is neither foreseeable nor possible loss - rather, it is the maximum loss which probably will happen when, and if, the peril insured against actually occurs."

The definition which will be used in this paper is, simply, that probable maximum loss is the worst loss likely to happen. Since an actual loss could be greater than the PML, the PML depends upon (i) estimates of the likelihood that losses of various sizes will occur, (ii) the insured's risk acceptance level and (iii) the underwriter's risk aversion level. Note that for the same risk the insured and underwriter can have different estimates of the PML.

ORDER STATISTICS

Let $X_1, X_2, \ldots, X_n$ denote a random sample from a population with continuous cumulative distribution function F_X. Since F_X is continuous, the probability of any two sample values being equal is zero. Consequently, there exists a unique ordered arrangement of the sample. Let $X_{(1)}$ denote the smallest member of the set, $X_{(2)}$ the second smallest, etc. Then

$$X_{(1)} < X_{(2)} < \ldots < X_{(n)}$$

and these are called the _order statistics_ from the random sample $X_1, X_2, \ldots, X_n$. For $1 \leq r \leq n$, $X_{(r)}$ is called the _r^{th} order statistic_.

Order statistics are particularly useful for studying certain phenomenon because quite a few of the results concerning the properties of $X_{(r)}$ and the properties of functions of some subset of the order statistics are distribution-free. If an inference is distribution-free, assumptions regarding the underlying population are not necessary since the inference is based on a random variable with a distribution which is independent of the underlying population's distribution.

GENERAL RESULTS CONCERNING $X_{(n)}$

$X_{(n)}$ is the largest value of the sample. This is a good place to start since probable maximum loss is the worst loss likely to happen.

Distribution of $X_{(n)}$

The cumulative distribution function of X_n is given by

$$\begin{aligned} F_{X_{(n)}}(x) &= \Pr\{\, X_{(n)} \le x \,\} \\ &= \Pr\{\, \text{all } X_i \le x \,\} \\ &= F_X{}^n(x) \end{aligned} \qquad (1)$$

since the X_i's are independent. The corresponding density function is found by differentiating (1). It is easily verified that

$$f_{X_{(n)}}(x) = nf_X(x)F_X{}^{n-1}(x) \qquad (2)$$

where f_X is the density function corresponding to F_X.

Moments of $X_{(n)}$

The exact moments of $X_{(n)}$ can be derived from the following equation:

$$E(X_{(n)}^k) = \int_{-\infty}^{\infty} x^k f_{X_{(n)}}(x)\,dx$$

$$= \int_{-\infty}^{\infty} n x^k f_X(x) F_X^{n-1}(x)\,dx. \qquad (3)$$

This requires a specified distribution F_X and is of limited practical value due to the complexity of the integral involved.

There are large-sample approximations for the mean and variance of $X_{(n)}$ that are easily calculable. The approximations require two facts.

1. If $U_{(r)}$ denotes the r^{th} order statistic from a uniform distribution over the interval (0,1), then

$$X_{(r)} = F_X^{-1}(U_{(r)}).$$

2. The Taylor's series expansion of a function g(z) about a point μ is

$$g(z) = g(\mu) + \sum_{i=1}^{\infty} \frac{(z-\mu)^i}{i!} g^{(i)}(\mu)$$

where $g^{(i)}(\mu) = \left.\frac{d^i g(z)}{dz^i}\right|_{z=\mu}$. This series converges if

$$\lim_{n\to\infty} \frac{(z-\mu)^n}{n!} g^{(n)}(z_1) = 0$$

for $\mu < z_1 < z$.

The first requirement is due to the probability integral transformation and is proved in various statistical texts.[4] The second requirement is the standard Taylor's series expansion.

If the Taylor's series expansion is rewritten for a random variable Z with mean μ, and the expected value of both sides is taken, the result is

$$E[g(Z)] = g(\mu) + \frac{\mathrm{var}(Z)}{2!} g^{(2)}(\mu) + \sum_{i=3}^{\infty} \frac{E[(Z-\mu)^i]}{i!} g^{(i)}(\mu).$$

So, a first approximation to $E[g(Z)]$ is $g(\mu)$, and a second approximation is $g(\mu) + \frac{\mathrm{var}(Z)}{2!} g^{(2)}(\mu)$.

To find similar approximations for var $[g(Z)]$, form the difference $g(Z)-E[g(Z)]$, square it and take the expected value. The result is

$$\mathrm{var}[g(Z)] = \mathrm{var}(Z)[g^{(1)}(\mu)]^2 - \frac{1}{4}[g^{(2)}(\mu)]^2 \mathrm{var}^2(Z) + E[h(Z)]$$

where $E[h(Z)]$ involves third or higher central moments of Z.[5] A first approximation to $\mathrm{var}[g(Z)]$ is

[4] In particular, see Gibbons, Jean Dickinson, Nonparametric Statistical Inference, p. 23, New York 1971.

[5] Ibid, p. 35.

$\mathrm{var}(Z)[g^{(1)}(\mu)]^2$, and a second approximation is $\mathrm{var}(Z)[g^{(1)}(\mu)]^2 - \frac{1}{4}[g^{(2)}(\mu)]^2\mathrm{var}^2(Z)$.

In order to apply these results to $X_{(n)}$, g is defined so that

$$g(u_{(n)}) = x_{(n)} = F_X^{-1}(u_{(n)})$$

where $u_{(n)} = F_X(x_{(n)})$. The appropriate moments are

$$\mu = E[u_{(n)}] = \frac{n}{n+1}$$

and

$$\mathrm{var}[u_{(n)}] = \frac{n}{(n+1)^2(n+2)} .$$ [6]

The derivatives needed are

$$g^{(1)}(\mu) = \{f_X[F_X^{-1}(\tfrac{n}{n+1})]\}^{-1}$$

and

$$g^{(2)}(\mu) = -f_X'[F_X^{-1}(\tfrac{n}{n+1})]\{f_X[F_X^{-1}(\tfrac{n}{n+1})]\}^{-3}.$$ [7]

Substituting yields as first approximations:

$$E(X_{(n)}) \approx F_X^{-1}(\tfrac{n}{n+1}) \tag{4}$$

$$\mathrm{var}(X_{(n)}) \approx \frac{n}{(n+1)^2(n+2)}\{f_X[F_X^{-1}(\tfrac{n}{n+1})]\}^{-2}. \tag{5}$$

[6] Ibid, pp. 32-33.
[7] Ibid, p. 37.

Second approximations are similarly found by the appropriate substitutions.

Distribution-Free Bounds for $E(X_{(n)})$*

If a variate X has a finite variance, the expected value of $X_{(n)}$ can not be arbitrarily large even if the range of X is unbounded.

From Equation (3), the expected value of $X_{(n)}$ is

$$E(X_{(n)}) = \int_{-\infty}^{\infty} nxF_X^{\,n-1}(x)f_X(x)dx.$$

Let $u = F_X(x)$ and standardize X to have mean 0 and variance 1. This means

$$E(X_{(n)}) = \int_0^1 nx(u)u^{n-1}du,$$

$$\int_0^1 x(u)du = 0,$$

$$\int_0^1 [x(u)]^2 du = 1,$$

where $x(u)$ indicates that x is expressed as a function of u.

Schwartz's inequality states that

$$\int fg\ du \leq (\int f^2 du \int g^2 du)^{1/2}.$$

* David, Herbert A, Order Statistics pp. 56-59, New York 1981.

Let $f = x$ and $g = nu^{n-1}-1$. Then

$$\int_0^1 x(nu^{n-1}-1)\,du \le \left(\int_0^1 x^2 du \int_0^1 (nu^{n-1}-1)^2 du\right)^{1/2}.$$

Expanding yields

$$\int_0^1 xnu^{n-1}du - \int_0^1 x\,du$$

$$\le \left(\int_0^1 x^2 du\right)^{1/2} \left(\int_0^1 (n^2u^{2n-2}-2nu^{n-1}+1)\,du\right)^{1/2}.$$

Substituting for the various pieces gives

$$E(X_{(n)}) \le \left(\int_0^1 (n^2u^{2n-2}-2nu^{n-1}+1)\,du\right)^{1/2}.$$

Hence

$$E(X_{(n)}) \le \frac{n-1}{(2n-1)^{1/2}} \quad .$$

If the mean and variance of the population are μ and σ^2, respectively, the result becomes

$$E(X_{(n)}) \le \mu + \frac{(n-1)\sigma}{(2n-1)^{1/2}} \quad . \tag{6}$$

This result is distribution-free and requires only the knowledge of the mean and variance of the population, not its specific distribution.

GENERAL RESULTS FOR QUANTILES

Probable maximum loss has been defined as the worst loss likely to happen. If the sample under consideration has an unreasonably large loss, then using $X_{(n)}$ to estimate the PML would be unreasonable. In this case, quantiles could be used. The quantile approach would also be preferred if the insured was willing to accept more risk or the underwriter wanted to accept less risk. "More risk" and "less risk" used in this context are comparable to the risk level implied by using $X_{(n)}$ to estimate the PML.

A quantile of a continuous distribution $f_X(x)$ of a random variable X is a real number which divides the area under the probability density function into two parts of specified amounts. Denote the p^{th} quantile by κ_p for $0 \leq p \leq 1$. Then κ_p is defined as any real number solution to the equation

$$F_X(\kappa_p) = p.$$

It is assumed that there is a unique solution to this equation, as there would be if F_X is strictly increasing.

Point Estimate for κ_p[9]

It can be shown that the r^{th} order statistic is a consistent estimator of the p^{th} quantile where $\frac{r}{n} = p$ remains fixed. A definition which provides a unique $X_{(r)}$ to estimate the p^{th} quantile is to choose r so that

$$r = \begin{cases} np & \text{if np is an integer} \\ [np+1] & \text{if np is not an integer} \end{cases} \tag{7}$$

where [x] denotes the greatest integer not exceeding x.

Distribution-Free Confidence Interval for κ_p[10]

Since consistency is only a large-sample property, it is desirable to have an interval estimate for κ_p with a known confidence coefficient for a given sample size. The objective is to find two numbers r and s, $r < s$, such that

$$P(X_{(r)} < \kappa_p < X_{(s)}) = 1-\alpha$$

for some chosen number $0 < \alpha < 1$.

For all $r < s$,

$$P(X_{(r)} < \kappa_p < X_{(s)}) = P(X_{(r)} < \kappa_p) - P(X_{(s)} < \kappa_p).$$

[9] Op. Cit. Gibbons, pp. 40-41.
[10] Ibid, pp. 41-43.

Since F_X is a strictly increasing function,

$$X(r) < \kappa_p \text{ if and only if } F_X(X_{(r)}) < F_X(\kappa_p) = p.$$

Thus,

$$P(X_{(r)} < \kappa_p < X_{(s)}) = P[F_X(X_{(r)}) < p] - P[F_X(X_{(s)}) < p]$$

$$= \int_0^p n\binom{n-1}{r-1} x^{r-1}(1-x)^{n-r} dx$$

$$- \int_0^p n\binom{n-1}{s-1} x^{s-1}(1-x)^{n-s} dx.$$

If this formula is integrated by parts the necessary number of times, the result is

$$P(X_{(r)} < \kappa_p < X_{(s)}) = \sum_{i=r}^{s-1} \binom{n}{i} p^i (1-p)^{n-i}. \qquad (8)$$

This does not produce a unique solution for r and s. The narrowest interval is produced when $X_{(s)} - X_{(r)}$ is minimized. Alternatively, s-r could be minimized. Also, a confidence interval produced by

$$\sum_{i=r}^{s-1} \binom{n}{i} p^i (1-p)^{n-i} = 1-\alpha$$

is distribution-free.

The formula derived above can also be argued directly. For any p, $X_{(r)} < \kappa_p$ if and only if at least r of the sample values $X_1, X_2, \ldots, X_n$ are less than κ_p. The sample values are independent and can be classified according to whether they are less than κ_p. Thus, the n random variables can be considered the result of n independent trials of a Bernoulli variable with parameter p. The number of observations less than κ_p then has a binomial distribution with parameter p.

APPLICATION OF ORDER STATISTICS TO THE PML PROBLEM

The application of order statistics has various requirements depending on the approach taken. The PML can simply be estimated by $X_{(n)}$ if a reliable data set applicable to the particular problem is available. If the concern is to estimate the PML by using the expected value of $X_{(n)}$ or by constructing an interval around $X_{(n)}$ using the variance of $X_{(n)}$ and choosing the PML as the upper limit of this interval, the distribution of X, F_X, must be known (actually F_X^{-1}, f_X and f_X' are needed). If estimates of the mean and variance of F_X are available, derived either theoretically or from a data set, then the upper bound for $E(X_{(n)})$ could be used as the PML. If a data set is available but, for various reasons, the quantile approach is preferred, only the order statistics themselves are necessary to produce either a point estimate for the quantile or a confidence interval for the quantile. In the former case, the PML would be the quantile; in the latter case the PML would be the upper bound of the confidence interval.

$X_{(n)}$ as an Estimate for PML

Exhibit I contains a list of 100 claims that are representative of a particular problem in which a PML estimate is needed. $X_{(n)}$ in this case is $X_{(100)}$ or \$576,525. Consequently the PML is \$576,525.

$E(X_{(n)})$ as an Estimate for the PML

The use of $E(X_{(n)})$ as an estimate for the PML requires F_X^{-1}. Suppose it is assumed that the data has a lognormal distribution. The mean is \$212,521 and the standard deviation is \$110,506. The corresponding normal distribution has a mean of 12.14714 and a standard deviation of .48920. From Equation (4), the approximation for the expected value of $X_{(n)}$ is

$$E(X_{(n)}) \approx \Lambda_X^{-1}\left(\frac{n}{n+1}\right) = e^{[\sigma Z^{-1}(\frac{n}{n+1}) + \mu]}$$

where Λ_X is the lognormal distribution,

Z is the standard normal distribution,

μ is the mean of the normal distribution, and

σ is the standard deviation of the normal distribution.

If $n = 100$, the value of $Z^{-1}(.9901)$ is found from standard normal tables to be 2.33. The PML estimate is \$589,468.

The Upper Bound of an Interval Around $E(X_{(n)})$ Using Var $(X_{(n)})$ as an Estimate for the PML

It is possible to choose k so that

$$E(X_{(n)}) + k(\mathrm{var}(X_{(n)}))^{1/2}$$

produces a reasonable estimate of the risk that is acceptable. If the prior example is continued, the $\mathrm{var}(X_{(n)})$ can be approximated using Equation (5):

$$\mathrm{var}(X_{(n)}) \approx \frac{100}{(101)^2(102)} (\lambda_X(589,468))^{-2}$$

where λ_X is the density function corresponding to Λ_X. The formula for λ_X is

$$\lambda_X(x) = \frac{1}{x\sigma(2\pi)^{1/2}} e^{\{-\frac{1}{2\sigma^2}(\ln x-\mu)^2\}} .$$

The $(\mathrm{var}(X_{(n)}))^{1/2}$ is \$106,976 for this example. If k is chosen to be 2.0, the PML estimate is \$803,420.

The Distribution-Free Upper Bound of $E(X_{(n)})$ as an Estimate for the PML

The data shown in Exhibit I have a sample mean of \$212,521 and a sample standard deviation of \$110,506.

Consequently,

$$E(X_{(100)}) \leq 212{,}521 + \frac{99\ (110{,}506)}{(199)^{1/2}} .$$

The PML is thus \$988,044.

If sample data are not available, a mean, variance and number of claims could be chosen on some theoretical grounds and the upper bound calculated as shown above.

κ_p as an Estimate for the PML

Suppose it is decided that the .95 quantile will be used as the PML. If the sample data from Exhibit I are used, r is 95 (because .95x100 = 95) and the PML ($X_{(95)}$) is \$434,449.

The Distribution-Free Upper Bound of κ_p as an Estimate for the PML

The estimate of κ_p for p = .95 based on the sample data is \$434,449. Now a confidence interval is desired around this estimate so that α = .10. In other words, $r < s$ must be found so that

$$P(X_{(r)} < \kappa_p < X_{(s)}) = \sum_{i=r}^{s-1} \binom{n}{i} p^i (1-p)^{n-i} = .90.$$

s-r should also be minimized. Exhibit II shows $X_{(i)}$ and $\binom{n}{i}\ p^i (1-p)^{n-i}$ for i = 90, 91, ..., 100. There are two possibilities for r and s:

$$P(X_{(91)} < \kappa_{.95} < X_{(99)}) = .934732$$

and

$$P(X_{(92)} < \kappa_{.95} < X_{(99)}) = .899831.$$

The second is closer to .90 and s-r is 7. The first has an s-r of 8. Even though the probabilities are so close, and the second probability is slightly less than .90, the second answer would be chosen because s-r is minimized. The PML in this case is $X_{(99)}$ or $563,899.

In the above six examples a particular size of loss distribution was assumed. The PML estimates for the sample data are summarized in Exhibit III. While these estimates vary considerably, this is due to differing data and risk aversion considerations. The methods presented work equally well if the distribution of size of loss as a percentage of value is available. The former is more correct for liability insurance or for property insurance if the population has the same property value as the insured. The latter is more correct for property insurance where the property values differ among risks.

SUMMARY

This paper has presented two different approaches to the PML problem using order statistics: $X_{(n)}$ and quantiles. These approaches lead to six different methods for estimating the PML:

1. $X_{(n)}$,
2. $E(X_{(n)})$,
3. $E(X(n)) + k(var(X_{(n)}))^{1/2}$,
4. distribution-free upper bound of $E(X_{(n)})$,
5. $X_{(r)}$ as an estimate of κ_p, and
6. distribution-free upper bound of κ_p.

Methods 1, 5 and 6 require sample data. Methods 2 and 3 require assumptions about n and the underlying distribution of the population. Method 4 requires only estimates of n and the mean and variance of the population. The choice of method would depend on availability of data, willingness to make assumptions about the underlying population, and the amount of risk the insured is willing to accept or the underwriter is not willing to accept.

Ordered Sample Data

i	$X_{(i)}$	i	$X_{(i)}$
(1)	(2)	(3)	(4)
1	$ 19,874	51	$207,196
2	30,610	52	208,959
3	32,159	53	209,568
4	34,115	54	213,084
5	40,660	55	214,307
6	53,453	56	214,546
7	56,598	57	215,978
8	61,651	58	216,369
9	63,411	59	220,808
10	66,007	60	222,804
11	73,062	61	224,417
12	76,962	62	224,475
13	87,348	63	235,209
14	96,498	64	238,249
15	98,408	65	238,679
16	109,837	66	238,842
17	122,838	67	240,455
18	128,372	68	244,699
19	128,426	69	247,465
20	130,048	70	251,374
21	130,610	71	257,426
22	131,326	72	258,513
23	131,474	73	265,051
24	137,655	74	269,816
25	139,681	75	271,647
26	140,949	76	274,154
27	147,987	77	275,727
28	150,776	78	277,211
29	151,044	79	277,734
30	151,967	80	279,494
31	152,219	81	280,721
32	153,388	82	293,728
33	154,619	83	302,641
34	157,065	84	308,771
35	162,956	85	311,612
36	169,142	86	314,410
37	170,262	87	319,722
38	171,988	88	323,711
39	173,391	89	327,927
40	174,049	90	331,179
41	175,689	91	345,130
42	180,406	92	368,095
43	182,223	93	371,194
44	183,399	94	396,911
45	190,532	95	434,449
46	195,658	96	440,639
47	197,482	97	447,171
48	199,788	98	482,259
49	203,310	99	563,899
50	205,796	100	576,525

mean=$212,521
standard deviation=$110,506

Exhibit II

Binomial Probabilities for n = 100, p = .95

i (1)	$X_{(i)}$ (2)	$\binom{100}{i}(.95)^{i}(.05)^{100-i}$ (3)
90	331,179	.016716
91	345,130	.034901
92	368,095	.064871
93	371,194	.106026
94	396,911	.150015
95	434,449	.180018
96	440,639	.178143
97	447,171	.139576
98	482,259	.081182
99	563,899	.031161
100	576,525	.005921

Exhibit III

Summary of Example PML Calculations

Method		PML Estimate
1.	$X_{(n)}$	\$576,525
2.	$E(X_{(n)})$	589,468
3.	$E(X_{(n)}) + k(\mathrm{var}(X_{(n)}))^{1/2}$	803,420
4.*	upper bound of $E(X_{(n)})$	988,044
5.	$X_{(r)}$ as an estimate of κ_p	434,449
6.*	upper bound of κ_p	563,899

*These are distribution free.

ESTIMATING PROBABLE MAXIMUM LOSS WITH ORDER STATISTICS

by Margaret Wilkinson

REVIEWED BY Albert J. Beer

With the increased importance of utilizing quantitative analysis in risk management decision-making, Ms. Wilkinson's paper should provide our profession with a valuable use of the concept of probable maximum loss (PML) which has been a fixture of the insurance vernacular for decades. Previously, underwriters had used the PML (or other related tools) to establish the range for the "working layer of coverage". While it was always acknowledged that a larger loss was possible, the PML estimated the expected maximum loss potential for the risk, with the exposure beyond the PML being treated by a catastrophe (risk) load. Today, the dramatic increase in the amount of risk retained by insureds has made the pricing of large accounts more complex since the "buffer" of the working layer is no longer available. It is at these extreme values that the author's work with order statistics may provide a variety of applications.

Before I discuss the results of the paper, I would like to resolve what I percieve to be an ambiguity in the treatment of PML as defined by the author. In my opinion, any discussion of PML is unclear without a quantification of the term "probable". If a pair of dice are rolled, is it reasonable to say the total will "probably" be less than eight ($p = 21/36$)?.....less than ten ($p = 30/36$)?.....less than twelve ($p = 30/36$)? How certain of an outcome must one be in order to say it is probable? It is precisely this subjectivity that leads to the potential conflict between the insured and the carrier which is alluded to by the author. This dilemma could easily be resolved by quantifying the term "probable". McGuinness accomplishes this by means of a reference to a "stated proportion of all cases" which will be equaled or exceeded by the PML. This concept is similar to the confidence coefficient of a one-sided confidence interval. With these ideas in mind, I would suggest that the PML could be redefined as follows:

> Definition: PML_{α} is that amount (or proportion of total value) which will equal or exceed 100α % of all losses that are incurred.

For example, $PML_{.95}$ would represent that amount which would be expected to equal or exceed 95% of the losses incurred by the risk.

If the PML_{α} is so defined, an insured and underwriter who agree on the underlying loss distribution would arrive at the same PML_{α}. It is true that the respective risk aversion and risk acceptance levels would certainly effect the degree of satisfaction each would have at various α levels. However, at any fixed α point, there would be technical agreement on PML_{α}. The "negotiation" on the appropriate price for risk transfer would at least have a common starting point.

Ms. Wilkinson's definition of PML as the "worst loss likely to happen" does not include any quantification of the term "likely". Therefore, as is noted in the paper, the PML estimates that appear in Exhibit III are not approximating the same quantities. For example, the n'th sample order statistic $X_{(n)}$ is intended to be an estimator for the upper bound of the loss variate X. $X_{(n)}$, therefore, is more closely related to the maximum possible loss. Clearly, this is not the same concept McGuinness had in mind when he discussed the generalized PML. It may be noted that my suggested definition of PML_{α} allows for this degenerate case by choosing $\alpha = 1.00$. (Of course, it may not be technically possible to derive a $PML_{1.00}$ if the distribution has no finite upper bound.) On the other hand, using $X_{(95)}$ as an estimate for $k_{.95}$, the 95'th quantile, is equivalent to approximating $PML_{.95}$. I will try to demonstrate that the results displayed in Exhibit III are much more consistent than they appear .

Throughout this discussion , an attempt will be made to provide more general results derived from the author's excellent foundation. I hope these additional comments help to clarify any imprecision in the PML concept.

General Results Concerning $X_{(n)}$

This section concisely presents the theory upon which most of the remainder of the paper is based. In addition to the results which appear, the corresponding distribution for $X_{(r)}$ could be given by:

$$f_{X_{(r)}}(x) = \frac{n!}{(r-1)!(n-r)!}\,(F_X(x))^{r-1} f(x)\,(1 - F_X(x))^{n-r}$$

The reason for introducing this more general result is to allow for the derivation of properties of $X_{(r)}$ similar to those presented for $X_{(n)}$. In particular, it may be shown that the order statistics from a Uniform distribution over (0,1),with

$$u_{(r)} = F_X(x_{(r)})$$

have a Beta distribution with parameters a = r , b = n - r + 1.

Therefore, $$E(u_{(r)}) = \frac{r}{n+1}$$

$$Var(u_{(r)}) = \frac{r\,(n-r+1)}{(n+1)^2(n+2)} \qquad \text{for } r = 1,2,\ldots,n$$

Additionally, the first approximations displayed in the paper as (4) and (5) can be extended to :

$$E(X_{(r)}) \doteq F_X^{-1}\left(\frac{r}{n+1}\right)$$

$$\mathrm{Var}(X_{(r)}) \doteq \frac{r\,(n-r+1)}{(n+1)^2\,(n+2)} \left(f_X\left[F_X^{-1}\left(\frac{r}{n+1}\right)\right]\right)^{-2}$$

These results form the basis of the author's initial three estimates of PML. Using the generalized forms above (with $r = 100\alpha$), estimates for our PML_α may be computed as follows:

	Estimates for :			
Method	$PML_{.90}$	$PML_{.95}$	$PML_{.99}$	$PML_{1.00}$
1) $X_{(r)}$ from sample data	\$331,179	\$434,449	\$563,899	\$576,525
* 2) $E(X_{(r)})$	344,158	404,453	516,532	589,468
* 3) $E(X_{(r)}) + 2(\mathrm{Var}(X_{(r)}))^{\frac{1}{2}}$	399,632	482,839	662,380	803,420

* These methods assume an underlying lognormal distribution with μ = \$212,521 and σ = \$110,506.

It may be noted that the $PML_{1.00}$ estimates are those derived in the paper under the author's definition of PML.

Using $X_{(r)}$ As An Estimate for the PML

Although this is obviously the most convenient approach, it relates only to the data that are available from reported claims and may not be an accurate indication of the underlying exposure in the future. For example, immature loss history may not show any losses in excess of a few thousand dollars. Should the PML be chosen to be the largest claim paid to date?.....or the largest reported claim?.....or some other choic

From another point of view, suppose $X_{(99)}$ = \$400,000 and the larges claim $X_{(100)}$ = \$2,000,000. Is the \$2,000,000 loss catastrophic and, by definition, not probable? Clearly $X_{(n)}$ alone should not be used in any