A PRACTICAL GUIDE TO THE SINGLE PARAMETER PARETO DISTRIBUTION

STEPHEN W. PHILBRICK

Abstract

The actuarial literature has discussed several candidates for size-of-loss distributions—log normal, Weibull, multi-parameter Pareto, gamma, as well as others. However, despite the demonstrated success of these distributions, there is a dependence on techniques such as empirical data, judgment, or at times some unwieldy formulae. This suggests that there may be a need for a size-of-loss distribution that is relatively easy to apply in practice.

The one-parameter Pareto is an example of such a distribution. Its use may be restricted to the tail of a distribution, but it is easy to apply. The formulae for the mean, variance, and the variance of the aggregate loss distributions are simple in form and may be used as quick approximations in many cases.

I. INTRODUCTION

"The ultimate goal of model-building is either as a tool for communicating . . . or for predicting and making decisions . . ."

—William S. Jewell

Although model-building is common to many branches of science, there are important distinctions among the properties of various models. The laws of physics such as Newton's laws are attempts at mathematical models of reality. These efforts have been particularly successful because the major forces at work are few in number and often constant over time and position. Although technically there are many forces involved in, say, the movement of the planetary bodies, the dominant force of gravity dwarfs the other forces such as friction. Models can be developed based solely on the properties of the gravitational force which describe the motion of the planets to a very high degree of precision. In these situations, it is common to find mathematical models with few parameters that are highly accurate models of reality. It is appropriate, even if technically incorrect, to speak of the search for *the* correct mathematical model.

In the social sciences the situation is quite different. The forces involved in economics, for example, are numerous and usually not constant over time. Many forces exist that have the same order of magnitude; hence they cannot be ignored. Furthermore, in the social sciences it is often more difficult to do controlled experiments where one force is allowed to vary while all others are held constant. For these reasons, it is less appropriate to think of a search for *the* model in the social sciences than in the physical sciences. Although we might talk about such a concept theoretically, the practical reality is that any parameter-based model that completely describes an existing situation will require so many parameters as to make it unusable. In these situations, model-building requires a trade-off between accuracy and practicality.

Thus, the question "What is the appropriate loss distribution?" does not have a unique answer. It depends on the intended use of the distribution and the available data.

The question requires a cost-benefit analysis. Different models will have various costs related to:

- Mathematical complexity,
- Availability of computer/calculator software routines,
- Computer processing time requirements,
- Conceptual simplicity (ease of explanation to others), and
- Availability and accuracy of data.

Generally speaking, increasing sophistication of the model produces more accurate results. The selection of an appropriate model for a particular problem requires deciding whether the increased accuracy of the more complex model justifies the increased costs associated with it. Furthermore, in many situations the available data may be sparse or subject to inaccuracies. In these instances, a simple model may be preferred because the accuracy of results will not be materially improved by the use of a more complex model.

For example, suppose an actuary is trying to solve a typical risk management problem: the projection of losses for an individual risk. A common procedure in this analysis is to separate the projections of the large or excess losses from the projections of the more stable primary portions of the losses. Several characteristics of this situation make a simple model particularly appropriate.

- The projection of the limited losses may be accomplished without the need for a specific size-of-loss distribution. The moments of the data, using a frequency/severity or total loss approach, may be sufficient for a reasonable projection. It will then be necessary to fit a model only above a particular loss amount. Fitting a distribution to only a portion of the range will reduce the required complexity of the model.

- Inaccuracy of estimates of expected losses arises from a number of sources. Two major ones are:
  —Oversimplified models, and
  —Misestimated parameters.
  In a situation involving an individual risk, the number of large losses used to estimate the parameters will typically be less than the number involved in an insurance company or industry analysis. The errors arising from the sample size may dominate those arising from a less complex model. As a consequence, the simplicity of the less complex model may be preferred because the possible loss of accuracy is more than offset by the benefits of a simpler model.

- There may be a need to explain the loss projection process to people without extensive actuarial or statistical training. Although techniques should not, in general, be dictated by the sophistication of the audience, if competing models produce almost identical results, the ease of explanation of one may be an important consideration.

The remainder of this paper will be organized as follows:

- Section II—A discussion of the way distributions are depicted. An alternative to the "standard" representation will be presented.

- Section III—A discussion of the basic properties of the single parameter Pareto distribution.

- Section IV—Various methods of parameter estimation using empirical data.

- Section V—The results of trend on losses when a Pareto distribution is assumed.

- Section VI—A method to simulate Pareto losses.

- Section VII—Specific applications using a Pareto distribution.

The author would like to acknowledge the help of several people, including Jerry Miccolis and Claudia Leme, who reviewed early drafts of this paper; Kurt Reichle and John Yonkunas, who provided many helpful suggestions and encouragement; and Jerry Jurschak, who contributed some of the concepts embodied in this paper.

II. SIZE-OF-LOSS REPRESENTATIONS

Most texts on probability and statistics portray distributions (as well as density functions) using similar conventions. That is, the horizontal axis represents the value of the observations and the vertical axis represents the relative frequency (for density functions) or the cumulative frequency (for distributions). It is clear, from a mathematical point of view, that this choice is arbitrary. The axes could be switched without violating or changing any of the statistical concepts.

Despite the almost universal acceptance of this "standard" representation, the alternative representation turns out to be a clearer choice for size-of-loss distributions in some situations. The reason for this preference is that this representation can be developed in a "natural" way and will allow a number of concepts, such as loss limitation (truncation and censorship), to be applied in a more intuitive fashion. Appendix D contains a more detailed discussion of reasons for preferring this orientation.

In the following discussion, we will develop a size-of-loss representation where the y-axis is the horizontal axis and the x-axis is the vertical axis. We will refer to this representation as the "alternative" representation.

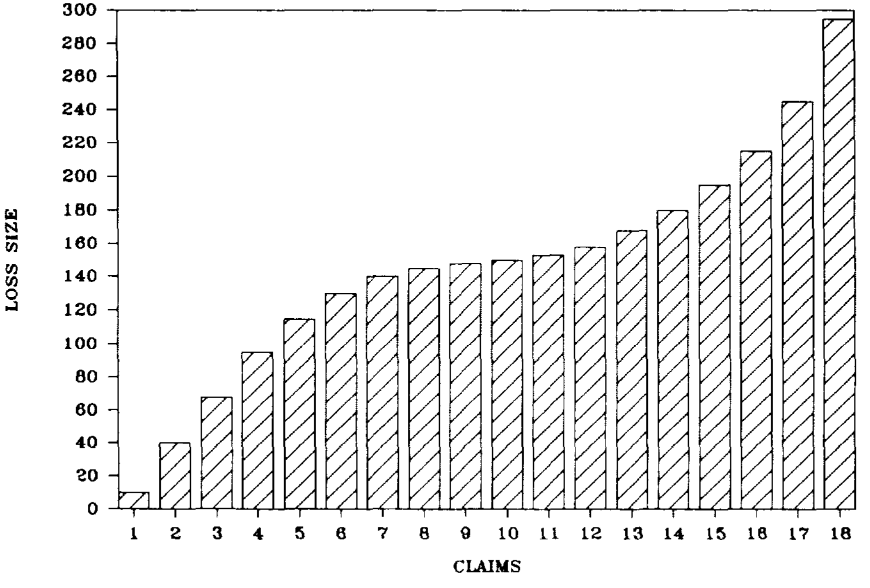
Because we have switched the axes rather than redefined them, the definitions of x and y will remain unchanged; that is, x refers to loss amounts and y refers to cumulative frequency.

Discrete Case

Consider a set of n losses from some arbitrary size-of-loss distribution where each loss has size S_i, $i = 1, 2, \ldots, n$. Represent each loss by a rectangle with

FIGURE 1

SIZE-OF-LOSS DISTRIBUTION



width one and height S_i. Arrange these losses from smallest to largest, each perpendicular to the y-axis. Figure 1 displays a typical example of such a procedure.

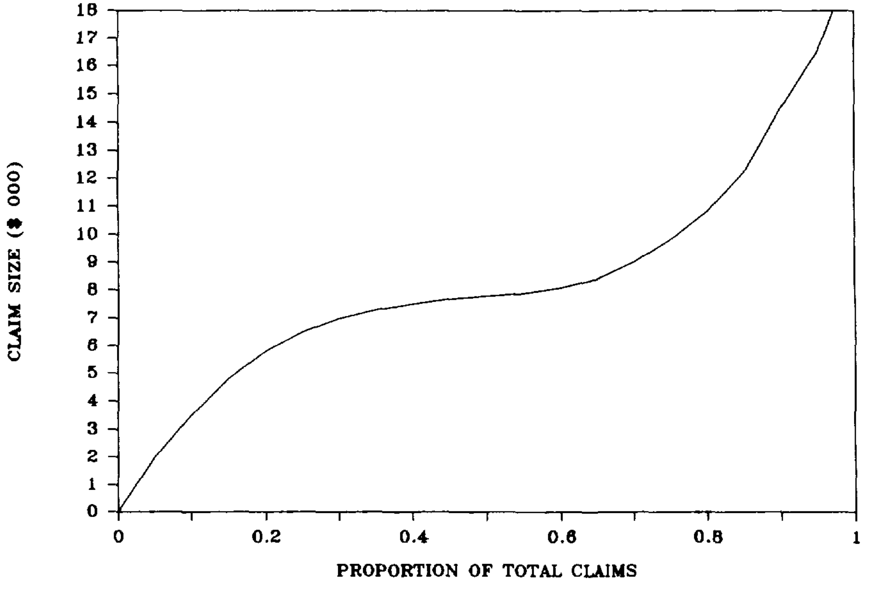
Define $G(y)$ to be the curve represented by the tops of each of the rectangles. Then, $G(y) = S_i$ for $i-1 < y \leq i$. Note that the interpretation of the random variable Y is the number of losses less than or equal to $G(y)$ (for integral values of y).

Continuous Case

When we consider the continuous case, the width of each loss is dy. The value of y ranges from 0 to 1 and represents the percentage of losses less than or equal to $G(y)$. A typical example would look like Figure 2.

FIGURE 2

CLAIM SIZE DISTRIBUTION



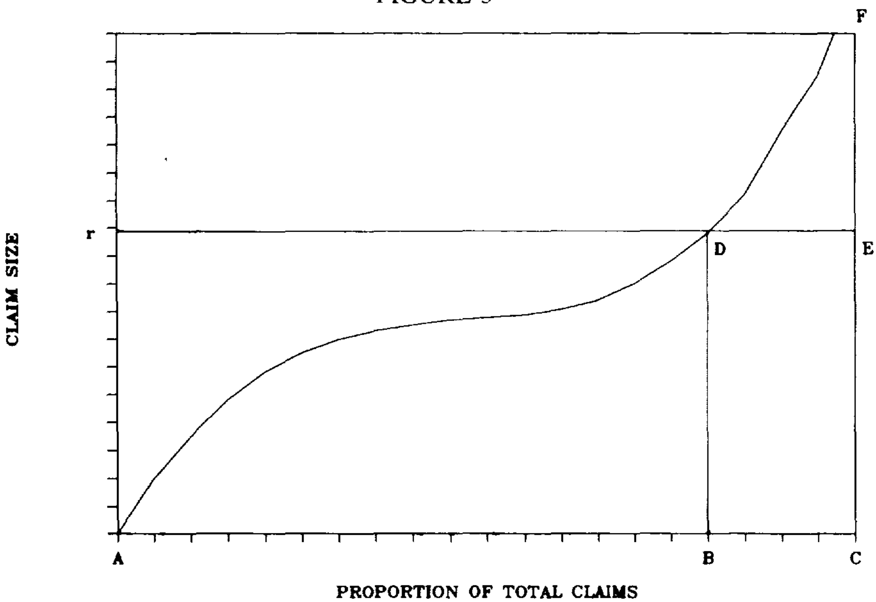
From this point on, the continuous version of the representation will be used. However, some of the concepts may be better understood if the original motivation of this representation is recalled, namely "stacking" individual losses along the y-axis.

When we work with a set of losses (whether actual or theoretical), we generally wish to partition these losses in some way. The most common partitions are "large" versus "small" and primary versus excess. These partitions can be graphically represented by defining areas under the curve $X = G(y)$.[1]

Generally, we will indicate the losses of interest by defining one or more straight lines on the graph (see Figure 3). When we define an area by a pair of lines parallel to the horizontal axis, we will refer to these losses as a "layer" of losses. Alternatively, if we use a pair of lines that are parallel to the vertical

[1] A third type of partition is described in Hewitt and Lefkowitz [H1]. That partition cannot be handled in this way.

FIGURE 3



axis, these will be referred to as "interval" losses. In this case, we are referring to those losses that correspond to an interval specified on the (horizontal) axis.

We could define the areas we are interested in by directly writing the integral over the appropriate limits. However, we can keep the notational complexity to a minimum if we adopt symbols for the areas that will be used most often.

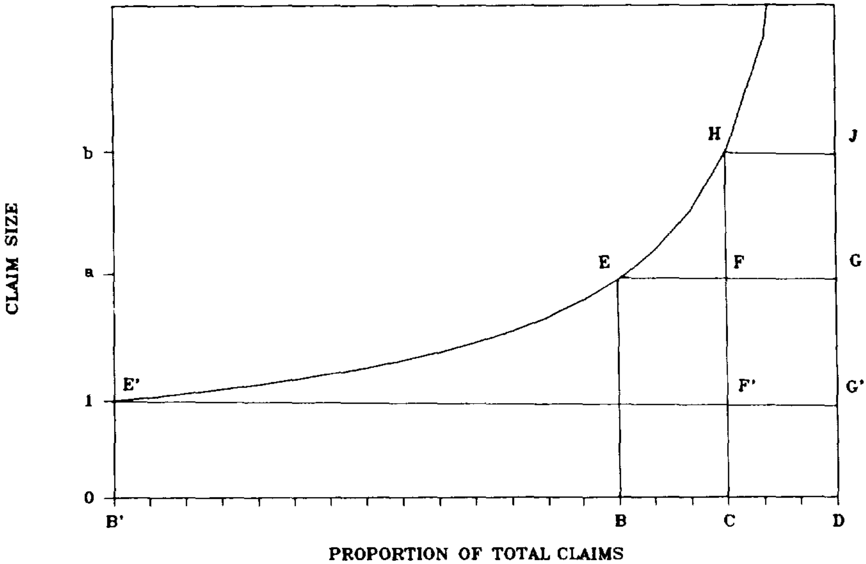
Given loss amount, r, we define[2]:

	Verbal	Mathematical
$T(r)$ —	The average claim size of all losses less than or equal to r; i.e., losses are *truncated* at amount r.	$\dfrac{\int_0^r x\, dF(x)}{\int_0^r dF(x)}$

[2] The reader may note that the notation used here is not entirely consistent with that developed in a discussion of LaRose [L1]. The notation developed by LaRose calculates claim amounts as percentages of the *average* claim. Unfortunately, the average claim size is not always well-defined, so a more general notation is required.

FIGURE 4

LARGE LOSS DISTRIBUTION



$C(r)$ —	The average claim size of all claims where the amount of each claim is limited to size r; i.e., losses are *censored* at amount r.	$\int_0^r x\, dF(x) + r[1 - F(r)]$

For example, if r is \$100,000, then $T(\$100{,}000)$ represents the average of all losses less than or equal to \$100,000. In Figure 3, this average would be represented by the ratio of the area bounded by ABD divided by the number of claims in the interval. The quantity $C(\$100{,}000)$ is the average of all claims where amounts greater than \$100,000 are capped or limited to \$100,000.

The preceding discussion is applicable to any size-of-loss distribution. Figure 3 applies to any distribution that is used to model the entire range of losses. In the remainder of this paper, we will work with the tail of a loss distribution that is applicable to "large" losses. Consequently, we will truncate the loss distribution at some value and remove each loss less than that value. A typical distribution representing the remaining "large" losses is shown in Figure 4.

Figure 4 is derived from Figure 3 by truncating the loss distribution at loss amount r. Physically, we remove the portion of the graph to the left of the vertical line BD, then renormalize our axes so that the y-axis is the cumulative percentage of the "large" losses, that is, losses greater than or equal to r. It should be emphasized that Figure 3 is not drawn to scale for typical loss distributions. If we select a lower limit r such as $25,000, the cumulative probability that a claim is less than $25,000 (which is represented by point B) is typically in excess of 90%. We will work only with the large losses in the remainder of this paper, so Figure 4 is the important figure to keep in mind.

III. BASIC PROPERTIES OF THE SINGLE PARAMETER PARETO

The Pareto distribution as described in Johnson and Kotz [J1] has cumulative distribution function:

$$F(x) = 1 - \left(\frac{k}{x}\right)^{a} \qquad k > 0;\ a > 0;\ x \geqslant k$$

This is also known as the "Pareto distribution of the first kind." Strictly speaking, this distribution has two parameters, k and a. In general, both k and a may be estimated from the data. However, the verbal definition of k is the lower bound of the data in question. Although there may be situations where this value must be estimated, in virtually all insurance applications this value will be selected in advance. The typical insurance application will be to model losses whose value is in excess of some pre-selected size, such as $25,000 or $100,000.

Furthermore, if we "normalize" our losses, that is, divide each loss by the selected lower bound, then the normalized lower bound is 1, and the parameter does not need to be stated explicitly. Finally, we will use q as the parameter, rather than a, to be consistent with ISO usage (ISO [I1], p. 34). The distribution can then be written as:

$$F(x) = 1 - x^{-q} \tag{1}$$

and the density function is

$$f(x) = qx^{-(q+1)} \tag{2}$$

This is the distribution that will be discussed in the remainder of this paper.[3] Typical values for q can range from .7 to 2.0, although values outside this

[3] See Appendix C for a discussion of alternative forms of the Pareto.

range are possible. A typical value for q of property losses is 1.0, while a typical value for casualty losses is 1.5 (based upon empirical evidence). Note that a *low* value of q corresponds to a distribution with high severity. Fire may not be thought of as a line with high severity, but that is because there are so many very small claims. Considering only larger claims, e.g., claims greater than $25,000, fire claims have a fairly "thick" tail. The density function for a Pareto with parameter $q = 1.5$ is shown in Figure 5; the corresponding c.d.f. is shown in Figure 6.

If we "flip" the x- and y-axes of the cumulative distribution, we will produce Figure 4. Note that the curve intersects the x-axis at $x=1$, because we have normalized the losses. The curve is asymptotic to $y=1$. As mentioned earlier, we can visualize the area under the curve as being made up of thin vertical rectangles whose height corresponds to the size of loss. Thus the total area under the curve represents the total losses, and the losses associated with various retentions or policy limits can be described by different areas under the curve.

The distribution as shown in Figure 4 is based upon the assumption that the lower limit is 1 and the expected frequency of claims greater than or equal to this value is also 1. Formulae will be derived under these assumptions. The necessary conversions to real problems are simple and straightforward. Examples of conversions will be given in most cases. Although it may seem awkward at first to work with a normalized distribution, it will soon become very natural. The motivation for using the normalized distribution should become clear when we analyze the losses contained in different layers.

Unlimited Claims

The formula for the average claim size is as follows:

$$\text{Unlimited Mean Claim Size} = \frac{q}{q-1} \qquad q > 1 \qquad (3)$$

Note that this formula also represents the expected total losses when the expected frequency is 1 (assuming independence of frequency and severity).

If the data being analyzed has a lower limit of $K per claim, then the mean size in "real" dollars is:

$$\text{Unlimited Mean Claim Size} = K\left(\frac{q}{q-1}\right) \qquad (3a)$$

FIGURE 5

Pareto Density

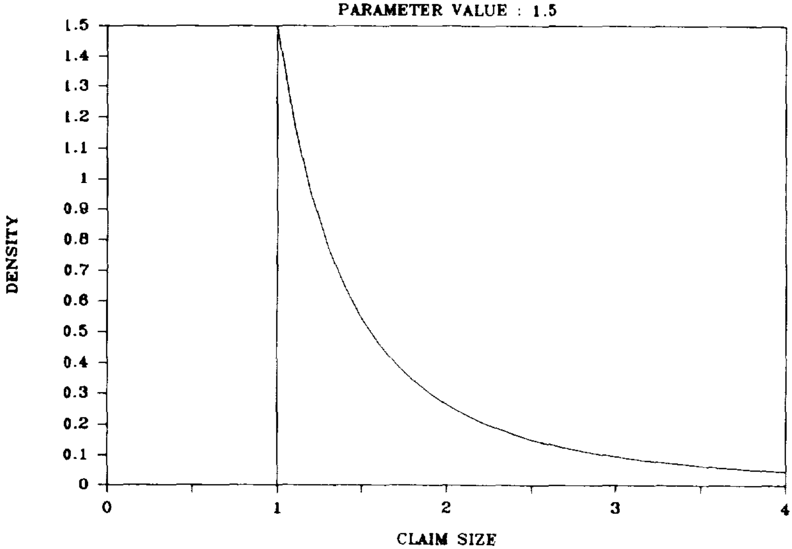
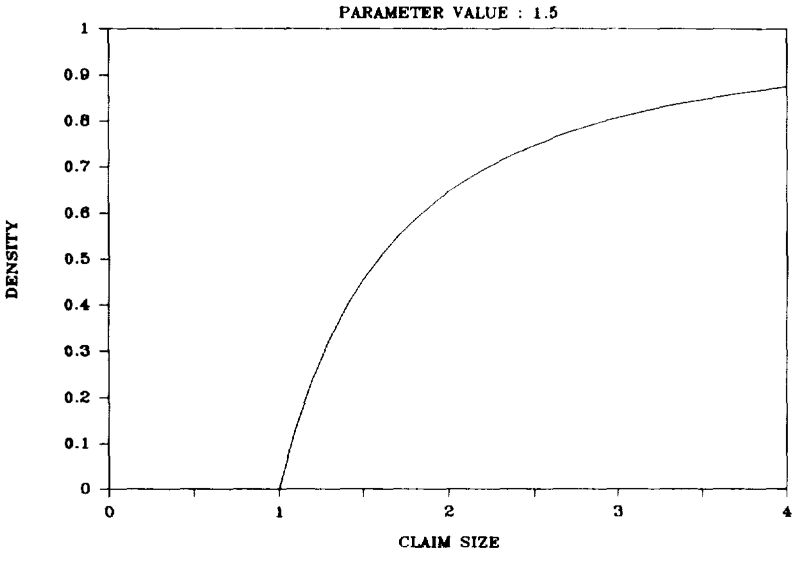


FIGURE 6

Pareto Cumulative Distribution



If we anticipate n claims greater than or equal to \$$K$ per claim, then the total expected losses are:

$$\text{Unlimited Expected Losses} = n\,K\left(\frac{q}{q-1}\right) \tag{3b}$$

For example, suppose we are analyzing claims where the lower limit is \$25,000. That is, all claims are greater than or equal to \$25,000. After normalizing our losses (dividing each by \$25,000) we conclude that a parameter value of $q = 1.5$ is appropriate. (A later section will discuss parameter estimation.) Then the normalized gross mean claim size is $1.5/(1.5 - 1) = 3$. In terms of "real" dollars, where $K = \$25{,}000$, the gross mean claim size is $3 \times \$25{,}000 = \$75{,}000$.

If we expect 7 claims to exceed \$25,000, then our gross expected losses are $7 \times \$75{,}000 = \$525{,}000$. (Again, it should be remembered that we are analyzing the large claims only. The expected losses arising from claims less than \$25,000 are assumed to be estimated separately.)

We may wish to calculate the net losses, for example, if we have a \$25,000 deductible. The formula for the net mean claim size is derived from the gross mean claim size simply by subtracting 1:

$$\text{Net Mean Claim Size}\left(\frac{q}{q-1}\right) - 1 = \frac{1}{q-1} \qquad q > 1 \tag{4}$$

The conversion to "real" dollars and total losses follows the same approach as above. For example, with $q = 1.5$, $K = \$25{,}000$ and a frequency of 7, the net expected losses above \$25,000 are $7 \times \$25{,}000 \times 1/(1.5 - 1) = \$350{,}000$.

Censored Claims

If we impose an upper limit (such as a policy limit) with value b, then the formula for the average loss limited to \$$b$ per claim is:

$$C(b) = \frac{q - b^{1-q}}{q-1} \qquad q \neq 1 \tag{5}$$

If $q = 1$ we can calculate the formula using L'Hôpital's Rule:

$$C(b) = 1 + \ln b \qquad q = 1 \tag{6}$$

In the case where we want net losses, we can simply subtract 1 from each formula. Note especially, when $q = 1$ the average loss with upper limit b is simply $\ln b$.

Continuing our previous example (with $q = 1.5$ and lower limit of \$25,000), if we impose an upper limit of \$500,000, then $b = 20 \times (500{,}000/25{,}000)$. The average claim whose value is greater than \$25,000 but limited to \$500,000 can be calculated using (5):

$$\frac{1.5 - 20^{1-1.5}}{1.5 - 1} = 2.553$$

In "real" dollars, the average claim is 2.553× \$25,000 = \$63,820 (calculations here and subsequently are performed without rounding at intermediate steps). If we are pricing reinsurance for the net layer (\$475,000 xs \$25,000), then we would subtract 1 first to calculate the net claim size: 1.553 × \$25,000 = \$38,820. Assuming we expect 7 claims over \$25,000, the expected losses in the layer are 7 × \$38,820 = \$271,738.

Truncated Claims

The situation described above (with an upper *censorship* limit) arises naturally in practice because of the existence of policy limits and the way companies commonly write excess of loss reinsurance. Another way to limit losses is to *truncate* the losses at some value. This means that losses in excess of the truncation point are "thrown away," rather than simply "capped" at the limit. Note that this is different from censorship in two ways:

1. More dollars are removed when losses are truncated at a value because the entire loss above the limit is removed.
2. Truncation affects the frequency. Censorship removes the excess portion of a claim, but does not affect the number of claims. Truncation removes the *entire* claim, so the formulae for average values must reflect the reduced claim count.

The concept of truncation arises rarely in property/casualty policy language (with the rare exception of franchise deductibles). However, the concept may arise in the analysis of experience. For example, it might be appropriate to separate losses into large versus small (rather than primary versus excess). In this case, the limit chosen to distinguish between large and small losses will be a *truncation* point, rather than a *censorship* point.

The formula for the average claim size with lower value 1 and truncation point b is:

$$T(b) = \frac{q(1 - b^{1-q})}{(q - 1)(1 - b^{-q})} \qquad q \neq 1 \qquad (7)$$

The comparable formula for the case $q = 1$ is:

$$T(b) = \frac{b \ln b}{b - 1} \qquad q = 1 \qquad (8)$$

Continuing our example, suppose we are interested in the losses larger than \$25,000 but ignoring all losses greater than \$500,000 (rather than including the first \$500,000 of those claims greater than \$500,000). With $q = 1.5$ and $b = 20$, the average claim size is calculated with (7):

$$\frac{1.5\,(1 - 20^{-.5})}{.5\,(1 - 20^{-1.5})} = 2.356$$

In "real" dollars, the average claim is 2.356 × \$25,000 = \$58,888. If we expect 7 claims over \$25,000 we can calculate the total dollars for the interval. Given 7 claims over \$25,000, we expect $7 \times F(20) = 7 \times .9888 = 6.922$ claims in the interval between \$25,000 and \$500,000. We have already calculated the average of those claims, so we multiply the frequency by the average claim size to yield the total dollars: 6.922 × \$58,888 = \$407,606.

The formula for the average truncated size follows directly from the definition of truncated claims given earlier. However, we can simplify the formula and reduce the amount of calculation by adopting a slightly non-standard convention. Note that, in our example, one of the terms in the denominator for the average truncated claim size is $1 - b^{-q}$. Note also that the number of claims in the *interval* is calculated by multiplying the expected frequency (above the lower limit) by $F(b)$ which is $1 - b^{-q}$. Obviously, these terms cancel out when the total dollars in the interval are calculated.

Define $T'(b)$ to be the average claim size, where the denominator is not just the claims in the interval, but the number of claims above the lower limit. In other words, use the same denominator as in the censored situation. The motivation is two-fold:

(1) The formulae will be simpler.
(2) It is more likely that we will have an estimate of the total number of claims above a limit than that we will have an estimate of the number of claims in an interval.

The formula for the revised "average" truncated claims size is:

$$T'(b) = \frac{q(1 - b^{1-q})}{q - 1} \qquad q \neq 1 \qquad (9)$$

When $q = 1$, the formula simplifies to:

$$T'(b) = \ln b \qquad q = 1 \qquad (10)$$

Redoing the example above, the "average" claim size is calculated using:

$$\frac{1.5\ (1 - 20^{-.5})}{.5} = 2.329$$

In "real" dollars, the "average" claim size is 2.329 × \$25,000 = \$58,229. Multiplying this by the number of claims expected over \$25,000 yields 7 × \$58,229 = \$407,606.

In summary, if we are interested in the true average claim size, we use formula (7) or (8). However, if the calculation of the average claim size is simply an intermediate step in the calculation of the total dollars, we may prefer to use alternative formula (9) or (10).

Next we will look at the excess portion of the distribution. In this case, we are interested in the total losses or average claim size of claims greater than some limit b. In terms of Figure 4, the area of interest is bounded by HJK. Rather than directly calculate the total losses and average losses in this layer, we will exploit a powerful property of the Pareto distribution. If we renormalize the excess portion by dividing each loss by b and dividing the excess frequency by $1 - F(b)$, the resulting distribution will have a shape identical to that in Figure 4. (This renormalization is the result of a scale change to both axes. For more discussion of scale changes, see Venter [V1].) Thus, we may use the formulae already calculated, although keeping careful track of the appropriate factors to convert back to "real" dollars.

The average gross claim size is still $q/(q - 1)$ and the average net claim size is $1/(q - 1)$. In terms of our first renormalization, the average gross claim size is $b\ (q/(q - 1))$ and in "real" dollars, the average is $bK\ (q/(q - 1))$. The total dollars involved in claims greater than b can be calculated by multiplying by the frequency of claims greater than b which is $1 - F(b) = b^{-q}$.

In practice this works out easier than the formulae would indicate. Continuing our example ($q = 1.5$, $K = \$25{,}000$, frequency over \$25,000 = 7), suppose we are interested in the losses in excess of \$100,000 per claim. We don't actually perform the renormalization; we simply use the formula for net average claim size ($1/(q - 1)$) and substitute $q = 1.5$ into the formula yielding a net claim size of 2. Multiply by \$100,000 (it isn't necessary to multiply first by \$25,000, then by 4) to produce the average net claim size of \$200,000. To calculate the total dollars, recall that the ratio of claims exceeding $100,000 is calculated by using the cumulative distribution $1 - F(b) = b^{-q} = 4^{-1.5} = .125$. Multiply this by the expected frequency over $25,000 of 7 yielding .875 claims expected to exceed $100,000. Thus, the expected excess losses are $.875 \times \$200{,}000 = \$175{,}000$.

This concept is important, as it allows us to quickly calculate the total losses and average claim sizes for arbitrary layers and intervals. As another example, suppose we continue our assumption that losses over $25,000 have a Pareto distribution with $q = 1.5$ and the expected frequency of claims over $25,000 is 7. Suppose we are asked to analyze the layer between $75,000 and $187,500 (i.e., $112,500 xs $75,000). The first step is to calculate the value of b, which is simply $187{,}500/75{,}000 = 2.5$. We can use (5) to calculate the gross average (censored) claim sizes:

$$\frac{1.5 - 2.5^{-.5}}{.5} = 1.735$$

The net average claim size is .735 or $.735 \times \$75{,}000 = \$55{,}132$ in "real" dollars. The frequency of claims is $7 \times (1 - F(75{,}000/25{,}000)) \times F(187{,}500/75{,}000) = 7 \times (3^{-1.5}) \times (1 - 2.5^{-1.5}) = 7 \times (.192) \times (.747) = 1.006$, so the expected losses in the layer are $1.006 \times \$55{,}132 = \$55{,}482$.

Next, we will calculate the variance of the individual claim amounts as well as the total loss variance. The formulae shown above for expected values are sufficient for pricing on an expected value basis or some function of the expected value. However, there are methods of pricing that include risk loading based upon variance, as well as other risk theoretic analyses that require the calculation of variances. (See Gerber [G1] for a discussion of various pricing approaches.)

Again, this is one of the motivations for the use of the Pareto. The calculation of total loss variance is a fundamental issue in risk theory, yet the procedures necessary to calculate the variance generally involve complex formulae or, more likely, computerized estimation techniques. The formulae associated with the single parameter Pareto are often easy to evaluate and may provide, at the very least, a reasonable first approximation.

Recall that the variance can be calculated as the second moment minus the square of the mean. The formula for the n^{th} moment of the Pareto distribution with no upper limit is

$$n^{\text{th}} \text{ moment} = \frac{q}{q + n} \qquad (11)$$

Thus, the second moment is $q/(q + 2)$ and the formula for the variance of a single claim is:

$$\text{Variance} = \left(\frac{q}{q-2}\right) - \left(\frac{q}{q-1}\right)^2 \qquad q > 2 \qquad (12)$$

Again, we have the problem that the variance is undefined for typical values of q. But if we restrict ourselves to reasonable upper limits, the variance will always be finite. If we impose upper limit b, then the variance of losses within the layer is:

$$\text{Variance} = \frac{q - 2b^{2-q}}{q-2} - \left(\frac{q - b^{1-q}}{q-1}\right)^2 \qquad q \neq 1,\ q \neq 2 \qquad (12a)$$

The formula simplifies in the cases where $q = 1$ or 2 as follows:

$$\text{Variance} = 2b - 1 - (1 + \ln b)^2 \qquad q = 1 \qquad (12b)$$

$$\text{Variance} = 1 + 2 \ln b - ((2b - 1)/b)^2 \qquad q = 2 \qquad (12c)$$

These formulae apply in either the net or unlimited layer cases.

To convert these results to "real" dollars, multiply by K^2 where K is the lower bound of the losses. It is important to realize that these formulae reflect only the variance associated with the loss severity. The total loss variance also reflects the variability of frequency, which will be covered shortly.

We will continue the example where the lower limit is \$25,000 and $q = 1.5$. As we have shown earlier, the gross mean claim size is 3 and the net mean claim size is 2 when no upper limit is imposed. However, the variance is not defined in this case. With an upper limit of \$500,000, $b = 20$ and the variance of a single claim is calculated by substituting into (12a) with $q = 1.5$ and $b = 20$. The result is 8.372. In "real" dollars, the variance is $8.372 \times (\$25{,}000)^2 = 5.23 \times 10^9$. This means that the standard deviation is \$72,335.

The claim size variance is rarely useful by itself. The major motivation for calculating this formula is because it is needed in the formula for the total loss variance. This refers to the variability of total losses, arising either from frequency or severity. The variance we will calculate is also sometimes called "process variance," because it relates to the possible variations in results arising from the loss causing process. This is to be distinguished from "parameter variance," which relates to the variations arising from the possibility that the parameters used differ from the "true" parameters. Parameter variance is beyond the scope of this paper.

Calculation of the total loss variance is necessary if a risk loading will be used that is a function of either the total loss variance or standard deviation. In addition, the variance can be used to specify percentiles of the total loss distribution using the Cornish-Fisher expansion [M1] or other techniques [L3]. For example, we may wish to determine the probability that total losses will exceed $1,000,000 when the expected losses are $600,000.

The general formula for the total loss variance is given in various sources including Mayerson, Jones and Bowers [M3]:

$$\sigma^2 = M_f \, \sigma_s^2 + M_s^2 \, \sigma_f^2 \qquad (13)$$

where M_f, σ_f^2, M_s, and σ_s^2 represent the mean and variance of the frequency and severity distributions respectively.

If we make the reasonable assumption that the claim frequency follows a Poisson distribution, then $M_f = \sigma_f^2$ and we can simplify (13):

$$\sigma^2 = M_f(\sigma_s^2 + M_s^2) \qquad (14)$$

Again, recalling that the variance can be expressed as the second moment less the square of the mean, we note that the expression in parentheses above simplifies to the second moment of the severity. Thus, the total loss variance can be simply calculated as the product of the expected claim frequency and the second moment (mean of the squares) of the loss severity.

We have seen the formula for the second moment of the severity in the case of no upper limit earlier (11). In this case, the total loss variance is:

$$\sigma^2 = M_f \frac{q}{q+2} \qquad q > 2 \qquad (15)$$

where M_f is the expected claim frequency.

We have seen earlier that the severity variance is the same in the case of the unlimited and net layers. This is not the case for the total loss variance. If we have upper censorship point b, the total loss variance for the unlimited layer is:

$$\sigma^2 = M_f \frac{q - 2b^{2-q}}{q-2} \qquad q \neq 2 \qquad (16)$$

In the case of the net layer, the total loss variance is:

$$\sigma^2 = M_f \left\{ \frac{q - 2b^{2-q}}{q-2} - 2\left(\frac{q - b^{1-q}}{q-1}\right) + 1 \right\} \qquad q \neq 2,\; q \neq 1 \qquad (17)$$

The expression in the parentheses may be recognized more quickly if we recall that $E[(X-1)^2] = E[X^2] - 2\,E[X] + 1$. Formula (16) for the case where $q = 2$ is shown in Appendix A. As before, to convert the results to "real" dollars, we multiply by K^2 where K is the lower limit of the losses used to normalize the values.

If we have the truncated case, with truncation point b, the total loss variance for the unlimited layer:

$$\sigma^2 = M_f \frac{q\,(1 - b^{2-q})}{(q-2)(1-b^{-q})} \qquad q \neq 2 \qquad (18)$$

Note carefully: the definition of M_f in this case is the expected number of claims greater than the lower limit, not simply the number between the lower limit and b. The situation with truncation point b and a *net* layer is almost never seen in practice, so it will not be discussed.

Continuing our example, suppose we are pricing the losses in excess of \$25,000 but censored at \$500,000. As we have seen earlier, the expected losses in this layer are \$271,738 (assuming the expected number of claims is 7). Suppose we wish to add a risk loading that is a function of the total loss variance. We can calculate the total loss variance using (17). Substituting the parameters into the formula yields a variance of 75.48. In "real" terms, this is $75.48 \times \$25{,}000^2$. The standard deviation of this value is \$217,199. We won't go into methods for calculating a factor to multiply by the variance to arrive at an appropriate risk load, but, even without such methods, the total loss variance can be used to compare the relative risk on different treaties.

Finally, we note that the formulae derived in this paper are only applicable to the portion of losses above the selected lower limit. In practical situations, it is necessary to combine the results of the analysis of the large losses with the results of the analysis of the small losses. Clearly, the expected losses of the two portions of the analysis can simply be added together. The overall average claim cost can be calculated as the weighted average of the means of each portion, where the weights are the expected number of claims. The variances of the severity cannot be combined so easily, although, if one recalls that the second moments can be weighted by claim counts, the formula for the combined severity variance follows easily. If we assume a Poisson distribution for the frequency of the small losses, then the total loss variance of the small losses will be of the same form as the large losses, specifically, the mean claim frequency multiplied by the second moment of the severity, so the total loss variance of the entire distribution is simply the sum of the total loss variance of each portion.

IV. PARAMETER ESTIMATION

Numerous articles in the actuarial and statistical literature (e.g., Patrik [P1], p. 62.) discuss the attractive properties of the maximum likelihood estimate (MLE). However, the MLE is often difficult to calculate in practice.

One of the attractive properties of the Pareto distribution is the ease of calculation of the maximum likelihood estimate of the parameter. Consider a set of n losses, each greater than or equal to some value K, which are normalized by dividing each loss by K. Denote this set by $(X_i), i = 1,2, \ldots, n$. The MLE of q is

$$q = \frac{n}{\Sigma \ln X_i} \tag{19}$$

Note that an alternative formula is

$$q = \frac{n}{\ln \Pi X_i} \tag{20}$$

These formulae are equivalent, but the second is easier to calculate. Note also that the MLE of q is such that $e^{1/q}$ is the geometric mean of the X_i. If we use the 25 losses contained in Appendix B, the estimated parameter is $q = 25/26.16 = .955$.

Although the MLE has attractive properties and is easy to calculate, we will examine the use of alternative methods. Probably the most common method is matching of moments. We have shown that the mean of the unlimited Pareto distribution is $q/(q - 1)$. If this is equated to the sample mean of the values in Appendix B, we have

$$\frac{q}{q-1} = 6.202$$

$$q = 1.192$$

This value is not particularly close to the true value. The discrepancy arises, not because of the relatively small sample, but from the method itself. If the formula for the mean is examined, it will be clear that a value of 1.0 could never result. If the true value of the parameter of the distribution is 1.0 or smaller, the method of moments will always produce too high a result. Because in many situations the value of the parameter may be close to or even less than 1.0, the method of moments may not be an appropriate method.

Another method of parameter estimation is based on quantiles.[4] Using the formula for the c.d.f.,

$$F(x) = 1 - x^{-q}, \tag{21}$$

we can equate the sample values of $F(x)$ to their theoretical values. Although this method of estimation is somewhat less efficient[5] than MLE, it is much faster and may be used as a quick method for approximating the parameter when only a rough estimate is needed. In our example the median, or 13th largest loss, is \$55,843 or 2.234 when normalized. Solving $.5 = 1 - 2.234^{-q}$ for q is straightforward yielding $q = 0.826$. If we look at the other two quartiles, which are approximately the 6th and 19th largest losses, we solve the equations

$$.25 = 1 - (1.311)^{-q}$$

$$.75 = 1 - (3.955)^{-q}$$

which yield estimates

$$q = 1.062$$

$$q = 1.008$$

A more important use of this method is when the individual claim sizes are not available (or not easily available), and only grouped statistics are available. Suppose that the losses in Appendix B had been incurred, but the only information was as follows:

Interval (000)	Frequency
25–100	20
100–1,000	4
1,000–∞	1

[4] Quantiles is the general term which includes the median, quartiles, and percentiles as special cases.

[5] For a discussion of efficiency, see Hoel, Port and Stone [H5], *Introduction to Statistical Theory*, page 16.

Using the information that 80% of the losses are no greater than \$100,000, we solve the following:

$$.8 = 1 - 4^{-q}$$

yielding

$$q = 1.161$$

This estimate is remarkably good when one considers the limited information available.

Alternative methods of parameter estimation are discussed in Quandt [Q1].

To this point, in this section we have assumed that there is no upper limitation on the loss data either by an upper censorship point created by policy limits or an upper truncation limit where certain values may be missing.

There are several reasons for suspecting that actual data has some type of limitation. In the case of insurance company data, the losses may be censored due to reinsurance agreements. In some cases, gross losses are available, but in others only net losses may be available in a usable form. Even if the losses are gross to reinsurance, there may be limitations due to policy limits.[6] Most casualty coverages have policy limits.[7]

One of the advantages of working directly on an individual risk is that these limitations can be overcome. Although the primary source for data is usually insurance company records, it is usually possible to make the appropriate adjustments whenever losses have been limited.

This does not totally remove the problems of limitations. In the case of property insurance, there is an upper bound to the amount of loss, namely the total value of the property. There seems to be no useful upper bound to liability situations, but most actual data suggests that the tail of the Pareto is still somewhat too "thick" at extremely high loss amounts. In other words, the theoretical density at high loss amounts is larger than empirical experience tends to indicate. Rather than discard the Pareto, it is easier to postulate that the distribution is censored or truncated at some high, but finite, value. As we have seen earlier, any upper limitation (either censorship point or truncation point)

---

[6] A discussion of the impact of policy limits can be found in Patrik [P1].

[7] Exceptions include workers' compensation coverage A and no-fault PIP in some states.

will produce formulae for the mean claim size that are finite for all possible values of q.

If we assume a censorship point c, then the density function is unchanged between 1 and c but will have a mass point at c and will be zero for all values greater than c. Let $f(x)$ be the unlimited Pareto density, that is

$$f(x) = q\,x^{-(q+1)}$$

Let $f_c(x)$ be the density function censored at c. Then,

$$f_c(x) = f(x) \qquad 1 \le x < c$$

$$f_c(x) = \int_c^{\infty} f(x)dx \qquad x = c$$

$$f_c(x) = 0 \qquad x > c$$

If we wish to consider the distribution truncated (above) at t, then the density function at all points less than or equal to t will have to be proportionately increased so that the total area under the curve still equals one, and the new density function is zero for all values greater than t. Let $f_t(x)$ be the distribution truncated above at t. Then,

$$f_t(x) = \frac{f(x)}{\int_0^t f(x)dx} \qquad 1 \le x \le t$$

$$f_t(x) = 0 \qquad x > t$$

Assume that we have n losses, of which m are less than the censorship limit c and $n - m$ are equal to c. The maximum likelihood estimate is

$$q = \frac{n - m}{\sum^{n-m} \ln X_i + (m)\ln c}$$

Suppose we have the loss data in Appendix B except that each loss is censored at \$100,000. Then,

$$q = \frac{20}{13.104 + 5(1.386)} = .998$$

Note that the MLE approach produces the parameter of the unlimited distribution; censorship is handled through definition of the density function.

V. EFFECT OF TREND

One of the practical problems with fitting size-of-loss distributions is the proper way to handle adjustments for trend and development. With most distributions, inflation of losses will change one or more of the parameters. In Hogg and Klugman, [H2] page 180, there is a table that shows the parameters of various distributions after the application of a trend factor. In each case (including the Pareto and generalized Pareto), the parameters are changed due to inflation.

However, the parameter of the Pareto distribution in this paper is unchanged due to trend. This result appears counterintuitive. After all, each of the formulae for mean claim size is a function of the parameter. If the parameter is unchanged, then the estimated average claim sizes must be unchanged. This appears unreasonable for several reasons.

First, it is obvious that, under influence of trend, the overall average claim size increases. This is true, but note that the distribution in question does not apply to the entire range of losses. It is not simply *better* suited for modeling excess losses, it does not fit small losses well at all. The typical size-of-loss distribution starts out with a small frequency of very small losses, growing to a larger frequency of intermediate losses, then a decreasing frequency of larger losses. The maximum density for the Pareto is always at the leftmost value, and the density is always decreasing as we move to larger claim sizes. Thus, the fact that the overall average claim increases with trend is simply evidence that the single parameter Pareto is not likely to fit the entire range.

Second, it may be recalled that trend is assumed to have a *leveraged* effect on excess losses, where the Pareto is supposed to fit so well. This is true (see Miccolis [M1]), but the leveraged effect is on the total excess *dollars,* not necessarily on the *average* excess claim size. It may seem ironic, but the major effect of trend is to increase the *frequency* of an excess claim, rather than its severity. This may be more obvious if we recall that a size-of-loss distribution is, by definition, the distribution of the relative *frequencies* of various sizes of claims.

Third, and most imporant, a review of empirical excess average claim sizes will show that they have been increasing over time for most coverages. This point is conceded and is inconsistent with an assumption that a Pareto fits the entire excess distribution to infinity. As has been noted earlier, the Pareto has "too thick" a tail, and, in most practical applications, an upper bound should be used. If one looks at the average excess claim *with* a reasonable upper limit, the average claim size will *not* be materially increasing over time.

Because this point is important, we will explore it in more detail. Consider the losses contained in Appendix B. These have been generated from a Pareto distribution with $q = 1.0$. The appendix contains both the normalized values and the raw dollars, which indicate that the raw losses represent losses greater than or equal to $25,000. If we calculate the MLE of these losses (assuming we did not know how they were generated), we would estimate the parameter to be .955. As can be verified, this value will produce average claim sizes for various layers of intervals (with reasonable upper bounds) that are reasonably close to the theoretical values. Specifically, we can use this parameter to estimate the average claim for layers or intervals where the lower limit is $50,000. Thus, this parameter can be thought of as the appropriate parameter for the size-of-loss distribution for claims greater than $50,000.

Suppose these losses were from year zero, and we wished to project losses for year n. Suppose further that the annual trend, $1 + i$, is such that $(1 + i)^n = 2.00$. If we were to trend each of our losses in Appendix B by this trend factor and use these losses to calculate a parameter to fit losses in excess of $50,000, it should be obvious that the estimated parameter would still be exactly .955.

What may be less than obvious is the fact that this parameter can be used for losses in excess of $25,000 in year n. This means that the losses between $25,000 and $50,000 in year n, which correspond to losses less than $25,000 in year 0, must be distributed in such a way that the Pareto distribution will still fit the distribution above $25,000 (to the upper limit) in year n.

As may be guessed, the requirement is that the Pareto distribution must fit the losses in year 0 as low as $12,500 ($25,000/2.00). In general, if we are using losses greater than K from year 0 to estimate a parameter to use in year n, we must assume that the Pareto distribution (with the same parameter) provides a reasonably good fit to losses in year 0 which are as small as $K/(1 + i)^n$. Experience has shown that this is typically true for casualty losses as low as $5,000 to $10,000 (higher for medical malpractice), so values of lower limits in the oldest year of experience greater than $25,000 will typically work. Of course, it is prudent to check the fit at the lower end of the range if possible.

We have gone over this point in some detail because it leads to an extraordinary result: to calculate the MLE of the Pareto parameter, given individual losses greater than a single fixed value K arising from several different years, it is not necessary to adjust the losses for trend.

For example, suppose the following data are available:

1978	100,000,	150,000,	225,000		
1979	109,000,	140,000,	180,000,	240,000	
1980	105,000,	115,000,	170,000,	290,000	
1981	104,000,	121,000,	160,000,	200,000,	300,000

Suppose we are interested in projecting losses for 1984 and the annual trend, $1+i = 1.1$. Under typical methods of analysis, we would trend each of the losses to a common date. The trend factor for 1978 would be $(1.1)^6 = 1.77$. But if we did not have any data on losses less than $100,000 for older years, we would have to use a lower limit of $177,000. Several of the losses in more recent years would then have to be thrown out, because their trended value is less than $177,000.

With the Pareto distribution, we can use all of the data points, *if* we have reason to believe that the Pareto distribution fits losses as low as $100,000/1.77 in 1978. But note that if we assume that the Pareto will fit above $100,000 in 1984, this is equivalent to assuming that it fits equally well above $100,000/1.77 in 1978 (assuming trend affects all claim sizes approximately the same).

Of course, this will allow us to estimate the parameter of the distribution, which will allow us to calculate the *average* severities for 1984. This is only half the problem, as we also need to estimate the frequency of claims to arrive at estimates of the total loss dollars. We cannot simply use the raw historical frequencies of claims greater than $100,000 to estimate our future frequency. We can, however, calculate an adjustment factor that will allow us to put each of the historical frequencies on a comparable basis.

The calculation of this factor can be shown most easily with a concrete example. Suppose we expect 10 claims greater than $25,000 in year n, where $q = 1.5$. Recall the formula for the distribution is $F(x) = 1 - x^{-q}$. How many claims in year n are expected to exceed $1.1 \times \$25,000 = \$27,500$? We calculate this using the distribution function, $F(1.1) = 1 - 1.1^{-1.5} = .133$. This means 13.3% of the claims will be less than $27,500, or 86.7% will be greater than $27,500. Thus, we expect $10 \times .867 = 8.67$ claims greater than $27,500. This means that we expect $1/.867 = 1.153$ claims over $25,000 for every claim greater than $27,500. If we now examine year $n - 1$, the $27,500 claim in year n would be $25,000 in year $n - 1$, and the $25,000 claim would be $\$25,000/1.1 = \$22,727$ in year $n - 1$. Clearly, for every claim greater than $25,000 in year $n - 1$, we would expect 1.153 claims greater than $22,727.

So if we multiply the number of claims greater than \$25,000 in year $n - 1$ by 1.153, we have the best estimate of the number of claims greater than \$22,727 in year $n - 1$, which corresponds to the number of claims that, if trended, would exceed \$25,000 in year n.

Typically, the frequency of claims in each year will be related to an exposure base such as number of beds (hospital malpractice), number of employees (workers' compensation), etc. When using the Pareto distribution, the first step is to multiply the raw frequency of claims greater than the underlying limit by the adjustment factor, then divide through by the exposure. The resulting values may be averaged, or perhaps a regression analysis will be performed. (Note that inflation sensitive exposure bases such as sales or payroll must also be put on a comparable basis.) The adjustment factor for n years of trend at annual inflation rate $1 + i$ with parameter q is simply $(1 + i)^{nq}$. The following table displays the factors for various combinations of i and q. Each value in the table is the one-year adjustment factor.

	q				
$1 + i$	1.00	1.20	1.50	1.80	2.00
1.05	1.050	1.060	1.076	1.092	1.103
1.08	1.080	1.097	1.122	1.149	1.166
1.10	1.100	1.121	1.154	1.187	1.210
1.12	1.120	1.146	1.185	1.226	1.254
1.15	1.150	1.183	1.233	1.286	1.323

VI. SIMULATION OF LOSSES

One type of analysis frequently performed by actuaries involves Monte Carlo simulation of results based upon a particular model of the loss process. One advantage of this Pareto distribution is the ease with which it can be simulated.

One method for simulating values for a function involves inverting the cumulative distribution. This is not always possible with some functions, but it is particularly easy with the Pareto. The cumulative distibution is

$$F(X) = 1 - X^{-q}$$

Thus

$$F^{-1}(Y) = (1 - Y)^{-1/q}$$

where Y has the uniform distribution.

A moment's reflection will reveal that $(1 - Y)$ is symmetric when Y has the uniform distribution, so we can replace $1 - Y$ with Y. Thus, if we can generate a uniform random variable Y, then $Y^{-1/q}$ will have a Pareto distribution with parameter q.

Consequently, we find that even hand-held calculators, such as the HP-15, can be used to simulate Pareto losses. For example, the following values in the first column were generated from a calculator with a random number generator. The second column contains the normalized loss when $q = 1.5$, and the third column contains the "real" claim amount if the lower limit is $25,000.

(1) Random Value from Uniform Distribution	(2) Normalized Pareto Value	(3) "Real" Dollars
.19875	2.93630	73,407
.73616	1.22655	30,664
.52174	1.54298	38,575
.97358	1.01801	25,450
.26635	2.41562	60,390
.54727	1.49462	37,366
.85879	1.10682	27,670
.31708	2.15058	53,764
.38295	1.89630	47,407
.23006	2.66341	66,585

VII. APPLICATIONS

In this section, we will discuss several applications of the Pareto distribution. In some of the cases, we will use actual data from published sources for two reasons: first, to demonstrate that this distribution works well with "real" data, and second, so that this distribution can be compared to those used in the original source of the data.

Application 1

Consider the OL&T BI claims for policy year 1976 contained in Appendix F of Patrik [P1]. We will fit the Pareto to losses greater than $25,000. There are 90 losses in this exhibit. Individual losses are not shown, but the ranges are quite narrow, as they are $5,000 ranges up to $100,000, and $10,000 thereafter. We can use the average claim size in the range as a reasonable proxy for the individual claim amounts (with wider ranges, we might need to make adjustments). The sum of the normalized logs (dividing each claim by $25,000) is 81.2; thus, our estimate of q is 1.108. Note that there are no claims greater than $500,000. We would expect $90\ (1 - F(20)) = 90 \times (20^{-1.108}) = 90 \times .0362 = 3.26$ claims greater than $500,000 if the Pareto fit all the way to infinity. This is evidence that the theoretical tail overstates the actual tail. We can calculate the expected average claim size with an upper limit of $500,000, using (5) with $b = 20$. This yields an estimate of $25,000 × 3.559 = $88,975. The actual average claim size is $89,703.

Application 2

Consider the 40 wind-related catastrophes in 1977 listed in Hogg and Klugman [H2] page 64. Only claims of $2,000,000 or more were included. These values, recorded in millions, are as follows:

2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
2, 2, 3, 3, 3, 3, 4, 4, 4, 5,
5, 5, 5, 6, 6, 6, 6, 8, 8, 9,
15, 17, 22, 23, 24, 24, 25, 27, 32, 43

If we calculate the MLE of the parameter using (19) or (20), the result is $q = .976$. This tends to confirm the statement made earlier that a typical parameter value for property is 1.0. In the same reference, on page 68, are 31 wind catastrophes over $1 million for 1971. The MLE for these losses is $q = .959$.

Application 3

Suppose we have the following hypothetical information for the professional liability experience of a hospital. Assume that the hospital has a $25,000 retention and that information on claims less than the retention is either unavailable or unreliable.

Accident Year	1978	1979	1980	1981
# Occupied Beds	200	200	260	260
Individual Claims Greater than $25,000	127,000	71,000	34,000	55,000
	28,000	119,000	26,000	43,000
	32,000	135,000	38,000	40,000
	103,000	42,000	93,000	42,000
		37,000	40,000	50,000
		55,000	34,000	31,000
			228,000	30,000
			57,000	29,000
			27,000	29,000
			36,000	137,000
				61,000

Suppose we are interested in projecting the experience for 1984 for the layer $225,000 excess of $25,000. Assume that external data leads us to believe that the severity trend has been 20% annually between 1978 and 1981, but is projected to be 15% annually between 1981 and 1984. We also estimate 240 occupied beds in 1984.

First, as noted earlier, we can use all 31 losses in the analysis. Each loss is normalized by dividing by $25,000. The MLE of the parameter is calculated using (19) or (20). The sum of the logs is 22.024, so the estimate of the parameter is 31/22.024 = 1.408.

We can calculate the average claim size in the layer $225,000 xs $25,000 using formula (1)

$$XC(b) = \frac{q - b^{1-q}}{q - 1}$$

With $b = 250/25 = 10$ and $q = 1.408$, the result is 2.493, which corresponds to an average claim size of $25,000 × 2.493 = $62,326. Thus, we expect that the average claim, greater than $25,000 but limited to $250,000, will be $62,326. The amount within the insured layer will be $62,326 − $25,000 = $37,326 per claim.

To estimate the frequency of claim within the layer, we first calculate the frequencies in terms of claims per 100 beds. The resulting ratios are:

Year	# Claims/100 Beds
1978	2.00
1979	3.00
1980	3.85
1981	4.23

We now have to adjust the frequency for trend, so that each year will be on a comparable basis. We will convert each frequency to the frequency that would be expected in 1981, using the adjustment factor in the section on trend, $(1 + i)^{nq}$ where $1 + i = 1.20$, $q = 1.408$, and n is the number of years between each year and 1981. For example, the adjustment factor for 1978 is $(1.20)^{3 \times 1.408} = 2.16$. This means that for every claim that exceeded $25,000 in 1978, we would expect 2.16 claims over $25,000 in 1981. The adjustment factors and the adjusted frequencies are shown in the following:

Year	Raw Frequency	Adjustment Factor	Adjusted Frequency
1978	2.00	2.16	4.32
1979	3.00	1.67	5.01
1980	3.85	1.29	4.98
1981	4.23	1.00	4.23

We can calculate a simple average of the adjusted frequencies to arrive at an estimate of the frequency of claims greater than $25,000 for 1981. This value is 4.63. Alternative methods to calculate an overall frequency could be used. For example, it might be appropriate to use the number of occupied beds as weights. If the adjusted frequencies show a pronounced trend over time, then the frequencies are being affected by something other than changes in claim sizes and further analysis is indicated.

We now calculate the frequency appropriate for 1984. Based upon the assumption of a 15% annual trend, the adjustment factor is $(1.15)^{3 \times 1.408} = 1.805$. Thus, our estimated frequency for 1984 is $1.805 \times 4.63 = 8.36$ claims per 100 occupied beds. Using our assumption that there will be 240 occupied beds in 1984, we expect $8.36 \times 2.4 = 20.07$ claims greater than $25,000 in

1984. Thus, our expected losses in the layer \$225,000 xs \$25,000 are 20.07 × \$37,326 = \$749,301.

Application 4

Finally, we note that the fact that a typical value of q for property losses is 1.0 and the formula for the average loss when $q = 1.0$ is so simple, allows us to easily provide a rough estimate of the average claim size for various layers. Suppose we are asked to quote a reinsurance cover on a book of property business for the layer \$2,750,000 xs \$250,000. The ratio of the upper limit to the lower limit is 3,000/250 = 12, so an estimate of the gross mean claim size is $1 + \ln 12 = 2.485$ or \$621,000. The net mean claim size would be \$371,000. This could be used as a rough estimate for discussion purposes. More refined analysis can be performed if both parties to the intended transaction are still interested.

BIBLIOGRAPHY

[B1] Benktander, G., "On the Correlation in Results from Different Layers in Excess Reinsurance," *Astin Colloquium*, 1961, pp. 203–209.

[G1] Gerber, Hans U., *An Introduction to Mathematical Risk Theory*, Huebner Foundation Monograph Number 8, Richard D. Irwin, Inc., 1979.

[H1] Hewitt, Charles C., Jr. and Lefkowitz, Benjamin, "Methods for Fitting Distributions to Insurance Loss Data," *PCAS*, LXVI, 1979, pp. 139–160.

[H2] Hogg, Robert V. and Klugman, Stuart A., *Loss Distributions*, J. Wiley and Sons, Inc., 1984.

[H3] Haung, J. S., "A Note on Order Statistics from Pareto Distribution," *Scandinavian Actuarial Journal*, 1975, pp. 187–190.

[H4] Hastings, N. A. J. and Peacock, J. B., *Statistical Distributions*, J. Wiley and Sons, Inc., 1974.

[H5] Hoel, Paul G., Port, Sidney C. and Stone, Charles, J., *Introduction to Statistical Theory*, Houghton Mifflin Co., 1971.

[I1] Insurance Services Office, "Report of the Increased Limits Subcommittee: A Review of Increased Limits Ratemaking," 1980.

[J1] Johnson, Norman L. and Kotz, Samuel, *Continuous Univariate Distributions-1*, Houghton Mifflin Co., 1970.

[K1] Kendall, M.G. and Stuart, A., *The Advanced Theory of Statistics, Vol. 2*, (2nd ed.) Mafner Publishing Co., 1967.

[L1] LaRose, J. Gary, "A Note on Loss Distributions," *PCAS*, LXIX, 1982, pp. 15–29, and discussion by Stephen W. Philbrick, *PCAS*, LXXI, 1984, pp. 1–7.

[L2] Lwin, Thaung, "Empirical Bayes Approach to Statistical Estimation in the Paretian Law," *Scandinavian Actuarial Journal*, 1974, pp. 221–236.

[L3] Lau, Hon-Shiang, "An Effective Approach for Estimating the Aggregate Loss of an Insurance Portfolio," *Journal of Risk and Insurance*, LI, No. 1, 1984, pp. 20–30.

[M1] Miccolis, Robert S., "On the Theory of Increased Limits and Excess of Loss Pricing" and discussion by Sheldon Rosenberg, *PCAS*, LXIV, 1977, pp. 27–73.

[M2] Malik, H. J., "Estimation of the Parameters of Pareto Distribution," *Metrika*, Vol. 15, 1970, pp. 126–132.

[M3] Mayerson, A. L., Jones, D. A., and Bowers, N. L., Jr., "On the Credibility of the Pure Premium." *PCAS*, LV, 1968, pp. 175–185.

[P1] Patrik, Gary, "Estimating Casualty Insurance Loss Amount Distributions," *PCAS*, LXVII, 1980, pp. 57–109.

[Q1] Quandt, R. E., "Old and New Methods of Estimation and the Pareto Distribution," *Metrika*, Vol. 10, 1966, pp. 55–82.

[S1] Snader, R. H., "Fundamentals of Individual Risk Rating and Related Topics," Study Note for Parts 9b & 10, Casualty Actuarial Society.

[V1] Venter, Gary, "Scale Adjustments to Excess Expected Losses," *PCAS*, LXIX, 1982, pp. 1–14.

APPENDIX A

SUMMARY OF FORMULAE

This appendix contains a summary of the most commonly used formulae. It begins with the formulae used to calculate the maximum likelihood estimates of the parameter. Formulae are shown later for the mean and total loss variance (under the assumption of a Poisson frequency). The formula for the variance of severity alone is not given, because the primary use for this formula is to derive the formula for the total loss variance.

It should be noted that "K" is used to represent the lower bound of the distribution in nominal or "real" dollars. This is the value used to normalize the distribution. The letter "n" is used in the formula for the MLE to denote the actual number of losses used in the calculation. In the calculation of the expected losses, "n" is used to denote the expected number of claims in the period of interest. The letter "b" is used to denote an upper limit to losses, either a censorship or truncation point.

Density $\quad f(x) = qx^{-1-q}$

Distribution $\quad F(x) = 1 - x^{-q}$

Maximum Likelihood Estimates

Unlimited $\quad q = n/\ln \prod^{n} x_i$

or

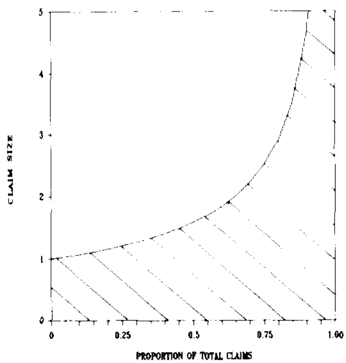
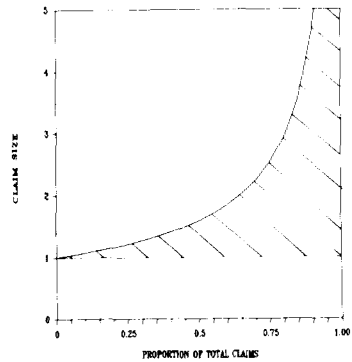
$$q = n/ \sum^{n} \ln x_i$$

Censored at b $\quad q = (n-m)/ \sum^{n-m} \ln x_i + m \ln b$

Truncated at b $\quad q = n/ \sum^{n} \ln x_i - (n \ln b/(b^q - 1))$

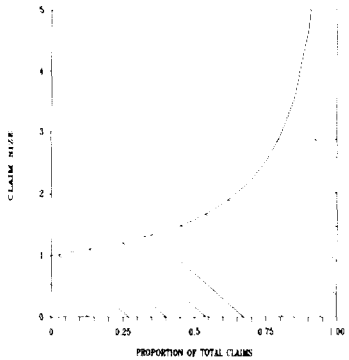
Note that q is on both sides of the equation; thus, it must be solved using numerical methods.

No Upper Limit

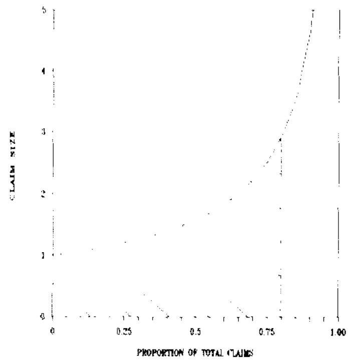
Gross Layer	Net Layer
	

	Gross Layer	Net Layer
Mean Claim Size $q \neq 1$	$\frac{q}{q-1}$	$\frac{1}{q-1}$
"Real" Mean Claim Size	$K\frac{q}{q-1}$	$\frac{K}{q-1}$
"Real" Expected Losses	$nK\frac{q}{q-1}$	$\frac{nK}{q-1}$
Total Loss Variance	$\frac{q}{q-2}$	$\left(\frac{q}{q-2}\right) - \left(\frac{2q}{q-1}\right) + 1$
Total Loss Variance in "Real" Dollars Where Expected Number of Claims is n	$nK^2\left(\frac{q}{q-2}\right)$	$nK^2\left(\frac{q}{q-2}\right) - \left(\frac{2q}{q-1}\right) + 1$

Censored Distribution
(Gross Layer)



Truncated Distribution
(Gross Layer)



	Censored Distribution	Truncated Distribution
Mean Claim Size		
$q \neq 1$	$\dfrac{q - b^{1-q}}{q - 1}$	$\dfrac{q\,(1 - b^{1-q})}{(q - 1)\,(1 - b^{-q})}$
$q = 1$	$1 + \ln b$	$(\ln b)\,/\,(1 - b^{-1})$
"Real" Mean Claim Size		Multiply appropriate formula by K
Expected Losses		Multiply appropriate formula by nK
Total Loss Variance		
$q \neq 2$	$\dfrac{q - 2b^{2-q}}{q - 2}$	$\dfrac{q\,(1 - b^{2-q})}{(q - 2)\,(1 - b^{-q})}$
$q = 2$	$1 + 2 \ln b$	$(2 \ln b)\,/\,(1 - b^{-2})$
Total Loss Variance in Real Dollars		Multiply appropriate formula by nK^2

Net Layer—Mean formula can be calculated by observing that $E[X - 1] = E[X] - 1$. Variance formulae can be calculated by noting that the variance is equivalent to $E[X^2]$, and using the relationship $E[(X - 1)^2] = E[X^2] - 2E[X] + 1$.

APPENDIX B

SIMULATED PARETO LOSSES

25 pseudo-random losses from a Pareto distribution with $q = 1$ and a lower limit of $25,000.

	Amount of Loss	Normalized Amount of Loss
1	69,976	2.799
2	62,913	2.517
3	25,766	1.031
4	39,800	1.592
5	97,739	3.910
6	36,356	1.454
7	139,665	5.587
8	34,749	1.390
9	45,716	1.829
10	96,353	3.854
11	1,847,213	73.889
12	25,231	1.009
13	48,057	1.922
14	31,744	1.270
15	98,882	3.955
16	209,031	8.361
17	214,700	8.588
18	396,323	15.853
19	32,772	1.311
20	45,190	1.808
21	32,044	1.282
22	55,843	2.234
23	99,601	3.984
24	29,900	1.196
25	60,463	2.419

APPENDIX C

VARIOUS FORMS OF THE PARETO

The Pareto distribution is mentioned in a large number of statistical texts and technical papers. Although many distributions (e.g., Poisson and normal) have a fairly standard notation, there is a wide variety of formulations of the Pareto distribution. This appendix will present a brief survey of some of the alternatives.

Johnson and Kotz [J1] contains one of the most thorough treatments of this distribution, as it devotes an entire chapter to the Pareto distribution. This reference includes a discussion of the history of the distribution, which can be traced to the Italian born, Swiss professor of economics, Vilfredo Pareto. Three main representations of the cumulative distribution are given:

Johnson and Kotz $F_x(x) = 1 - \left(\frac{K}{x}\right)^a \qquad K > 0;\ a > 0;\ x \geq K$

Johnson and Kotz $F_x(x) = 1 - \frac{K_1}{(x + c)^a}$

Johnson and Kotz $F_x(x) = 1 - \frac{K_2 e^{-bx}}{(x + c)^a}$

The first is referred to as the "Pareto distribution of the first kind," the second as the "Pareto distribution of the second kind," and the third as the "Pareto distribution of the third kind." Johnson and Kotz note that the first two formulations are Pearson Type VI distributions.

Patrik [P1] uses a form of the Pareto distribution of the second kind:

Patrik $F(x|\beta,\delta) = 1 - \left(\frac{\beta}{x + \beta}\right)^\delta$

Hogg and Klugman [H2] discuss two formulations. The first is referred to as the Pareto distribution and has the cumulative distribution:

Hogg and Klugman $F(x) = 1 - \left(\frac{\lambda}{\lambda + x}\right)^\alpha \qquad \alpha > 0 \quad \lambda > 0$

The second is referred to as the generalized Pareto distribution and has a cumulative distribution as follows (where B (·) refers to the beta distribution):

Hogg and Klugman $F(x) = B\left(K, \alpha: \frac{x}{\lambda + x}\right)$

The density function is as follows:

Hogg and Klugman $f(x) = \frac{\Gamma(\alpha + K)\,\lambda^{\alpha}\; x^{K-1}}{\Gamma(\alpha)\,\Gamma(K)\,(\lambda + x)^{K+\alpha}}$

They note that the Pareto distribution is a special case of the generalized Pareto when $K = 1$.

Formulations by authors who work primarily with the cumulative distribution include:

Huang	$G(X; a, v) = 1 - a^{v} x^{-v}$	$x > a, a > 0, v > 0$
Benktander	$F(x) = 1 - x^{-\alpha}$	$x \geq 1$
Quandt	$F(x) = 1 - \left(\frac{K}{x}\right)^{a}$	$K > 0,\ a > 0,\ x \geq K$

Other authors present this distribution in terms of the density function:

Malik	$f(x) = v\, a^{v} x^{-v-1}$	$a > 0, v > 0, x \geq a$
Lwin	$f(x \mid \lambda, a) = \lambda\, a^{\lambda} x^{-\lambda-1}$	$a > 0, \lambda > 0, x > a$
Kendall and Stuart	$dF = \frac{K}{x^{\alpha}} dx$	$0 < K \leq x \leq \infty, \quad \alpha > 1$
Hastings and Peacock	$f(x) = cx^{-c-1}$	$1 \leq x, \quad c > 0$

Finally, the ISO uses a Pareto distribution in the "Report of the Increased Limits Subcommittee: A Review of Increased Limits Ratemaking" [I1]. In that paper, they use "q" as a parameter. For that reason, "q" has been selected as the parameter in this paper.

APPENDIX D

REASONS FOR PREFERRING "ALTERNATIVE" REPRESENTATION

Although we typically portray density and distribution functions with the loss size along the horizontal axis and the density or cumulative probability along the vertical axis, there are a number of logical reasons for preferring the "alternative" representation, as portrayed in Figures 1, 2, 3, and 4.

1. In the standard representation, a loss limit is a vertical line and the excess losses lie to the right of the line. In my representation, a loss limit would be a horizontal line, and excess losses would lie *above* the line. It seems more intuitive to think of excess losses lying *above* a line.
2. In my representation, losses eliminated by a deductible would be *below* the line representing the deductible amount, rather than to the left of a line.
3. If we apply a trend factor to the cumulative distribution of losses, the new line is *below* the old line in the standard representation but *above* it in my representation. It makes more sense to think of inflation as producing a new curve *above* the old one.

Finally, I would note that this alternative representation is not new. It is essentially equivalent to that used in Snader [S1] to depict the insurance charge and savings.