

Program Manual for FIT, DFIT, CR and Related Programs

Stuart A. Klugman

May 20, 1998

Contents

1	Introduction	1
1.1	Guarantees, Disclaimers, and Help	1
1.2	Installing the programs	2
1.3	Running the programs	2
1.4	Using the menus	3
2	FIT	5
2.1	Data format	5
2.2	Grouped data	5
2.3	Individual data	8
2.4	Distributions	8
2.5	Fitting the model	9
2.6	Results	10
2.7	Plots	11
3	DFIT	13
3.1	Data format	13
3.2	Distributions	14
3.3	Fitting the model	14
3.4	Results	14
3.5	Plots	15
4	CR...	17
4.1	Distributions	17
4.2	Computing and printing the probabilities	19
4.3	FREQADJ	21
4.4	Ruin	21
5	Comprehensive example	23
5.1	Models for the three data sets	24
5.1.1	Model for aggregate small losses	24
5.1.2	Model for the amount of individual large losses	25
5.1.3	Model for the number of large losses per family	26

5.2	A model for aggregate losses	26
5.2.1	Aggregate small losses	26
5.2.2	Aggregate large losses	27
5.2.3	Total aggregate losses	27
6	Miscellaneous	29
6.1	DOS Command	29
6.2	Menu Options	30
6.2.1	FIT and DFIT	30
6.2.2	CR	30
6.2.3	RUIN	31
6.3	Printing	31
A	An Inventory of Continuous Distributions	33
A.1	Introduction	33
A.2	Transformed beta family	36
A.2.1	Four parameter distribution	36
A.2.2	Three parameter distributions	37
A.2.3	Two parameter distributions	39
A.3	Transformed gamma family	42
A.3.1	Three parameter distributions	42
A.3.2	Two parameter distributions	43
A.3.3	One parameter distributions	45
A.4	Other distributions	46
B	An Inventory of Discrete Distributions	49
B.1	Introduction	49
B.2	The (a,b,0) class	50
B.3	The (a,b,1) class	51
B.3.1	The zero-truncated sub-class	51
B.3.2	The zero-modified sub-class	53
B.4	Compound class	53
B.4.1	Some compound distributions	54
B.5	A hierarchy of discrete distributions	55
C	The SimplexMethod	57
D	Frequency and Severity Relationships	59
E	Limited Expected Value Calculations	61
F	The Recursive Formula	63

G	Discretization of the Severity Distribution	65
G.1	The method of rounding	65
G.2	Mean-preserving	66
G.3	Undiscretization of a discretized distribution	66

Chapter 1

Introduction

The programs FIT and DFIT find maximum likelihood estimates for 20 continuous and 16 discrete distributions respectively. They also produce estimates of the covariance matrix of the estimators as well as graphs that illustrate the quality of the fit. The chi-square goodness-of-fit test is also performed. The program CR evaluates probabilities for the collective risk model. Frequencies used can be any of the models used in DFIT while severities can be any of the models from FIT or a model entered by the user. In all cases the probabilities are obtained using recursive formulas. Prior to the recursion, individual or aggregate deductibles, limits, or coinsurances may be applied.

1.1 Guarantees, Disclaimers, and Help

The software and its documentation are provided “as is” and Stuart Klugman disclaims all warranties, either express or implied, including, but not limited to, implied warranties of fitness for a particular purpose. Stuart Klugman does not warrant that the software will meet the user’s requirements or that the operation of the software will be uninterrupted or error free.

Stuart Klugman will not be liable for any indirect, special, or consequential damages in connection with or arising from the performance of or use of this software and its documentation. Stuart Klugman shall in no event be liable for any damages whatsoever (including without limitation, damages for loss of profits or other pecuniary loss) arising out of the use or of the inability to use the software and its documentation, even if Stuart Klugman has been advised of the possibility of such damages.

There are a variety of reasons for the above disclaimer. First of all, no program can be guaranteed to be free of bugs. The best we can hope for is that every time it fails to work properly, I will correct the problem. It is not possible to anticipate every data set which may be used and so every conceivable situation cannot be tested. Second, it is impossible for the program to verify that the data you have entered means what you think it does. That is why help from me is available. Third, I can only assure you that the programs run on my computer. Although there is a great deal of compatibility between systems, yours is not

exactly like mine.

Because I have great faith in the ability of these programs to perform as promised, I am prepared to offer you unlimited assistance in helping them to work for you. This frequently takes the form of you supplying your input file and my either explaining the sequence of steps required to make the program provide the answer, or my informing you how the file as constructed does not match the reality of your data or of your problem, or that a bug has been discovered which will promptly be resolved.

Finally, I trust you, the user, are sufficiently familiar with the underlying statistical principles behind these programs so that you may use them wisely as you conduct your business.

1.2 Installing the programs

The programs are in the files FIT.EXE, DFIT.EXE, CR.EXE, FREQADJ.EXE, CRC.EXE, CRC2.EXE, and RUIN.EXE and should work with any IBM compatible computer. If the 80x87 coprocessor is installed or is built in to your main processor, the program will detect and use it. If the coprocessor is not present, it is emulated, so all users enjoy the additional accuracy provided, but the coprocessor itself is needed to enjoy the extra speed. In order for the graphics to work one of the following files should be in the same directory as the program.

File	Graphics adapter
CGA.BGI	IBM CGA and MCGA
EGAVGA.BGI	IBM EGA and VGA
HERC.BGI	Hercules monochrome
PC3720.BGI	IBM 3270PC

I believe that if you have higher resolution capabilities (such as super-VGA or IBM 8514) the program will read the EGAVGA.BGI file and produce VGA resolution output.

To install the program, copy the program files and the appropriate graphics driver to a directory on your hard disk (or you can run from a floppy, but it will take longer to load).

1.3 Running the programs

Any program can be run from any directory either by typing the complete path to the program or having the program's directory in your path statement. To run, just type the program name, e.g., FIT, DFIT, etc., at the DOS prompt. If all is working well you should see a display in the middle of your screen that reads (depending on the program)

Severity Fit – Version 4.1
by Stuart Klugman
Copyright, June 1996
Do not distribute beyond your immediate
employment area
Press any key to continue

or

Discrete Fit – Version 2.1
by Stuart Klugman
Copyright, June 1996
Do not distribute beyond your immediate
employment area
Press any key to continue

or

Collective Risk – Version 3.0
by Stuart Klugman
Copyright, June 1996
Do not distribute beyond your immediate
employment area
Press any key to continue

The programs CRC and CRC2 do not use windows, the information appears directly on the screen. Both are at Version 2.1, June 1996. The same holds for FREQADJ but it is at version 1.1, June 1996. RUIN does use windows and is at Version 1.2, June 1996. Immediate employment area is to be interpreted as the local office where you work. For example, the local office of your consulting firm, or the cluster of buildings that are the home office of your insurance company. Pressing any key at this point will start the program. The distribution disk contains three data sets: FIT1.DAT and FIT2.DAT contain grouped data and can be used with FIT while FIT3.DAT contains individual data for FIT. DFIT1.DAT and DFIT2.DAT can be used to try out DFIT.

1.4 Using the menus

The neutral state for each program (other than CRC, CRC2, and FREQADJ) has a window in the upper left hand corner. For FIT and DFIT it displays the current model, parameter

values, and negative loglikelihood value. For CR the neutral state window displays the current frequency and severity models, parameter values, and coverage modifications. For RUIN the input parameters are displayed. The message Select an option is at the bottom of the window. Any letter (upper or lower case) entered at this time will cause a command to be executed. The letter M will display the menu options. Command letters can also be entered when the menu is being displayed. The option X terminates the program. A list of the options for each program appears in at the end of the manual.

Chapter 2

FIT

FIT obtains maximum likelihood estimates for severity data that may be grouped or individual (ungrouped). The data may be truncated from below and censored or truncated from above. In all cases the model that is found is for losses starting at zero (ground-up). For each fitted model, values of the distribution function and the limited expected value can be computed.

2.1 Data format

Data may be entered with the program or using any word processor that produces ASCII output with no special characters. The DOS (5.0 or higher) editor (EDIT) is good for this purpose. The MS-Windows editor Notepad will also work. To see what a data file looks like, view any of the sample files using one of these editors, or have them listed on the screen by typing `type fit1.dat` or have one printed out by typing `copy prn fit1.dat`.

Use option C to create a data set. You will be prompted for the name of the file that will hold the data. If the file is to be saved on a directory other than the current one, the complete path must be specified.

2.2 Grouped data

For grouped data, at each subsequent prompt enter the lower limit for the interval, a space, the number of observations in that interval, a space, and the average of the observations in that interval (if you do not have the averages, you can so indicate and the program will insert the midpoints). The averages are used only for computing the empirical limited expected values; they have no influence on the maximum likelihood estimates. The lower limit entered on any line is both the lower limit for that interval and the upper limit for the previous interval. For the purpose of maximum likelihood estimation it does not matter which endpoints are included in an interval, but you should be consistent. The last line

should begin with the upper limit of the ...nal interval and then -10 for the other two entries. If the ...nal interval has in...nity as its upper limit, enter -1 for the boundary.

If values above the ...nal grouping were possible, but just not observed in this sample, an interval with no observations and in...nite upper limit should be included. The data set in FIT1.DAT is an example. To repeat, if the ...nal interval entered has a ...nite upper limit, the program will take this to mean that values above that limit were not possible (truncation from above) as opposed to just did not occur in the sample.

If the data set is truncated from below (that is, there was a deductible), reflect this by simply having the ...rst interval begin with the deductible. Note that the program is expecting the amount of the loss, not the amount paid after the deductible was subtracted.¹

Once again, if the interval from 0 to x had no observations, but observations were possible, that interval must be included. Starting with an interval from x>0 to y indicates that values below x could not be observed.

Example 2.2.1 The following examples indicate some possibilities:

```

0   200   50
100 100   200
300  50   400
500  0    700
i 1  i 10 i 10

```

Observations above 500 were possible, there just were not any observed.

```

0   200   6
12  100   18
24  50    30
36  i 10  i 10

```

Observations above 36 were impossible (truncation from above). This might be data on the time in months from policy issue to claim report. There were no policies issued more than 36 months ago.

```

100 100   150
200  50   250
500  10   750
i 1  i 10  i 10

```

Values below 100 were impossible (truncation from below, as in a deductible). There may have been a policy limit of 500 (censoring from above) with 10 observations at the limit. Had there been a limit, the average of 750 for losses above 500 would not have been observable.

¹The amounts paid after subtracting the deductible could be entered, but then the model found would be for amounts paid, not for ground-up losses.

For maximum likelihood estimation, the values in the third column are not used. Their only purpose is to provide empirical limited expected values for comparison to the model values. Had there been a limit, the value of 750 would be used to calculate the the empirical limited expected value at infinity (the sample mean) to compare to the model mean. With a policy limit this comparison is not valid and so the value can be ignored.

You can also input data that is from various groupings. Terminate each grouping with the entries $xx \ -5 \ -1$ where xx is the upper limit of the last group in that set (use -1 in place of xx for infinity). The data set FIT2.DAT is an example. There is one more feature that can be incorporated into the data set. When there are multiple limits (for example) the data can be easily entered if the original frequency tables were kept separate.

Example 2.2.2 Consider data collected from policies with two limits, 1,000 and 5,000:

0	100	50
100	50	300
500	60	750
1000	30	2000
i 1	i 5	i 1
0	30	50
100	20	300
500	25	750
1000	10	3000
5000	5	13000
i 1	i 10	i 10

There were 240 losses on policies with the 1,000 limit and 90 losses with the 5,000 limit.

Example 2.2.3 Now suppose the two groups had not been kept separate. The data set would then be entered as:

0	130	50
100	70	300
500	85	750
1000	10	3000
5000	5	13000
i 1	i 1	i 1
1000	30	2000
i 1	i 10	i 10

There is now just one group with 320 losses. The -1 terminator in the second entry in line 6 indicates that the group is unchanged, but it contains some overlapping intervals. If there is truncation from above, this must be revealed by having the entry at the ...rst -1 be the truncation point. It should be noted that when there are overlapping intervals, the empirical probabilities and LEV's are adjusted so that they can be comparable.

For option P the current model is utilized to provide the needed interpolation. When graphing, the current model is also used, so if several models are to be placed on one graph you should have what you think is the best model as the current model. The best results are obtained if the maximum possible number of entries are presented prior to the ...rst -1 as was done in the above example.

The maximum number of lines is 500.

2.3 Individual data

For individual data, values can be entered via option C although this may be painful. The data format is as follows (all numbers must be in ASCII code, which is easy to generate with your spreadsheet or word processor).

Line 1 – The number -1, to indicate individual data follows.

Line 2 – 0 if the loss entries are not sorted, 1 if they are increasing.

Line 3 – The number of observations in the data set.

Subsequent lines – The loss (total, not the payment after subtracting the deductible), followed by the deductible (0, if no deductible), followed by the limit (0, if no limit). The three numbers may be separated by spaces or by a comma. If the limit is a censoring limit (observations above the limit are included, but their value is unknown) enter the limit as is. If the limit is a truncating limit (observations above the limit are excluded) enter the negative of the limit. Any entries in the data set that are below the deductible or above a truncating limit will be ignored.

The sample file FIT3.DAT is an example. The maximum number of data points is 3000.

To view the data set use option V. Errors can be corrected with option E. This is not a complete editor, but does allow you to change, delete, or add an entry for grouped data. For individual data any observation can be changed or an observation added. If major work needs to be done it is better to use an ASCII word processor such as DOS-Edit or Windows-Notepad. This can be done without leaving FIT via option O (DOS command, discussed later).

2.4 Distributions

Option S allows you to select the distribution.

2.5 Fitting the model

You must begin by specifying starting values for the parameters. Possible values for some one and two parameter distributions can be found via option B. There is no guarantee that the program will find suitable (or even legitimate) values. If the algorithm (method of moments or percentile matching, depending on the distribution) produces illegal parameter values, no starting values will be provided. Sometimes the parameters are legitimate, but strongly inconsistent with the data, causing the likelihood to be impossible to compute. This will be indicated by a negative loglikelihood of 1E100 (overflow), 1E200 (underflow) or 1E300 (both). In either case, alternative starting values must be created by you, using option N. For three and four parameter models, start a two or three parameter member of the same family that is a special case. Option B will then select the parameters from your earlier start and then set the additional parameter equal to 1. The Appendix can help you identify the special cases.

To enter your own starting values, use option N. You must have already selected a distribution with option S. This option can be used at any time to change the current iterate. If you are desperate for starting values try setting μ equal to the sample mean and all the other parameters equal to 2.

Two methods of doing the minimization (of the negative loglikelihood) are available. The first is the method of scoring. It is a modification of the Newton-Raphson method for finding the roots of an equation. Option I will perform one iteration of this method. The new and old loglikelihood values will be displayed so you can see if the procedure is converging. This method requires good starting values, but when it converges, it does so quickly. Continue iterating until you are satisfied with the results. If the method yields illegal parameter values, the new iterate will not be accepted and you must either get new starting values or switch to the simplex method. The scoring method is only approximate when models whose cumulative distribution function involves the incomplete gamma or beta functions are used. The final iterations must be done by the simplex method. Also, the variances and correlation matrices that are produced are approximate because they are based on approximate derivatives. The distributions for which this is the case are the gamma (5), inverse gamma (6), generalized Pareto (11), transformed gamma (12), inverse transformed gamma (13), and transformed beta (14). For the inverse Gaussian distribution (1), exact derivatives can be written in terms of the standard normal distribution function. Because this must be approximated, the scoring method will converge quickly, but to a slightly incorrect result. As with the previous distributions, it is best to finish with the simplex method. The scoring method is not available for individual data.

The other method is the simplex method. Use option L to initiate it. Follow the instructions. An initial simplex with values increased by 5 percent seems reasonable, but 10 to 20 percent may be more appropriate if your starting values are not good. The better you believe your starting values to be the smaller the percent that can be used. At each iteration the new value is better than at least one of the previous values. If the scoring method is available, once the simplex method appears to have gotten close to the solution, it may be

a good idea to switch.

As a ...nal check, option G will compute the negative loglikelihood at a grid of values surrounding the current parameter value. You specify the percentage adjustment to make to each value and the output places a (*) next to each value that is better than the one at the current point.

You should be aware that maximum likelihood estimates do not exist for all models for some data sets. It may be that the maximum occurs when two or more parameters go to zero and/or in...nity. So it is a good idea to watch the parameter values as you iterate to see if this is happening. This usually means the model is a poor choice anyway or that a model with fewer parameters will do just as well. A word of warning – when the non- μ parameters in the three and four parameter members of the transformed beta family get large (in the hundreds or more) it is likely that you have not converged (and won't) and the program will spend a log time computing under option P as it tries to compute various quantities via nearly in...nite sums.

2.6 Results

Option R lists all models ...tted during the session along with the most recent loglikelihood values. If the chi-square goodness-of-...t test has been done (automatic with option P with grouped data) this command will also present the chi-square test statistic, the degrees of freedom and the p-value. For grouped data option P lists more details about the current model, including variances, a correlation matrix, limited expected values, and a chi-square goodness-of-...t test. For the latter, the individual contribution of each cell to the chi-square test statistic is presented. In the presentation the number is signed to indicate if the expected count was larger (j) or smaller (+). The program groups cells together to ensure that expected counts always exceed 5. Cells which have been grouped have an entry of -1000 displayed as the "contribution" to the test statistic. The number prior to the string of -1000's is the contribution for the grouped cells. For individual data, option P provides only the parameter values, negative loglikelihood, variances, and the correlation matrix. To obtain the other items, use option U to group the data. You can then use the parameters as previously set and option P to display the results for the grouped version. When option U is used, the grouped data are placed in a new ...le. All previous parameters and ...ts are retained, but to save them as connected to the grouped data ...le, option K must be re-run with a new ...le name for the saved results.

As you change from one model to another (via option S) the program remembers the latest parameter values from all previous models. The same output that is given by R plus the parameter values can be saved to a ...le. This ...le can also be read in to allow a ...tting session to be continued at a later time. Use option K to save or restore a session. It is a good idea to save your session often. FIT has very little in the way of error trapping, so if you or the program bomb, all your previous ...ts will be lost. The ...le FIT1.SAV contains the results of ...tting several distributions to the data in FIT1.DAT.

Option F gives $f(x)$, $F(x)$, and $E[X; x]$ at a speci...ed value of x . It also gives the mean

and standard deviation if they exist. This procedure ignores any truncation or censoring points that were in the original data set and computes values as if there was no deductible or limit. This option also allows you to create a table of values for $F(x)$ and $E[X; x]$ provided the entries are arithmetically spaced.

2.7 Plots

Option T creates graphs. This can only be done for grouped data. (Option U can do the grouping for individual data.) The ...ve options are:

1. Histogram and ...tted pdf for the current model.
2. Histogram and ...tted pdf's for the current model and up to three other models.
3. Empirical and ...tted cdf for the current model.
4. Empirical and ...tted LEV for the current model.
5. Empirical and ...tted LEV's for the current model and up to three other models.

You will also be given the opportunity to limit the presentation to a speci...ed range of loss values. This lets you "blow-up" the plot to view its most important features. The lower and upper loss limits must be group boundaries. When asked if you want a two-color plot a "Y" response will produce a plot with white lines on a black background. This is best for notebooks, gray-scale projectors, or if you will be sending the graph to the printer. The quality of the plot will depend on the graphics capabilities of your machine. After viewing, hit any key to continue. Warning - this procedure will not work if the range of loss values contains more than 150 groups.

See the end of the manual for information on printing the graphs.

Chapter 3

DFIT

DFIT finds maximum likelihood estimates for discrete data. All of the models assume that probabilities are associated with non-negative integers. Some of the distributions assume zero is not a possible value.

3.1 Data format

DFIT requires a file containing the sample data. Data may be entered with the program or using any word processor that produces ASCII output with no special characters. Try using your word processor to read the file DFIT1.DAT (or just type type DFIT1.DAT) to see what a data set looks like.

Use option C to create a data set. You will be prompted for the name of the file that will hold the data. If the file is to be on another directory or disk you must give the complete path. At each subsequent prompt, enter the number of times that value occurred in the sample. All of the distributions used in this program assume that the lowest possible value for the random variable is either 0 or 1. If 0 is not a possible value, you still must enter 0 as the number of times 0 was observed. When 0 is impossible, only models in the AB1 class "without zeros" should be used. To indicate that no more data will be entered, enter -10. The program assumes that the last entry prior to the -10 represents the count for that value and all larger values. The only exception is for the binomial distributions where the value of m is input by the user. It is assumed that the number of observations for values larger than the last entry up through m are all zero. If all of the values were recorded as seen, then the last entry prior to the -10 should be zero. This is the case for the two sample sets.

It is also possible for data to be grouped. In that case enter the frequency for the lower limit for the group and enter -1 for the frequency for other values in that group. However, the program requires that the first group contain only 0 (that is, the first group cannot be 0-3, for example). Other groups can be any size.

```

Example 3.1.1    3
                 6
                 i 1
                 i 1
                 7
                 i 1
                 i 1
                 5
                 i 1
                 3
                 i 10

```

The actual observations would be 3 at 0, 6 in the interval 1-3, 7 in the interval 4-6, 5 in the interval 7-8, and 3 in the interval 9-1 .

To view the data set, use option V. Errors can be corrected with option E. This is not a complete editor, but does allow you to change, delete, or add an entry. If major work needs to be done it is better to use a word processor. You can run another program without leaving DFIT via option O.

To use a data set, select it with option D. You will be prompted for the ...le name. The complete path needs to be speci...ed. If you are unsure of the ...le's location, use option O to view your directories.

The data set can have a maximum of 500 entries.

3.2 Distributions

Option S allows you to select the distribution.

3.3 Fitting the model

Fitting is done exactly is with FIT. The scoring method is always available and is a good choice unless the starting values are poor. If useful starting values are hard to obtain, consider setting any μ or σ parameter to the sample mean and all other parameters to 1.

3.4 Results

Option R lists all models ...tted during the session along with the most recent loglikelihood values as well as results from the chi-square goodness-of-...t test for those models that have been examined with option P. Option P lists more details about the current model, including a correlation matrix. As you change from one model to another (via option S) the program remembers the latest parameter values from all previous models. The same output that is given by R along with the parameter values can be saved to a ...le. This ...le can also be

read in to allow a ...tting session to be continued at a later time. Use option K to save or restore a session. It is a good idea to save your session often. DFIT has very little in the way of error trapping, so if you or the program bomb, all your previous ...ts will be lost. DFIT1.SAV contains the results of some ...ts using DFIT1.DAT. Option F will allow you to compute the exact and cumulative probabilities for any value. It also gives the mean and standard deviation for the model.

3.5 Plots

Option T allows you to view the results of the estimation. There are two options. They are

1. Bar charts for the data and the current model.
2. Bar charts for the data, for the current model, and for up to three other models.

Except for the smaller number of options, all else is the same as for FIT.

Chapter 4

CR/FREQADJ/CRC/CRC2/RUIN

The program CR computes probabilities for the model $S = X_1 + \dots + X_N$ where N is a random variable called the frequency and the X_i are identically distributed random variables called the severity. It is assumed that all the random variables are independent. The frequency model can be selected from any of those available in the program DFIT and the same parameterizations are used. The severity model can be any of those used in the program FIT or it can be an arbitrary distribution created by the user. The severity must be a discrete distribution with probabilities at equally spaced values. Because the distributions in FIT are all continuous, the program creates a discrete distribution that is a close approximation. The user can decide what the spacing should be.

The severity distribution may be modified by imposing a deductible, limit, and or co-insurance. The same is true for the aggregate distribution. In addition to the probabilities, the program also produces limited expected values and limited expected squared values for the aggregate distribution.

The program FREQADJ is used to change the parameters of the frequency distribution in order to reflect modifications due to truncation.

The programs CRC and CRC2 are general purpose convolution programs. CRC forms the convolution of two identical distributions while CRC2 forms the convolution of two different distributions.

The program RUIN computes infinite time, discrete process ruin probabilities.

4.1 Distributions

Option F allows you to select the frequency distribution. A menu appears offering the various choices. After your choice is made you will be prompted for the parameter values. The final option is called "Exactly One Claim" and provides a frequency distribution for which the probability of one claim is 1. This makes the output aggregate distribution the same as the (modified) severity distribution.

Option S allows you to select the severity distribution. A menu appears offering the various choices. The last choice is for a user specified distribution. If this one is selected,

you will be prompted for the name of a file holding the values for your distribution or you will be given the opportunity to create a file. The file must be in ASCII form. The first line should have the number of severity values minus 1 (this must be no more than 6500). The second line should have the true probability at zero. This is important if the severity distribution is a discretized version of a continuous or mixed distribution. Subsequent lines each have two numbers. The first is the loss amount and the second is the probability. The loss amounts must be evenly spaced, but need not start at zero. The first of these entries need be zero (that is, probability may start at any value), but if the first entry is zero, it should be at least as large as the probability assigned to be exactly at zero. This is the same format that is required for input to programs CRC, CRC2 and RUIN.

When initially selected, it is assumed that the severity distribution covers losses from 0 to an upper limit (the last value in the discretized version). A deductible, limit (censoring or truncating), and/or coinsurance may be imposed prior to the discretization. When the program does the modifications, the severity variable is always on a per payment basis. There are two kinds of deductible, the ordinary deductible and the franchise deductible. With both deductibles, when the loss x is below the deductible d , nothing is paid. However, when the loss is above d the ordinary deductible pays $x - d$ while the franchise deductible pays the full amount x . The franchise deductible is also useful when a fitted model is being used for severities above some amount and empirical data below.

With regard to the policy limit, it should be noted that the amount entered is the limit prior to the imposition of the deductible or coinsurance. The entered limit is not the maximum amount payable per claim. For example, with a deductible of 100, a coinsurance of .8, and a limit of 1,000, a loss of 700 results in a payment of $.8(700 - 100) = 480$ while a loss of 1,100 results in a payment of $.8(1,000 - 100) = 720$. The maximum payment is 720.

Once the distributions have been selected, the severity probabilities may be viewed by using option V. The values can be placed on the screen, printer, or be written to a file.

There are two methods available for performing the discretization. The method of rounding preserves total probability while the mean-preserving method ensures that the discretized distribution has the same mean as the original continuous model.

Example 4.1.1 The severity distribution is Pareto ($\alpha = 3, \mu = 10$) and the frequency for ground-up claims is zero-modified geometric ($p_0 = .6, \bar{p} = 2$). A discretization is desired with a span (interval) of 2. The first six discretized values, by the method of rounding are:

$$\begin{aligned} f_0 & \Pr(X \leq 1) = .248685 \\ f_1 & \Pr(1 < X \leq 3) = .246149 \\ f_2 & \Pr(3 < X \leq 5) = .158870 \\ f_3 & \Pr(5 < X \leq 7) = .092755 \\ f_4 & \Pr(7 < X \leq 9) = .057748 \\ f_5 & \Pr(9 < X \leq 11) = .037814 \\ & \vdots \end{aligned}$$

Now suppose an ordinary deductible of 4 is applied prior to the discretization. This gives the probability of a particular amount being paid given that a payment is to be made. The results are

$$\begin{aligned} g_0 & \Pr(4 < X \leq 5) = \Pr(X > 4) = :186963 \\ g_1 & \Pr(5 < X \leq 7) = \Pr(X > 4) = :254519 \\ g_2 & \Pr(7 < X \leq 9) = \Pr(X > 4) = :158460 \\ g_3 & \Pr(9 < X \leq 11) = \Pr(X > 4) = :103762 \\ & \vdots \end{aligned}$$

If a limit of 10 (6 above the deductible) and coinsurance of .75 are to be applied the only change is to concentrate all of the excess probability at g_3 . That is, $g_3 = :400058$.

For the last two cases, the frequency distribution must be modified to reflect the per payment basis. Using FREQADJ, the new zero-modified geometric parameters are $p_0 = :747049$, $\tau = :728863$.

4.2 Computing and printing the probabilities

Aggregate modifications can be imposed by selecting option A. This will not affect the recursions as the modifications are imposed only when option P is selected for viewing the results. Recursions will be carried out to the earliest of 3000 values, the rounding of all but about one hundred-millionth of the probability, or a value preset when option C was invoked. So the aggregate modifications can be changed and their effects noted without redoing the recursions (option C). However, if you know that there will be an aggregate limit that will not be raised, you can save time by only doing recursions to just a little bit past this limit.

Option C asks the program to compute the probabilities for the collective risk model. You will be prompted for the starting value for recursions (see the discussion below on modifying the recursions). If you enter zero, ordinary recursions giving true probabilities will be done. You can also enter the aggregate amount at which aggregate loss calculations are to stop. If you enter -1 the program will terminate on its own, when either .9999999 of the probability has been found or 6500 values have been set. You should always select -1 when recursions start at a value above zero.

Output is always computed at the same spacing as used for the severity distribution. This can cause some problems when the expected frequency is high. For example, a spacing of 5 may work well for a severity distribution, creating 500 intervals (so nearly all the probability is assumed to be below 2,500). If the expected frequency is 100 claims, this would require perhaps 25,000 intervals of width 5 for the aggregate distribution, well beyond the memory of the program. To solve this problem there are two possible approaches.

Begin the `...rst` method by running CR with the expected frequency divided by a power of 2 (2, 4, 8, etc.) to make it reasonable. For the Poisson or any Poisson-x distribution, divide the Poisson parameter λ by this number. For the negative binomial, or any negative binomial-x distributions, divide the value of r by this number (recall that the geometric

distribution is just another negative binomial distribution, with $r = 1$). For the binomial distribution, divide m by this number (the result must be a whole number). Obtain the aggregate distribution and when complete, write the results to disk using the option that allows for output suitable for input to the convolution program. Then run CRC using this file as input. The output will be the convolution of the aggregate distribution with itself. The output can be written to disk in a format suitable for re-input to CRC or in a form that matches the output from CR. This process must be repeated as many times as the power of 2 that was used (for example, if division was by 8, run CRC 3 times).

The program CRC2 is a general purpose convolution program that produces the convolution of two different discretized distributions. An example would be two separate aggregate distributions for two lines of business, each produced by program CR. Input is in the same format as for CRC.

The second method is to start the recursive calculations at a value higher than zero. This would be appropriate if the high frequency has pushed the probabilities well above zero. A reasonable place to start is four to six standard deviations below the mean. The program temporarily assigns a probability of one to this value and then proceeds recursively. Recursions continue until the allotted 6,500 spaces are used or the probabilities become relatively small. Once the recursions are completed the probabilities are rescaled to add to one. The major drawback to this approach is that you cannot be sure that all of the probability has been accounted for when the recursions stop (when recursions start at zero the true probabilities are obtained and so the remaining probability can be placed at the last value). The program will provide a warning when it appears that there is a problem.

Option P will print the cumulative probabilities, the limited expected values, and the limited expected squared values. You have the option of printing the results on the screen, on the printer, or to a file. As noted above, you can also write the output to disk in a form suitable for input into CRC, CRC2, or RUIN. This form is also suitable for input as a user specified severity distribution for CR itself. Finally, you also have the option of printing a table of percentiles.

Example 4.2.1 Use program CR to obtain the aggregate distribution of losses from the model described in the previous example. Impose the individual ordinary deductible of 4, censoring limit of 10 and coinsurance of .75. The output shown below is from option P with output written to disk and then imported to this word processor.

```
Deductible applied prior to discretization:
Frequency: Truncated Geometric with zero
Beta = 7.2886300E-0001
p0 = 7.4704900E-0001
Severity: Pareto
Alpha = 3.0000000E+0000
Theta = 1.0000000E+0001
Discretization interval = 1.5000000E+0000
Discretization method: Rounding
```

Individual ordinary deductible: 4.0000000E+0000

Individual censoring limit: 1.0000000E+0001

Individual coinsurance: 0.750000

No aggregate deductible

No aggregate limit

Aggregate coinsurance: 1.000000

AggLoss	PDF	CDF	LEV	LEV2
0.00	0.498032667	0.747049000	0.0000	0.0000
5.00	0.112863233	0.916343849	0.8694	3.7214
10.00	0.042364652	0.979890827	1.0978	6.9186
15.00	0.009977457	0.994857012	1.1548	8.2800
20.00	0.002586018	0.998736040	1.1688	8.7539
25.00	0.000636661	0.999691031	1.1722	8.9035
30.00	0.000155213	0.999923851	1.1730	8.9483
35.00	0.000038508	0.999981614	1.1732	8.9612
40.00	0.000009282	0.999995536	1.1733	8.9648
45.00	0.000002244	0.999998902	1.1733	8.9658
50.00	0.000000555	0.999999735	1.1733	8.9661
55.00	0.000000134	0.999999936	1.1733	8.9662
60.00	0.000000032	0.999999984	1.1733	8.9662
65.00	0.000000011	1.000000000	1.1733	8.9662
70.00	0.000000000	1.000000000	1.1733	8.9662

Mean = 1.1733

Std. Dev. = 2.7549

4.3 FREQADJ

This program is a small utility that can be used to adjust the frequency distribution. This usually happens when the original frequency distribution is based on data from one deductible and/or truncating limit and the discretized distribution is appropriate for a different deductible/limit combination. The program asks for the current frequency distribution, the current severity distribution, the current deductible/limit combination and the new deductible/limit combination. The output is simply the new parameters for the frequency distribution. If needed in the middle of running program CR, option O can be used to run this program and then return to CR.

4.4 Ruin

This program uses convolutions to compute ...nite time ruin probabilities for a discrete time model. The input is the initial surplus, the annual premium (assumed paid at the beginning of the year), the interest rate, and the ...le containing the aggregate loss distribution. The

latter should be a file produced by CR, CRC, or CRC2 using the output for convolutions option. It is assumed that all claims are paid at the end of the year. The program will then generate the distribution of surplus after a given number of years (option R allows you to select the number of additional years to run the model). The output (view with option P) is the probability of being ruined within that many years as well as the mean and standard deviation of surplus, given that ruin has not occurred.

Chapter 5

Comprehensive example

The format of this example is probably unlike any practical setting. It has been selected in order to illustrate many of the capabilities of these programs. The numbers themselves are made-up and are not an attempt to mimic any real data.

The problem is to learn about aggregate losses on family dental coverage. Three data sets are available. The first represents losses on small, individual claims. For each family the annual aggregate total of losses below 250 was recorded. That is, individual losses above 250 were excluded from the total.

Amount of loss	Number of losses
0-250	71
250-350	327
350-400	167
400-450	123
450-500	97
500-600	128
600-700	103
700-800	67
800-1,000	68
1,000-1,500	25
1500-	1

Individual losses above 250 were recorded individually. There were 384 such losses which were grouped as follows

Amount of loss	Number of losses
250–500	251
500–750	80
750–1,000	25
1,000–1,500	15
1,500–2,000	8
2,000–2,500	3
2,500–	3

The final piece of information is data on the number of large (excess of 250) claims by one family in one year:

Number of losses	Number of families
0	95
1	130
2	100
3	38
4	16
5	3
6	1
7	1

The three data files that correspond to this information are available as MAN1.DAT, MAN2.DAT, and MAN3.DAT respectively.

5.1 Models for the three data sets

5.1.1 Model for aggregate small losses

The first step in the fitting process is to use FIT to obtain a model for the amount of aggregate small losses in one year. After fitting all possible models (a few did not converge), the best models with each number of parameters were selected (the four parameter model did not converge). In each case the model with smallest negative loglikelihood was also the model with the smallest \hat{A}^2 -value for the goodness-of-fit test.

No. param.	Model	-loglikelihood	chi-square	df	p-value
1	Exponential	3100.94	1173.14	9	near 0
2	Inverse gamma	2506.30	32.98	7	near 0
3	Inverse trans. gamma	2505.32	31.99	7	near 0

The likelihood ratio test indicates that the inverse gamma model is an improvement over the exponential (twice the difference of the negative loglikelihoods is greater than 3.84, the 5% critical value with one degree of freedom), while the inverse transformed gamma model is not an improvement over the inverse gamma. Figure 5.1.1 shows that the inverse gamma model is satisfactory, even though it does not pass the goodness-of-fit test.

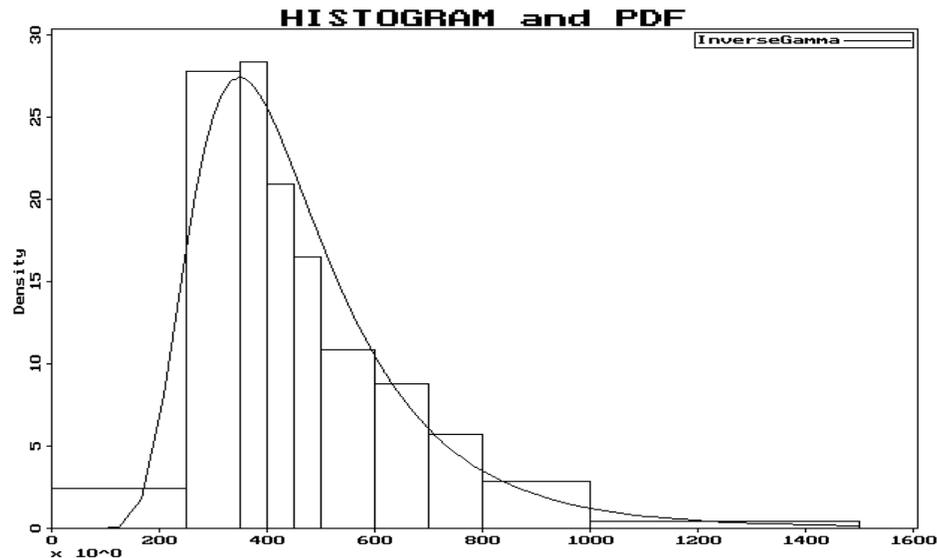


Figure 5.1.1: Inverse gamma distribution

The parameter estimates are $\hat{\theta} = 6.73021$ and $\hat{\mu} = 2692.53$. Further information about this model, including standard deviations of the estimators can be found with option P.

5.1.2 Model for the amount of individual large losses

Using program FIT, a number of models were successfully fitted to the data. The best fitting models with each number of parameters are summarized below.

No. param.	Model	-loglikelihood	chi-square	df	p-value
1	Exponential	427.08	9.82	2	.0074
2	Inverse Weibull	411.15	1.75	3	.6270
3	Burr	411.01	1.37	2	.5032

Identical reasoning to that from the previous subsection indicate that the inverse Weibull is the distribution of choice. This one also passes the goodness-of-fit test and had the highest p-value of any model. The quality of the fit is evident from Figure 5.1.2

The parameters for the selected inverse Weibull model are $\hat{\lambda} = 2.38148$ and $\hat{\mu} = 317.676$.

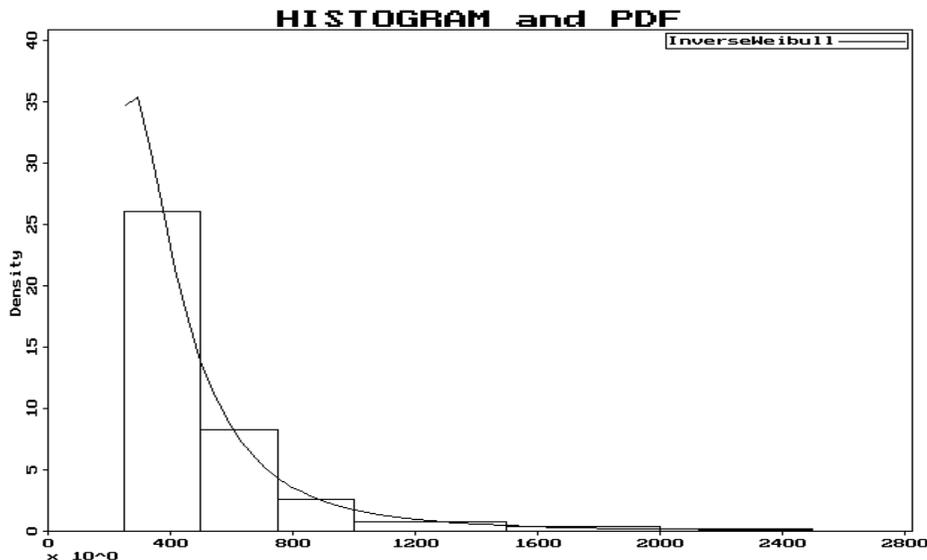


Figure 5.1.2: Inverse Weibull distribution

5.1.3 Model for the number of large losses per family

The data were entered into DFIT. Of the models which converged, the best were

No. param.	Model	-loglikelihood	chi-square	df	p-value
1	Poisson	574.97	1.34	4	.8549
2	Poisson-Inv. G.	574.96	1.40	3	.7066
2	Polya-Aeppli	574.96	1.39	3	.7077
3	Poisson-ETNB	574.88	1.53	2	.4646

Two two-parameter models are presented as one had the best loglikelihood number (Poisson-inverse Gaussian, in the third decimal place) and the other had the best chi-square value. However, the likelihood ratio test indicates that the Poisson model cannot be rejected in favor of one with more parameters and as well, it has the highest p-value for the goodness-of-fit test.

The parameter is $\lambda = 1:39583$.

5.2 A model for aggregate losses

5.2.1 Aggregate small losses

These have already been modeled with the inverse gamma distribution. To use this distribution we need only discretize it. This can be done with program CR, entering the inverse

gamma as the severity distribution, discretizing with no modifications (a span of 5 was used with the method that matches the first moment), selecting a frequency of exactly one claim, hitting C to make the transfer, and then using option P to write the distribution to disk in a format that can be used for later input.

5.2.2 Aggregate large losses

The severity for individual large losses is inverse Weibull. However, this is a ground-up distribution and we want to include only those losses in excess of 250. This can be done with a franchise deductible of 250. Because the model for the number of claims was based on those in excess of 250 there is no need to adjust the frequency distribution. The results are then saved to disk in a format that can be used for later input. A span of 25 was used with the method that matches the first moment.

5.2.3 Total aggregate losses

This is just the convolution of the two distributions found previously. This is done with program CRC2. The results are then saved to disk and then re-read into program CR with a frequency distribution of "exactly one claim." Then output items of interest can be obtained. For this problem the following moments and percentiles were requested:

Mean	1,230
Std.Dev.	969
50th pctile	1,032
75th pctile	1,574
90th pctile	2,241
95th pctile	2,765
99th pctile	4,269

Modifications can then be entered as aggregate, not individual, modifications. This was done here to evaluate the layer from 2,500 to 4,000. The mean turns out to be 48 and the standard deviation 224.

Chapter 6

Miscellaneous

6.1 DOS Command

Option O lets you execute a DOS command without leaving the program. After selecting this option you can either enter the DOS command (e.g. "DIR") and then be returned to the program as soon as execution is completed, or hit -return- and be placed back in DOS (for example, to do word processing) and then be returned to the program only when you type EXIT in response to a DOS prompt (>).

6.2 Menu Options

6.2.1 FIT and DFIT

Option	Action
B	Starting values
C	Create a new data set
D	Read in a data set
E	Edit a data set
F	Functional values from the current model
G	Likelihood values from a grid around the current values
I	One iteration of the scoring method
K	Save or restore the results of a ...tting session
L	Simplex minimization
M	Display this menu
N	New starting values
O	DOS command
P	Print results for the current model
R	Summarize results for all models used
S	Select a model
T	Plot PDF, CDF, or LEV
U	Group individual data (FIT only)
V	View the data
X	Terminate the program

6.2.2 CR

Option	Action
A	Input aggregate modi...cations
C	Compute aggregate probabilities
F	Select the frequency distribution
M	View this menu
O	DOS command
P	Print results
S	Select the severity distribution
V	View the severity probabilities
X	Terminate the program

6.2.3 RUIN

Option	Action
A	Get aggregate distribution
M	View this menu
O	DOS interface
P	Print results
R	Compute ruin probabilities
V	Select initial values
X	Terminate the program

6.3 Printing

There are four ways to print the graphs.

(a) The DOS 5.0 and above graphics command will work with various printers. If you have one, run the graphics command prior to running FIT or DFIT. Hitting the PrintScreen key (or perhaps Shift or Alt-PrintScreen) will cause the graph to be printed.

(b) There are a number of commercial screen capture routines. These can be used to write your graph to disk in some format that can be read by your word processor. You can then use your word processor to edit and print the graph.

(c) The distribution disk contains two programs, LJPS.EXE and LJPSRUN.EXE. Both are designed to enable you to print the contents of a graphics screen on an HP Laser Jet or Desk Jet printer. They will not work with a Hercules graphics card, nor will they work with Super-VGA. They will work with CGA, EGA, and regular VGA cards (Even if you have Super-VGA, I believe FIT and DFIT only use the resolution of regular VGA, so these programs will probably work anyway.) When running, all you need to do is hit PrintScreen (or Shift or Alt-PrintScreen). These programs will work with any screen, not just those produced by FIT and DFIT.

The program LJPS is a memory resident program. Run it by typing LJPS at the DOS prompt. Once run, it will always be the program by which screens dumps are handled. There is no way to unload this program.

LJPSRUN is not memory resident, but instead invokes a secondary DOS session. Run it by typing LJPSRUN at the DOS prompt. Later on you may type EXIT at the DOS prompt to cancel the program and the secondary DOS session. A more useful way to use this program is to type LJPSRUN FIT or LJPSRUN DFIT at the DOS prompt. This will create the secondary session, run the print screen handler, and then run the requested program. When the requested program is terminated, the secondary session will be ended and the print screen handler unloaded.

Both programs assume your printer is connected to LPT1, the ...rst parallel printer port. If your printer is connected to LPT2 or LPT3, type LJPS -p2 or LJPS -p3 when running the ...rst program or LJPSRUN -p2 or LJPSRUN -p3 when running the second one. You may still append a program name afterward when running LJPSRUN.

(d) If you are using Windows 3.0 or higher in enhanced mode, the following is available. First, run (D)FIT from within Windows. The default PIF works just ...ne. So just call the program using the RUN command from the FILE menu in either program manager or ...le manager. Create the desired graph (for best results, use the "two colors" option when making the graph and then hit Alt-PrtScr to copy the screen to the Windows clipboard. Then exit (D)FIT or at least switch to the Paintbrush program. Be sure the window is maximized. Choose Image Attributes from the Options menu and change the size of the picture to the resolution of your screen, using pels as the units. For example, with EGA use 640 and 350. From the View menu choose Zoom Out. From the Edit menu choose Paste. Cross hatches should appear. Place the cross-hair cursor in the upper left corner of the drawing area and click the left mouse button. The picture should now appear. Choose Zoom In from the View menu.

The colors must now be changed. The general routine is to select the old color on the color palette and click the left mouse button. Select the new color on the color palette and click the right mouse button. Then double click the color erase tool (the eraser on the left) to change every occurrence of the old color to the new color. Because the graph takes up more room than can be displayed, you will have to click the horizontal scroll bar to reveal the rest of the picture and double click the eraser tool to change the rest of it. Keep changing colors until you have black lines on a white background (unless, of course, you have a color printer, in which case you would have been printing already).

At this point you can print the graph directly from Paintbrush, or save it in PCX or BMP format for later importation to any program that recognizes these formats. Or, you can copy the graph back to the clipboard for pasting into other Windows applications (like a word processor) by ...rst choosing Zoom Out from the View menu, selecting the pick tool (the scissors with a rectangle attached), placing the cross-hair cursor on the upper left corner of the picture, pressing the left mouse button, dragging the cursor to the lower right corner to outline the entire graph, releasing the left mouse button, and then choosing Copy from the Edit menu.

Appendix A

An Inventory of Continuous Distributions

A.1 Introduction

Descriptions of the models are given below. First a few mathematical preliminaries are presented that indicate how the various quantities can be computed.

The incomplete gamma function¹ is given by

$$\Gamma(\alpha; x) = \frac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-t} dt; \quad \alpha > 0; \quad x > 0$$

$$\text{with } \Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt; \quad \alpha > 0:$$

When $\alpha \leq 0$ the integral does not exist. In that case, define

$$\Gamma(\alpha)G(\alpha; x) = \int_x^{\infty} t^{\alpha-1} e^{-t} dt; \quad x > 0:$$

Integration by parts produces the relationship

$$\Gamma(\alpha)G(\alpha; x) = \frac{x^{\alpha} e^{-x}}{\alpha} + \frac{\Gamma(\alpha + 1)}{\alpha} G(\alpha + 1; x)$$

which allows for recursive calculation because for $\alpha > 0$, $\Gamma(\alpha)G(\alpha; x) = \Gamma(\alpha)[1 - \Gamma(\alpha; x)]$. When α is zero or a negative integer, this recursive formula requires the evaluation of

$$\Gamma(0)G(0; x) = \int_x^{\infty} t^{-1} e^{-t} dt = E_1(x)$$

¹Some references, such as [1], denote this integral $P(\alpha; x)$ and define $\Gamma(\alpha; x) = \int_x^{\infty} t^{\alpha-1} e^{-t} dt$. Note that this definition does not normalize by dividing by $\Gamma(\alpha)$. When using software to evaluate the incomplete gamma function, be sure to note how it is defined.

where r is the smallest integer such that $b + r + 1 > 0$. The ...rst argument must be positive (that is, $a + r + 1 > 0$).

Numerical approximations for both the incomplete gamma and the incomplete beta function are available in many statistical computing packages as well as in many spreadsheets as they are just the distribution functions of the gamma and beta distributions. The following approximations are taken from [1]. The suggestion regarding using different formulas for small and large x when evaluating the incomplete gamma function is from [3]. That reference also contains computer subroutines for evaluating these expressions. In particular, it provides an effective way of evaluating continued fractions.

For $x < \alpha + 1$ use the following series expansion

$$\Gamma(\alpha; x) = \frac{x^\alpha e^{-x}}{\Gamma(\alpha)} \sum_{n=0}^{\infty} \frac{x^n}{\Gamma(\alpha + 1) \Gamma(\alpha + n)}$$

while for $x > \alpha + 1$ use the following continued fraction expansion

$$\Gamma(\alpha; x) = \frac{x^\alpha e^{-x}}{\Gamma(\alpha)} \cfrac{1}{x + \cfrac{1}{1 + \cfrac{1}{x + \cfrac{2}{1 + \cfrac{2}{x + \cfrac{3}{\ddots}}}}}}$$

The incomplete gamma function can also be used to produce cumulative probabilities from the standard normal distribution. Let $\Phi(z) = \Pr(Z \leq z)$ where Z has the standard normal distribution. Then for $z \geq 0$, $\Phi(z) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^z e^{-t^2/2} dt$ while for $z < 0$, $\Phi(z) = 1 - \Phi(-z)$.

The incomplete beta function can be evaluated by the following series expansion

$$I_x(a; b) = \frac{x^a (1-x)^b}{B(a, b)} \sum_{n=0}^{\infty} \frac{x^n (1-x)^n}{(a+n) \Gamma(a+n) \Gamma(b-n)}$$

The gamma function itself can be found from

$$\log \Gamma(\alpha) \doteq (\alpha - \frac{1}{2}) \log \alpha - \alpha + \log(\sqrt{2\pi}) + \frac{1}{12\alpha} - \frac{1}{360\alpha^3} + \frac{1}{1260\alpha^5} - \frac{1}{1680\alpha^7} + \frac{1}{1188\alpha^9} - \frac{691}{360 \cdot 360\alpha^{11}} + \frac{1}{156\alpha^{13}} - \frac{3617}{122 \cdot 400\alpha^{15}} + \frac{43 \cdot 867}{244 \cdot 188\alpha^{17}} - \frac{174 \cdot 611}{125 \cdot 400\alpha^{19}}$$

For values of α above 10 the error is less than 10^{-19} . For values below 10 use the relationship

$$\log \Gamma(\alpha) = \log \Gamma(\alpha + 1) - \log \alpha$$

The distributions are presented in the following way. First the name is given along with the parameters. Next the density function ($f(x)$) and distribution function ($F(x)$) are given. These are followed by the k th moment, the k th limited expected value (when there is simplification, the formula is given for integer k) and the mode. These are followed by the derivatives of the distribution function with respect to the parameters. For distributions

involving the incomplete gamma and beta functions some of these derivatives are difficult to compute and so an approximation should be used. Finally, for some distributions formulas for starting values are given. Within each family the distributions are presented in decreasing order with regard to the number of parameters. The Greek letters used are selected to be consistent. Any Greek letter that is not used in the distribution means that that distribution is a special case of one with more parameters, but with the missing parameters set equal to 1. Unless specifically indicated, all parameters must be positive.

Except for two distributions, inflation can be recognized by simply inflating the parameter μ . That is, if X has a particular distribution, then cX has the same distribution type, with all parameters unchanged except μ is changed to $c\mu$. For the lognormal distribution, σ^2 changes to $\sigma^2 + \log(c)$ with μ unchanged while for the inverse Gaussian both λ and μ are multiplied by c .

For several of the distributions, starting values are suggested. They are not necessarily good estimators, just places from which to start an iterative procedure to maximize the likelihood or other objective function. These are found by either the methods of moments or percentile matching. The quantities used are:

$$\text{Moments: } m = \frac{1}{n} \sum_{i=1}^n x_i \quad t = \frac{1}{n} \sum_{i=1}^n x_i^2$$

Percentile matching: $p = 25\text{th percentile}$, $q = 75\text{th percentile}$

For grouped data or data that have been truncated or censored, these quantities may have to be approximated. Because the purpose is to obtain starting values and not a useful estimate, it is often sufficient to just ignore modifications. For three and four parameter distributions starting values can be obtained by using estimates from a special case, then making the new parameters equal to 1. An all-purpose starting value rule (for when all else fails) is to set the scale parameter (μ) equal to the mean and all other parameters equal to 2.

A.2 Transformed beta family

A.2.1 Four parameter distribution

A.2.1.1 Transformed beta - $(\theta; \mu; \alpha; \beta)$

$$f(x) = \frac{\beta (\alpha + \beta) \mu^{\alpha+\beta} (x-\mu)^{\alpha-1}}{\Gamma(\alpha) \Gamma(\beta) \mu^{\alpha+\beta} [1 + (x-\mu)^{\alpha}]^{\alpha+\beta}}$$

$$F(x) = I_{\left(\frac{x-\mu}{\mu}\right)^{\alpha}}(\alpha; \beta); \quad u = \frac{(x-\mu)^{\alpha}}{1 + (x-\mu)^{\alpha}}$$

$$E[X^k] = \frac{\mu^k \Gamma(\alpha + k) \Gamma(\beta)}{\Gamma(\alpha) \Gamma(\beta)}; \quad \alpha < k < \alpha + \beta$$

$$E[(X \wedge x)^k] = \frac{\mu^k i (\lambda + k) i^{(\alpha)} i^{k-1}}{i^{(\alpha)} i (\lambda)} - (\lambda + k; \alpha; i^{k-1}; u) + x^k [1 - F(x)]; \quad k > i \lambda$$

$$\text{mode} = \mu \frac{\lambda - i + 1}{\alpha + 1}; \quad \lambda > 1; \text{ else } 0$$

$$\frac{\partial F}{\partial \alpha} = - {}_2(\lambda; \alpha; u)$$

$$\frac{\partial F}{\partial \mu} = i f(x) \frac{x}{\mu}$$

$$\frac{\partial F}{\partial \lambda} = f(x) \frac{x}{\mu} \log \frac{\mu x}{\mu}$$

$$\frac{\partial F}{\partial i} = - {}_1(\lambda; \alpha; u)$$

A.2.2 Three parameter distributions

A.2.2.1 Generalized Pareto - $\alpha; \mu; \lambda$

$$f(x) = \frac{i (\alpha + \lambda) \mu^\alpha x^{\lambda-1}}{i^{(\alpha)} i (\lambda) (x + \mu)^{\alpha+\lambda}}$$

$$F(x) = - (\lambda; \alpha; u); \quad u = \frac{x}{x + \mu}$$

$$E[X^k] = \frac{\mu^k i (\lambda + k) i^{(\alpha)} i^{k-1}}{i^{(\alpha)} i (\lambda)}; \quad i \lambda < k < \alpha$$

$$E[X^k] = \frac{\mu^k \lambda (\lambda + 1) \dots (\lambda + k - 1)}{(\alpha - i + 1) \dots (\alpha - i + k)}; \text{ if } k \text{ is an integer}$$

$$E[(X \wedge x)^k] = \frac{\mu^k i (\lambda + k) i^{(\alpha)} i^{k-1}}{i^{(\alpha)} i (\lambda)} - (\lambda + k; \alpha; i^{k-1}; u) + x^k [1 - F(x)]; \quad k > i \lambda$$

$$\text{mode} = \mu \frac{\lambda - i + 1}{\alpha + 1}; \quad \lambda > 1, \text{ else } 0$$

$$\frac{\partial F}{\partial \alpha} = - {}_2(\lambda; \alpha; u)$$

$$\frac{\partial F}{\partial \mu} = i f(x) \frac{x}{\mu}$$

$$\frac{\partial F}{\partial \lambda} = - {}_1(\lambda; \alpha; u)$$

A.2.2.2 Burr - $\mathbb{R}; \mu; \circ$

$$f(x) = \frac{\mathbb{R}^\circ (x=\mu)^\circ}{x[1 + (x=\mu)^\circ]^{\mathbb{R}+1}}$$

$$F(x) = 1 - u^\mathbb{R}; \quad u = \frac{1}{1 + (x=\mu)^\circ}$$

$$E[X^k] = \frac{\mu^k i (1 + k=\circ)_i (\mathbb{R} - i - k=\circ)}{i (\mathbb{R})}; \quad i^\circ < k < \mathbb{R}^\circ$$

$$E[(X \wedge x)^k] = \frac{\mu^k i (1 + k=\circ)_i (\mathbb{R} - i - k=\circ)}{i (\mathbb{R})} - (1 + k=\circ; \mathbb{R} - i - k=\circ; 1 - u) + x^k u^\mathbb{R}; \quad k > i^\circ$$

$$\text{mode} = \mu \frac{\tilde{\mathbb{A}}^\circ i - 1}{\mathbb{R}^\circ + 1}; \quad \circ > 1; \text{ else } 0$$

$$\frac{\partial F}{\partial \mathbb{R}} = -i u^\mathbb{R} \log u$$

$$\frac{\partial F}{\partial \mu} = -i f(x) \frac{x}{\mu}$$

$$\frac{\partial F}{\partial \circ} = f(x) \frac{x}{\circ} \log \frac{\mu x}{\mu}$$

A.2.2.3 Inverse Burr - $\zeta; \mu; \circ$

$$f(x) = \frac{\zeta^\circ (x=\mu)^\zeta}{x[1 + (x=\mu)^\circ]^{\zeta+1}}$$

$$F(x) = u^\zeta; \quad u = \frac{(x=\mu)^\circ}{1 + (x=\mu)^\circ}$$

$$E[X^k] = \frac{\mu^k i (\zeta + k=\circ)_i (1 - i - k=\circ)}{i (\zeta)}; \quad i \zeta^\circ < k < \circ$$

$$E[(X \wedge x)^k] = \frac{\mu^k i (\zeta + k=\circ)_i (1 - i - k=\circ)}{i (\zeta)} - (\zeta + k=\circ; 1 - i - k=\circ; u) + x^k [1 - u^\zeta]; \quad k > i \zeta^\circ$$

$$\text{mode} = \mu \frac{\tilde{\mathbb{A}}^\circ i - 1}{\circ + 1}; \quad \zeta^\circ > 1; \text{ else } 0$$

$$\frac{\partial F}{\partial \zeta} = u^\zeta \log u$$

$$\frac{\partial F}{\partial \mu} = -i f(x) \frac{x}{\mu}$$

$$\frac{\partial F}{\partial \circ} = f(x) \frac{x}{\circ} \log \frac{\mu x}{\mu}$$

A.2.3 Two parameter distributions

A.2.3.1 Pareto - $\alpha; \mu$

$$f(x) = \frac{\mu^\alpha}{(x + \mu)^{\alpha+1}}$$

$$F(x) = 1 - \frac{\mu}{x + \mu}$$

$$E[X^k] = \frac{\mu^k \alpha (\alpha - 1) \dots (\alpha - k + 1)}{\alpha!}; \quad \alpha > k > 1$$

$$E[X^k] = \frac{\mu^k k!}{(\alpha - k)!}, \text{ if } k \text{ is an integer}$$

$$E[X^{-1}] = \frac{\mu}{\alpha - 1}; \quad \alpha > 1$$

$$E[X^{-x}] = \alpha \mu \log \frac{\mu}{x + \mu}; \quad \alpha = 1$$

$$E[(X^{-1})^k] = \frac{\mu^k \alpha (\alpha - 1) \dots (\alpha - k + 1)}{\alpha!} - [k + 1; \alpha - k; x = (x + \mu)] + x^k \frac{\mu}{x + \mu}; \text{ all } k$$

mode = 0

$$\frac{\partial F}{\partial \alpha} = \alpha \frac{\mu}{x + \mu} \log \frac{\mu}{x + \mu}$$

$$\frac{\partial F}{\partial \mu} = \alpha f(x) \frac{x}{\mu}$$

$$\alpha = 2 \frac{t_i - m^2}{t_i - 2m^2}; \quad \hat{\mu} = \frac{mt}{t_i - 2m^2}$$

A.2.3.2 Inverse Pareto - $\lambda; \mu$

$$f(x) = \frac{\lambda \mu x^{\lambda-1}}{(x + \mu)^\lambda}$$

$$F(x) = \frac{\mu}{x + \mu}$$

$$E[X^k] = \frac{\mu^k \lambda (\lambda - k)}{\lambda}; \quad \lambda > k > 1$$

$$E[X^k] = \frac{\mu^k (k - \lambda)!}{(\lambda - k)!}, \text{ if } k \text{ is a negative integer}$$

$$E[(X \wedge x)^k] = \mu^k \int_0^{x/(x+\mu)} y^{\zeta+k-1} (1-y)^{\zeta-1} dy + x^k \int_{x/(x+\mu)}^1 \frac{x}{x+\mu} dy; \quad k > \zeta$$

$$\text{mode} = \mu \frac{\zeta-1}{2}; \quad \zeta > 1; \text{ else } 0$$

$$\frac{\partial F}{\partial \zeta} = \frac{x}{x+\mu} \log \frac{x}{x+\mu}$$

$$\frac{\partial F}{\partial \mu} = \int_0^1 f(x) \frac{x}{\mu}$$

A.2.3.3 Loglogistic - $\alpha; \mu$

$$f(x) = \frac{\alpha(x-\mu)^\alpha}{x[1+(x-\mu)^\alpha]^2}$$

$$F(x) = u; \quad u = \frac{(x-\mu)^\alpha}{1+(x-\mu)^\alpha}$$

$$E[X^k] = \mu^k \int_0^1 (1+u^\alpha)^{-1} (1-u^\alpha)^{-1} u^k du; \quad \alpha < k < \infty$$

$$E[(X \wedge x)^k] = \mu^k \int_0^{x/\mu} (1+u^\alpha)^{-1} (1-u^\alpha)^{-1} u^k du + x^k \int_{x/\mu}^1 (1-u^\alpha)^{-1} u^k du; \quad k > \alpha$$

$$\text{mode} = \mu \frac{\alpha-1}{\alpha+1}; \quad \alpha > 1; \text{ else } 0$$

$$\frac{\partial F}{\partial \alpha} = f(x) \frac{x}{\mu} \log \frac{x}{\mu}$$

$$\frac{\partial F}{\partial \mu} = \int_0^1 f(x) \frac{x}{\mu}$$

$$\hat{\alpha} = \frac{2 \log(3)}{\log(q) - \log(p)}; \quad \hat{\mu} = \exp \frac{\log(q) + \log(p)}{2}$$

A.2.3.4 Paralogistic - $\alpha; \mu$

This is a Burr distribution with $\alpha = \alpha$.

$$f(x) = \frac{\alpha^2(x-\mu)^\alpha}{x[1+(x-\mu)^\alpha]^{\alpha+1}}$$

$$F(x) = 1 - \frac{1}{1+(x-\mu)^\alpha}$$

$$E[X^k] = \frac{\mu^k i (1 + k = \textcircled{R}) i (\textcircled{R} i k = \textcircled{R})}{i (\textcircled{R})}; \quad i \textcircled{R} < k < \textcircled{R}^2$$

$$E[(X \wedge x)^k] = \frac{\mu^k i (1 + k = \textcircled{R}) i (\textcircled{R} i k = \textcircled{R})}{i (\textcircled{R})} - (1 + k = \textcircled{R}; \textcircled{R} i k = \textcircled{R}; 1 i u) + x^k u^{\textcircled{R}}; \quad k > i \textcircled{R}$$

$$\text{mode} = \mu \frac{\textcircled{R} i 1}{\textcircled{R}^2 + 1} \uparrow_{1 = \textcircled{R}}; \quad \textcircled{R} > 1; \text{ else } 0$$

$$\frac{\partial F}{\partial \textcircled{R}} = i u^{\textcircled{R}} [\log u + \textcircled{R} (1 i u) \log(\mu = x)]$$

$$\frac{\partial F}{\partial \mu} = i f(x) \frac{x}{\mu}$$

Starting values can use estimates from the loglogistic (use \circ for \textcircled{R}) or Pareto (use \textcircled{R}) distributions.

A.2.3.5 Inverse paralogistic - $\zeta; \mu$

This is an inverse Burr distribution with $\circ = \zeta$.

$$f(x) = \frac{\zeta^2 (x = \mu)^{\zeta^2}}{x [1 + (x = \mu)^{\zeta}]^{\zeta + 1}}$$

$$F(x) = u^{\zeta}; \quad u = \frac{(x = \mu)^{\zeta}}{1 + (x = \mu)^{\zeta}}$$

$$E[X^k] = \frac{\mu^k i (\zeta + k = \zeta) i (1 i k = \zeta)}{i (\zeta)}; \quad i \zeta^2 < k < \zeta$$

$$E[(X \wedge x)^k] = \frac{\mu^k i (\zeta + k = \zeta) i (1 i k = \zeta)}{i (\zeta)} - (\zeta + k = \zeta; 1 i k = \zeta; u) + x^k [1 i u^{\zeta}]; \quad k > i \zeta^2$$

$$\text{mode} = \mu (\zeta i 1)^{1 = \zeta}; \quad \zeta > 1; \text{ else } 0$$

$$\frac{\partial F}{\partial \zeta} = u^{\zeta} [\log u i \zeta (1 i u) \log(\mu = x)]$$

$$\frac{\partial F}{\partial \mu} = i f(x) \frac{x}{\mu}$$

Starting values can use estimates from the loglogistic (use \circ for ζ) or inverse Pareto (use ζ) distributions.

A.3 Transformed gamma family

A.3.1 Three parameter distributions

A.3.1.1 Transformed gamma - $(\theta; \mu; \zeta)$

$$\begin{aligned}
 f(x) &= \frac{\zeta u^{\theta} e^{-u}}{x^{\theta} \Gamma(\theta)}; \quad u = (x-\mu)^{\zeta} \\
 F(x) &= I_{\theta}(\theta; u) \\
 E[X^k] &= \frac{\mu^k I_{\theta}(\theta + k = \zeta)}{\Gamma(\theta)}; \quad k > \theta / \zeta \\
 E[(X - \mu)^k] &= \frac{\mu^k I_{\theta}(\theta + k = \zeta)}{\Gamma(\theta)} I_{\theta}(\theta + k = \zeta; u) + x^k [1 - I_{\theta}(\theta; u)]; \quad k > \theta / \zeta \\
 \text{mode} &= \mu + \frac{\mu^{\theta/\zeta} \Gamma(1 - \theta/\zeta)}{\zeta}; \quad \theta/\zeta > 1; \text{ else } 0
 \end{aligned}$$

$$\begin{aligned}
 \frac{\partial F}{\partial \theta} &= -I_{\theta}^{\prime}(\theta; u) \\
 \frac{\partial F}{\partial \mu} &= -I_{\theta} f(x) \frac{x}{\mu} \\
 \frac{\partial F}{\partial \zeta} &= f(x) \frac{x}{\zeta} \log \frac{\mu x}{\mu}
 \end{aligned}$$

A.3.1.2 Inverse transformed gamma - $(\theta; \mu; \zeta)$

$$\begin{aligned}
 f(x) &= \frac{\zeta u^{\theta} e^{-u}}{x^{\theta} \Gamma(\theta)}; \quad u = (\mu-x)^{\zeta} \\
 F(x) &= 1 - I_{\theta}(\theta; u) \\
 E[X^k] &= \frac{\mu^k I_{\theta}(\theta - k = \zeta)}{\Gamma(\theta)}; \quad k < \theta / \zeta \\
 E[(X - \mu)^k] &= \frac{\mu^k I_{\theta}(\theta - k = \zeta)}{\Gamma(\theta)} [1 - I_{\theta}(\theta - k = \zeta; u)] + x^k I_{\theta}(\theta; u) \\
 &= \frac{\mu^k I_{\theta}(\theta - k = \zeta)}{\Gamma(\theta)} G(\theta - k = \zeta; u) + x^k I_{\theta}(\theta; u); \quad \text{all } k \\
 \text{mode} &= \mu - \frac{\mu^{\theta/\zeta} \Gamma(1 - \theta/\zeta)}{\zeta}
 \end{aligned}$$

$$\frac{\partial F}{\partial \theta} = -I_{\theta}^{\prime}(\theta; u)$$

$$\frac{\partial F}{\partial \mu} = \int_0^x f(x) \frac{x}{\mu} \mu^{\frac{x}{\mu}-1} dx$$

$$\frac{\partial F}{\partial \lambda} = \int_0^x f(x) \frac{x}{\lambda} \log \frac{\mu x}{\lambda} \mu^{\frac{x}{\mu}-1} dx$$

A.3.2 Two parameter distributions

A.3.2.1 Gamma - $\alpha; \mu$

$$f(x) = \frac{(x/\mu)^\alpha e^{-x/\mu}}{\Gamma(\alpha)}$$

$$F(x) = \int_0^x f(t) dt$$

$$E[X^k] = \frac{\mu^k \Gamma(\alpha + k)}{\Gamma(\alpha)}, \quad k > \alpha$$

$$E[X^k] = \mu^k \frac{\Gamma(\alpha + k - 1)}{\Gamma(\alpha)}; \text{ if } k \text{ is an integer}$$

$$E[(X - \mu)^k] = \frac{\mu^k \Gamma(\alpha + k)}{\Gamma(\alpha)} \int_0^\infty f(t) dt + x^k [1 - F(x)]; \quad k > \alpha$$

$$= \alpha(\alpha + 1) \dots (\alpha + k - 1) \mu^k \int_0^\infty f(t) dt + x^k [1 - F(x)]; \quad k \text{ an integer}$$

$$\text{mode} = \mu(\alpha - 1); \quad \alpha > 1; \text{ else } 0$$

$$\frac{\partial F}{\partial \alpha} = \int_0^x f(t) \ln t dt$$

$$\frac{\partial F}{\partial \mu} = \int_0^x f(t) \frac{x}{\mu} dt$$

$$\hat{\alpha} = \frac{m^2}{t_i m^2}; \quad \hat{\mu} = \frac{t_i m^2}{m}$$

A.3.2.2 Inverse gamma - $\alpha; \mu$

$$f(x) = \frac{(\mu/x)^\alpha e^{-\mu/x}}{\Gamma(\alpha)}$$

$$F(x) = \int_0^x f(t) dt$$

$$E[X^k] = \frac{\mu^k \Gamma(\alpha - k)}{\Gamma(\alpha)}, \quad k < \alpha$$

$$E[X^k] = \frac{\mu^k}{(\alpha - 1) \dots (\alpha - k)}; \text{ if } k \text{ is an integer}$$

$$E[(X - \mu)^k] = \frac{\mu^k \Gamma(\alpha - k)}{\Gamma(\alpha)} [1 - F(\mu/x)] + x^k f(x)$$

$$\begin{aligned}
 &= \frac{\mu^k i^{(k)} G(i, k; \mu=x) + x^k i^{(k)} G(i, k; \mu=x)}{i^{(k)}}; \text{ all } k \\
 &= \frac{\mu^k}{(i-1) \dots (i-k)} G(i, k; \mu=x) + x^k i^{(k)} G(i, k; \mu=x); \text{ } k \text{ an integer} \\
 \text{mode} &= \mu = (i+1)
 \end{aligned}$$

$$\frac{\partial F}{\partial i} = i^{-1} i^{(k)} G(i, k; \mu=x)$$

$$\frac{\partial F}{\partial \mu} = i f(x) \frac{x}{\mu}$$

$$\hat{i} = \frac{2t_i m^2}{t_i m^2}; \quad \hat{\mu} = \frac{mt}{t_i m^2}$$

A.3.2.3 Weibull - $\mu; \lambda$

$$\begin{aligned}
 f(x) &= \frac{\lambda (x-\mu)^\lambda e^{-\lambda(x-\mu)^\lambda}}{x} \\
 F(x) &= 1 - e^{-\lambda(x-\mu)^\lambda} \\
 E[X^k] &= \mu^k i (1+k=\lambda); \quad k > i \lambda \\
 E[(X - \mu)^k] &= \mu^k i (1+k=\lambda) [1+k=\lambda; (x-\mu)^\lambda] + x^k e^{-\lambda(x-\mu)^\lambda}; \quad k > i \lambda \\
 \text{mode} &= \mu \frac{\lambda i - 1}{\lambda}; \quad \lambda > 1; \text{ else } 0
 \end{aligned}$$

$$\frac{\partial F}{\partial \lambda} = f(x) \frac{x}{\lambda} \log \frac{\mu x^\lambda}{\mu}$$

$$\frac{\partial F}{\partial \mu} = i f(x) \frac{x}{\mu}$$

$$\hat{\mu} = \exp \frac{\bar{A} g \log(p) i \log(q)}{g i - 1}; \quad g = \frac{\log(\log(4))}{\log(\log(4-3))}; \quad \hat{\lambda} = \frac{\log(\log(4))}{\log(q) i \log(\hat{\mu})}$$

A.3.2.4 Inverse Weibull - $\mu; \lambda$

$$\begin{aligned}
 f(x) &= \frac{\lambda (\mu-x)^\lambda e^{-\lambda(\mu-x)^\lambda}}{x} \\
 F(x) &= e^{-\lambda(\mu-x)^\lambda}
 \end{aligned}$$

$$\begin{aligned}
 E[X^k] &= \mu^k \Gamma(1 - k); \quad k < 1 \\
 E[(X + x)^k] &= \mu^k \Gamma(1 - k) \Gamma(1 - k; (\mu+x)^{-1}) + x^k \Gamma(1 - k); \quad \text{all } k \\
 &= \mu^k \Gamma(1 - k) G[1 - k; (\mu+x)^{-1}] + x^k \Gamma(1 - k) \\
 \text{mode} &= \mu \frac{1}{1}
 \end{aligned}$$

$$\begin{aligned}
 \frac{\partial F}{\partial \mu} &= f(x) \frac{x}{\mu} \log \frac{x}{\mu} \\
 \frac{\partial F}{\partial \mu} &= -f(x) \frac{x}{\mu}
 \end{aligned}$$

$$\hat{\mu} = \exp \left[\frac{\log(\log(4)) - \log(\log(4-3))}{\log(\log(4)) - \log(\log(4-3))} \right]; \quad g = \frac{\log(\log(4))}{\log(\log(4-3))}; \quad \hat{\mu} = \frac{\log(\log(4))}{\log(\log(4)) - \log(\log(4-3))}$$

A.3.3 One parameter distributions

A.3.3.1 Exponential - μ

$$\begin{aligned}
 f(x) &= \frac{e^{-x/\mu}}{\mu} \\
 F(x) &= 1 - e^{-x/\mu} \\
 E[X^k] &= \mu^k \Gamma(k + 1); \quad k > -1 \\
 E[X^k] &= \mu^k k!; \quad \text{if } k \text{ is an integer} \\
 E[X + x] &= \mu(1 + e^{-x/\mu}) \\
 E[(X + x)^k] &= \mu^k \Gamma(k + 1) \Gamma(k + 1; x/\mu) + x^k e^{-x/\mu}; \quad k > -1 \\
 &= \mu^k k! \Gamma(k + 1; x/\mu) + x^k e^{-x/\mu}; \quad k \text{ an integer} \\
 \text{mode} &= 0
 \end{aligned}$$

$$\frac{\partial F}{\partial \mu} = -f(x) \frac{x}{\mu}$$

$$\hat{\mu} = m$$

A.3.3.2 Inverse exponential - μ

$$f(x) = \frac{\mu e^{-\mu/x}}{x^2}$$

$$\begin{aligned}
 F(x) &= e^{-\mu/x} \\
 E[X^k] &= \mu^k (1 - k); \quad k < 1 \\
 E[(X \wedge x)^k] &= \mu^k (1 - k) G(1 - k; \mu/x) + x^k (1 - e^{-\mu/x}); \quad \text{all } k \\
 \text{mode} &= \mu/2
 \end{aligned}$$

$$\frac{\partial F}{\partial \mu} = - \int_0^x f(x) \frac{x}{\mu^2}$$

$$\hat{\mu} = - \int_0^x q \log(1 - F(x))$$

A.4 Other distributions

A.4.1.1 Lognormal - μ, σ^2 (μ can be negative)

$$\begin{aligned}
 f(x) &= \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right); \quad z = \frac{\log x - \mu}{\sigma} \\
 F(x) &= \Phi\left(\frac{\log x - \mu}{\sigma}\right) \\
 E[X^k] &= \exp\left(k\mu + \frac{k^2\sigma^2}{2}\right) \\
 E[(X \wedge x)^k] &= \exp\left(k\mu + \frac{k^2\sigma^2}{2}\right) \Phi\left(\frac{\log x - \mu - k\sigma^2}{\sigma}\right) + x^k [1 - \Phi\left(\frac{\log x - \mu - k\sigma^2}{\sigma}\right)] \\
 \text{mode} &= \exp\left(\mu - \sigma^2\right)
 \end{aligned}$$

$$\frac{\partial F}{\partial \mu} = \int_0^x f(x) \frac{x}{\mu}$$

$$\frac{\partial F}{\partial \sigma^2} = - \int_0^x z f(x) \frac{x}{\sigma^3}$$

$$\hat{\mu} = \frac{\int_0^x \log(t) f(t) dt}{\int_0^x f(t) dt}; \quad \hat{\sigma}^2 = \frac{\int_0^x (\log(t) - \hat{\mu})^2 f(t) dt}{\int_0^x f(t) dt}$$

A.4.1.2 Inverse Gaussian - μ, λ

$$\begin{aligned}
 f(x) &= \frac{\lambda^{3/2}}{\sqrt{2\pi x^3}} \exp\left(-\frac{\lambda(x - \mu)^2}{2x}\right); \quad z = \frac{x - \mu}{\sqrt{x}} \\
 F(x) &= \Phi\left(\frac{\lambda}{\mu} \sqrt{x}\right) + \exp(2\mu/\lambda) \Phi\left(\frac{\lambda}{\mu} \sqrt{\frac{\mu}{x}}\right); \quad y = \frac{x - \mu}{\sqrt{x}} \\
 E[X] &= \mu; \quad \text{Var}[X] = \frac{3\mu^3}{\lambda} \\
 E[X \wedge x] &= x \Phi\left(\frac{\lambda}{\mu} \sqrt{x}\right) + \mu \exp(2\mu/\lambda) \Phi\left(\frac{\lambda}{\mu} \sqrt{\frac{\mu}{x}}\right)
 \end{aligned}$$

$$\frac{\partial F}{\partial \mu} = \int_0^{\infty} \int_0^{\infty} \frac{z}{x} \frac{\mu}{x} \frac{x}{12} \exp(2\mu z) \frac{\mu}{x} \frac{x}{12} + \exp(2\mu z) \int_0^{\infty} \int_0^{\infty} \frac{z}{2(\mu x)^{1/2}} + \frac{2}{1} \exp(2\mu z) \frac{\mu}{x} \int_0^{\infty} \int_0^{\infty} \exp(2\mu z) \frac{\mu}{x} \frac{y}{2(\mu x)^{1/2}}$$

$$\hat{\mu} = m; \quad \hat{\mu} = \frac{m^3}{t_i m^2}$$

A.4.1.3 Single parameter Pareto - μ

$$f(x) = \frac{\mu^{\alpha}}{x^{\alpha+1}}; \quad x > \mu$$

$$F(x) = 1 - (\frac{\mu}{x})^{\alpha}; \quad x > \mu$$

$$E[X^k] = \frac{\mu^k}{\alpha - k}; \quad k < \alpha$$

$$E[(X - \mu)^k] = \frac{\mu^k}{\alpha - k} - \frac{k\mu^{\alpha}}{(\alpha - k)x^{\alpha-k}}$$

$$\text{mode} = \mu$$

$$\frac{\partial F}{\partial \alpha} = \int_0^{\infty} \int_0^{\infty} \frac{\mu}{x} \log \frac{\mu}{x}$$

$$\alpha = \frac{m}{m_i \mu}$$

Note - Although there appears to be two parameters, only α is a true parameter. The value of μ must be set in advance.

Appendix B

An Inventory of Discrete Distributions

B.1 Introduction

The sixteen models fall into three classes. The divisions are based on the algorithm by which the probabilities are computed. For some of the more familiar distributions these formulas will look different from the ones you may have learned, but they produce the same probabilities. After each name, the parameters are given. All parameters are positive unless otherwise indicated. In all cases p_k is the probability of observing k losses.

For finding moments, the most convenient form is to give the factorial moments. The j th factorial moment is $\mu^{(j)} = E[N(N-1)\cdots(N-j+1)]$. We have $E[N] = \mu^{(1)}$ and $\text{Var}(N) = \mu^{(2)} + \mu^{(1)} - \mu^{(1)^2}$.

The estimators which are presented are not intended to be useful estimators but rather for providing starting values for maximizing the likelihood (or other) function. For determining starting values, the following quantities are used (where n_k is the observed frequency at k (if for the last entry n_k represents the number of observations at k or more, assume it was at exactly k) and n is the sample size):

$$\hat{\mu} = \frac{1}{n} \sum_{k=1}^{\infty} kn_k; \quad \hat{\mu}^2 = \frac{1}{n} \sum_{k=1}^{\infty} k^2 n_k$$

When the method of moments is used to determine the starting value, a circumflex (e.g., $\hat{\mu}$) is used. For any other method, a tilde (e.g., $\tilde{\mu}$) is used. When the starting value formulas do not provide admissible parameter values, a truly crude guess is to set the product of all μ and λ parameters equal to the sample mean and all other parameters equal to 1. If there are two μ and/or λ parameters, an easy choice is to set each to the square root of the sample mean.

The last item presented is the probability generating function,

$$P(z) = E[z^N]:$$

B.2 The (a,b,0) class

The distributions in this class have support on 0, 1, For this class, a particular distribution is specified by setting p_0 and then using $p_k = (a + b/k)p_{k-1}$. Specific members are created by setting p_0 , a , and b . For any member, $p_{(1)} = (a + b)/(1 + a)$ and for higher j , $p_{(j)} = (aj + b)/(j + a)$. The variance is $(a + b)/(1 + a)^2$:

B.2.1.1 Poisson - λ

$$\begin{aligned} p_0 &= e^{-\lambda}; & a &= 0; & b &= \lambda \\ p_k &= \frac{e^{-\lambda} \lambda^k}{k!} \\ E[N] &= \lambda; & \text{Var}[N] &= \lambda \\ \hat{\lambda} &= \lambda \\ P(z) &= e^{-\lambda} e^{\lambda z} \end{aligned}$$

B.2.1.2 Geometric - \bar{p}

$$\begin{aligned} p_0 &= 1 - \bar{p}; & a &= \bar{p}; & b &= 0 \\ p_k &= \frac{\bar{p}^k}{(1 + \bar{p})^{k+1}} \\ E[N] &= \bar{p}; & \text{Var}[N] &= \bar{p}(1 + \bar{p}) \\ \hat{\bar{p}} &= \bar{p} \\ P(z) &= [1 - \bar{p}(z + 1)]^{-1} \end{aligned}$$

This is a special case of the negative binomial with $r = 1$.

B.2.1.3 Binomial - $q; m; (0 < q < 1; m \text{ an integer})$

$$\begin{aligned} p_0 &= (1 - q)^m; & a &= q; & b &= (m + 1)q \\ p_k &= \binom{m}{k} q^k (1 - q)^{m-k}; & k &= 0; 1; \dots; m \\ E[N] &= mq; & \text{Var}[N] &= mq(1 - q) \\ \hat{q} &= q \\ P(z) &= [1 - q(z + 1)]^m \end{aligned}$$

B.2.1.4 Negative binomial - ; r

$$\begin{aligned}
 p_0 &= (1 +)^{-r}; \quad a = -(1 +); \quad b = (r - 1)^{-1} = (1 +) \\
 p_k &= \frac{r(r + 1) \dots (r + k - 1)^{-k}}{k!(1 +)^{r+k}} \\
 E[N] &= r^{-1}; \quad \text{Var}[N] = r^{-1}(1 +) \\
 \Delta &= \frac{\Delta^2}{\Delta - 1}; \quad \hat{r} = \frac{\Delta^2}{\Delta^2 - \Delta} \\
 P(z) &= [1 - (z - 1)]^{-r}
 \end{aligned}$$

B.3 The (a,b,1) class

To distinguish this class from the (a; b; 0) class, the probabilities are denoted $\Pr(N = k) = p_k^M$ or $\Pr(N = k) = p_k^T$ depending on which sub-class is being represented. For this class p_0^M is arbitrary (that is, it is a parameter) and then p_1^M or p_1^T is a specified function of the parameters a and b. Subsequent probabilities are obtained recursively as in the (a; b; 0) class: $p_k^M = (a + b - k)p_{k-1}^M$; $k = 2; 3; \dots$ with the same recursion for p_k^M . There are two sub-classes of this class. When discussing their members, we often refer to the "corresponding" member of the (a; b; 0) class. This refers to the member of that class with the same values for a and b. The notation p_k will continue to be used for probabilities for the corresponding (a; b; 0) distribution.

B.3.1 The zero-truncated sub-class

The members of this class have $p_0^T = 0$ and therefore it need not be estimated. These distributions should only be used when a value of zero is impossible. The first factorial moment is $1_{(1)} = (a + b) / [(1 - a)(1 - p_0)]$ where p_0 is the value for the corresponding member of the (a; b; 0) class. For the logarithmic distribution (which has no corresponding member) $1_{(1)} = -1 / \log(1 -)$. Higher factorial moments are obtained recursively with the same formula as with the (a; b; 0) class. The variance is $(a + b) / [(1 - a)(1 - p_0)]^2$. For those members of the sub-class which have corresponding (a; b; 0) distributions, $p_k^T = p_k / (1 - p_0)$.

B.3.1.1 Zero-truncated Poisson - ;

$$\begin{aligned}
 p_1^T &= (e - 1)^{-1}; \quad a = 0; \quad b = \\
 p_k^T &= \frac{e^{-k}}{k!(e - 1)} \\
 E[N] &= (1 - e^{-1}); \quad \text{Var}[N] = (1 - e^{-1}) / (e - 1) = (1 - e^{-1})^2 \\
 \hat{\lambda} &= \log(n^{\wedge} = n_1)
 \end{aligned}$$

$$P(z) = \frac{e^{-z} i - 1}{e^{-z} i - 1}$$

B.3.1.2 Zero-truncated geometric - \bar{p}

$$\begin{aligned} p_1^T &= 1 - (1 + \bar{p}); \quad a = \bar{p} = (1 + \bar{p}); \quad b = 0 \\ p_k^T &= \frac{-k i - 1}{(1 + \bar{p})^k} \\ E[N] &= 1 + \bar{p}; \quad \text{Var}[N] = \bar{p}(1 + \bar{p}) \\ \Delta &= \bar{p} i - 1 \\ P(z) &= \frac{[1 - i - \bar{p}(z i - 1)]^{i - 1} i (1 + \bar{p})^{i - 1}}{1 - i (1 + \bar{p})^{i - 1}} \end{aligned}$$

This is a special case of the zero-truncated negative binomial with $r = 1$.

B.3.1.3 Logarithmic - \bar{p}

$$\begin{aligned} p_1^T &= \bar{p} = [(1 + \bar{p}) \log(1 + \bar{p})]; \quad a = \bar{p} = (1 + \bar{p}); \quad b = i - \bar{p} = (1 + \bar{p}) \\ p_k^T &= \frac{-k}{k(1 + \bar{p})^k \log(1 + \bar{p})} \\ E[N] &= \bar{p} = \log(1 + \bar{p}); \quad \text{Var}[N] = \bar{p} [1 + i - \bar{p} = \log(1 + \bar{p})] = \log(1 + \bar{p}) \\ \bar{p} &= \frac{n^{\bar{p}}}{n_1} i - 1 \text{ or } \frac{2(i - 1)}{\bar{p}} \\ P(z) &= 1 - i \frac{\log[1 - i - \bar{p}(z i - 1)]}{\log(1 + \bar{p})} \end{aligned}$$

This is a limiting case of the zero-truncated negative binomial as $r \rightarrow 0$.

B.3.1.4 Zero-truncated binomial - $q; m; (0 < q < 1; m \text{ an integer})$

$$\begin{aligned} p_1^T &= m(1 - i - q)^{m-1} q = [1 - i (1 - i - q)^m]; \quad a = i - q = (1 - i - q); \quad b = (m + 1)q = (1 - i - q) \\ p_k^T &= \frac{\binom{m}{k} q^k (1 - i - q)^{m-k}}{1 - i (1 - i - q)^m}; \quad k = 1; 2; \dots; m \\ E[N] &= mq = [1 - i (1 - i - q)^m] \\ \text{Var}[N] &= \frac{mq[(1 - i - q) i (1 - i - q + mq)(1 - i - q)^m]}{[1 - i (1 - i - q)^m]^2} \\ q &= \bar{p} = m \\ P(z) &= \frac{[1 + q(z i - 1)]^m i (1 - i - q)^m}{1 - i (1 - i - q)^m} \end{aligned}$$

B.3.1.5 Zero-truncated negative binomial - $r^-; r; (r > i - 1)$

$$\begin{aligned}
 p_1^T &= r^- = [(1 + \frac{r^-}{i})^{i-1}]; \quad a = \frac{r^-}{i} = (1 + \frac{r^-}{i}); \quad b = (r - i + 1)^- = (1 + \frac{r^-}{i}) \\
 p_k^T &= \frac{r(r+1)\dots(r+k-i)}{k![(1 + \frac{r^-}{i})^{i-1}]} \frac{r^-}{1 + \frac{r^-}{i}} \\
 E[N] &= r^- = [1 + \frac{r^-}{i}]^i \\
 Var[N] &= \frac{r^- [(1 + \frac{r^-}{i})^i (1 + \frac{r^-}{i} + r^-) (1 + \frac{r^-}{i})^i]}{[1 + \frac{r^-}{i}]^2} \\
 &= \frac{r^-^2}{i} + \frac{r^-^2}{i} \\
 P(z) &= \frac{[1 + \frac{r^-}{i} (z - 1)]^i}{1 + \frac{r^-}{i}}
 \end{aligned}$$

This distribution is sometimes called the extended truncated negative binomial distribution because the parameter r can extend below 0.

B.3.2 The zero-modified sub-class

A zero-modified distribution is created by starting with a truncated distribution and then placing an arbitrary amount of probability at zero. This probability, p_0^M , is a parameter. The remaining probabilities are adjusted accordingly. Values of p_k^M can be determined from the corresponding zero-truncated distribution as $p_k^M = (1 - p_0^M)p_k^T$ or from the corresponding $(a; b; 0)$ distribution as $p_k^M = (1 - p_0^M)p_k = (1 - p_0)$. The same recursion used for the zero-truncated sub-class applies.

The mean is $1 - p_0^M$ times the mean for the corresponding zero-truncated distribution. The variance is $1 - p_0^M$ times the zero-truncated variance plus $p_0^M(1 - p_0^M)$ times the square of the zero-truncated mean. The probability generating function is $P^M(z) = p_0^M + (1 - p_0^M)P(z)$ where $P(z)$ is the probability generating function for the corresponding zero-truncated distribution.

The maximum likelihood estimator of q_0 is always the sample relative frequency at 0.

B.4 Compound class

Members of this class are obtained by compounding one distribution with another. That is, let N be a discrete distribution, called the primary distribution and let $M_1; M_2; \dots$ be identically and independently distributed with another discrete distribution called the secondary distribution. The compound distribution is $S = M_1 + \dots + M_N$. The probabilities for the compound distributions are found from

$$p_k = \sum_{y=1}^k (a + by = k) f_y p_{k-y} = (1 - af_0)$$

for $n = 1; 2; \dots$ where a and b are the usual values for the primary distribution (which must be a member of the $(a; b; 0)$ class) and f_y is p_y for the secondary distribution. The only two primary distributions used here are Poisson (for which $p_0 = \exp[-(1 - f_0)]$) and geometric (for which $p_0 = 1/[1 + (1 - f_0)]$). As this information completely describes these distributions, only the names and starting values are given below.

The moments can be found from the moments of the individual distributions:

$$E[S] = E[N]E[M] \text{ and } \text{Var}[S] = E[N]\text{Var}[M] + \text{Var}[N]E[M]^2:$$

The probability generating function is $P(z) = P_{\text{primary}}(P_{\text{secondary}}(z))$.

In the following list the primary distribution is always named ...rst. For the ...rst, second, and fourth distributions, the secondary distribution is the $(a; b; 0)$ class member with that name. For the third and the last three distributions (the Poisson-ETNB and its two special cases) the secondary distribution is the zero-truncated version.

B.4.1 Some compound distributions

B.4.1.1 Poisson-binomial - $\mu; q; m$ ($0 < q < 1$, m an integer)

$$\hat{\mu} = \frac{\mu^2 - \mu}{m - 1}; \quad \hat{\sigma}^2 = \mu(mq) \text{ or } q = 0.5; \quad \tilde{\sigma}^2 = 2\mu - m$$

B.4.1.2 Poisson-Poisson - $\mu_1; \mu_2$

The parameter μ_1 is for the primary Poisson distribution and μ_2 is for the secondary Poisson distribution. This distribution is also called the Neyman Type A.

$$\tilde{\sigma}_1 = \tilde{\sigma}_2 = \mu_1 - \mu_2$$

B.4.1.3 Geometric-extended truncated negative binomial - $\mu_1; \mu_2; r$ ($r > \mu_1 + 1$)

The parameter μ_1 is for the primary geometric distribution. The last two parameters are for the secondary distribution. The truncated version is used so that the extension of r is available.

$$\tilde{\sigma}_1 = \tilde{\sigma}_2 = \mu_1 - \mu_2$$

B.4.1.4 Geometric-Poisson - $\mu; \mu_2$

$$\tilde{\sigma} = \tilde{\sigma}_2 = \mu_2 - \mu$$

B.4.1.5 Poisson-extended truncated negative binomial- $\lambda, \mu; r$

$$P = \frac{\lambda(K + 3\lambda^2 + 2\lambda) \mu^{n_1} (2\lambda^2 \mu)^{n_0}}{\lambda(K + 3\lambda^2 + 2\lambda) \mu^{n_1} (\lambda^2 \mu)^{n_0}}; \approx = \frac{\lambda^2 \mu}{\lambda(1 + \lambda)}; \tilde{\mu} = \frac{\lambda}{\lambda^\Delta}, \text{ or}$$

$$P = \frac{\lambda^2 n_1 = n \mu^{n_1} n_0 = n}{(\lambda^2 \mu)^{n_1} (n_0 = n) \log(n_0 = n) \mu^{n_0} (\lambda n_0 = n \mu^{n_1 = n})}; \approx = \frac{\lambda^2 \mu}{\lambda(1 + \lambda)}; \tilde{\mu} = \frac{\lambda}{\lambda^\Delta}$$

where $K = \frac{1}{n} \sum_{k=0}^1 k^3 n_k + 3 \frac{1}{n} \sum_{k=0}^1 k^2 n_k + 2 \lambda^3$

This distribution is also called the generalized Poisson-Pascal.

B.4.1.6 Polya-Aeppli - $\lambda; \mu$

$$\Delta = \frac{\lambda^2 \mu}{2\lambda}; \hat{\mu} = \frac{\lambda}{1 + \Delta}$$

This is a special case of the Poisson-extended truncated negative binomial with $r = 1$. It is actually a Poisson-truncated geometric.

B.4.1.7 Poisson-inverse Gaussian - $\lambda; \mu$

$$\tilde{\mu} = \mu \log(n_0 = n); \hat{\mu} = 4(\lambda \mu)^{-1}$$

This is a special case of the Poisson-extended truncated negative binomial with $r = \mu^{-1}$.

B.5 A hierarchy of discrete distributions

The following table indicates which distributions are special or limiting cases of others. For the special cases, one parameter is set equal to a constant to create the special case. For the limiting cases, two parameters go to infinity or zero in some special way.

Distribution	is a special case of	is a limiting case of
Poisson	ZM Poisson	Neg. bin., P-bin. P-inv. Gaussian Polya-Aeppli Neyman A
ZT Poisson	ZM Poisson	ZT Neg. bin.
ZM Poisson		ZM Neg. bin
Geometric	Neg. bin. ZM geom	Geometric-Poisson
ZT Geometric	ZT Neg. bin.	
ZM Geometric	ZM Neg. bin.	
Logarithmic		ZT Neg. bin.
ZM Logarithmic		ZM Neg. bin.
Binomial	ZM binomial 0	
Neg. bin.	ZM Neg. bin.	P-ETNB
P-inv. Gaussian	P-ETNB	
Polya-Aeppli	P-ETNB	
Neyman A		P-ETNB

Appendix C

The Simplex Method

The method (which is not related to the simplex method from operations research) was introduced for use with maximum likelihood estimation by Nelder and Mead in 1965 [2]. An excellent reference (and the source of the particular version presented here) is Sequential Simplex Optimization by Walters, Parker, Morgan, and Deming [4].

Let x be a $k + 1$ vector and $f(x)$ be the function in question. The iterative step begins with $k + 1$ vectors, $x_1; \dots; x_{k+1}$, and the corresponding functional values, $f_1; \dots; f_{k+1}$. At any iteration the points will be ordered so that $f_2 < f_3 < \dots < f_{k+1}$. When starting, also arrange for $f_1 < f_2$. Three of the points have names: x_1 is called worstpoint, x_2 is called second-worstpoint, and x_{k+1} is called bestpoint. It should be noted that after the first iteration these names may not perfectly describe the points. Now identify five new points. The first one, y_1 , is the center of $x_2; \dots; x_{k+1}$. That is, $y_1 = \frac{1}{k} \sum_{j=2}^{k+1} x_j$, and is called midpoint. The other four points are found as follows:

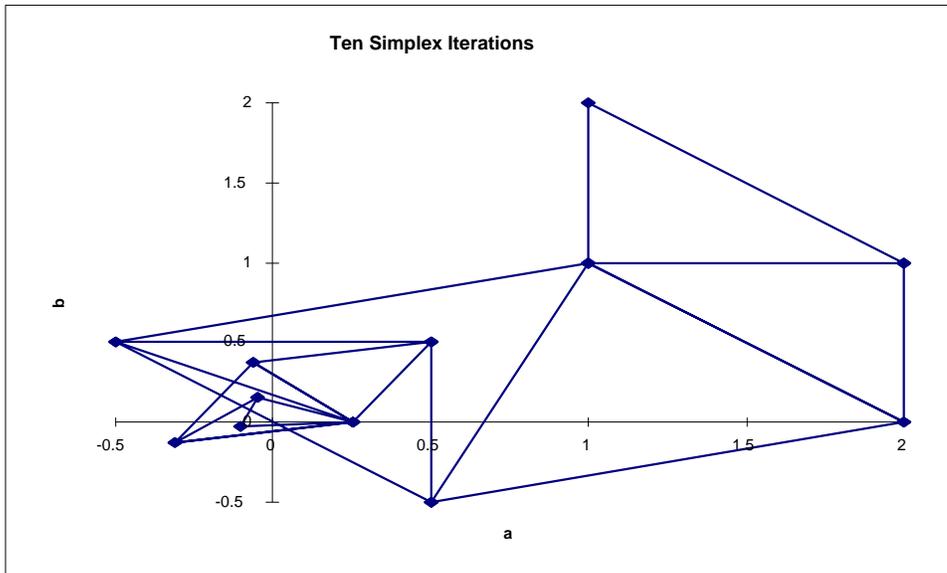
$$\begin{aligned} y_2 &= 2y_1 - x_1; && \text{refpoint} \\ y_3 &= 2y_2 - x_1; && \text{doublepoint} \\ y_4 &= (y_1 + y_2)/2; && \text{halfpoint} \\ y_5 &= (y_1 + x_1)/2; && \text{centerpoint.} \end{aligned}$$

Then let $g_2; \dots; g_5$ be the corresponding functional values (the value at y_1 is never used). The key is to replace worstpoint (x_1) with one of these points. The decision process proceeds as follows:

1. If $f_2 < g_2 < f_{k+1}$ then replace it with refpoint.
2. If $g_2 > f_{k+1}$ and $g_3 > f_{k+1}$ then replace it with doublepoint.
3. If $g_2 > f_{k+1}$ and $g_3 < f_{k+1}$ then replace it with refpoint.
4. If $f_1 < g_2 < f_2$ then replace it with halfpoint.
5. If $g_2 < f_1$ then replace it with centerpoint.

After the replacement has been made, the old secondworstpoint becomes the new worst-point. The remaining k points are then ordered. The one with the smallest functional value becomes the new secondworstpoint and the one with the largest functional value becomes the new bestpoint. In practice, there is no need to compute y_3 and g_3 until you have reached step 2. Also note that at most one of the pairs $(y_4; g_4)$ and $(y_5; g_5)$ needs to be obtained, depending on which (if any) of the conditions in steps 4 and 5 hold.

The following graph indicates the progress of the simplex. The problem is to maximize the function $f(x; y) = |a^2 - 2b^2|$. The starting simplex is $(1; 1)$, $(1; 2)$, $(2; 1)$ and the maximum is at $(0; 0)$.



Appendix D

Frequency and Severity Relationships

Let N be the number of losses random variable and X be the severity random variable. If there is a deductible of d imposed, there are two ways to modify X . One is to create Y , the amount paid per loss:

$$Y = \begin{cases} 0; & X \leq d \\ X - d & X > d \end{cases}$$

In this case the appropriate frequency distribution continues to be N .

An alternative approach is to create Y^* , the amount paid per payment:

$$Y^* = \begin{cases} \text{undefined} & X \leq d \\ X - d & X > d \end{cases}$$

In this case the frequency random variable must be altered to reflect the number of payments. Let this variable be N^* . Assume that for each loss the probability is $v = 1 - F_X(d)$ that a payment will result. Further assume that the incidence of making a payment is independent of the number of losses. Then $N^* = L_1 + L_2 + \dots + L_N$ where L_i is 0 with probability $1 - v$ and 1 with probability v . Probability generating functions yield the following relationships

N	Parameters for N^*
Poisson	$p_0^{M^*} = v$
ZM Poisson	$p_0^{M^*} = \frac{p_0^M (1 - v) + v e^{-\lambda} p_0^M e^{\lambda v}}{1 - v e^{-\lambda} (1 - p_0^M)}; p_0^{M^*} = v$
Binomial	$q^* = vq$
ZM binomial	$p_0^{M^*} = \frac{p_0^M (1 - q)^m + (1 - vq)^m p_0^M (1 - vq)^m}{1 - (1 - q)^m}; q^* = vq$
Negative binomial	$r^* = v^-; r^* = r$
ZM negative binomial	$p_0^{M^*} = \frac{p_0^M (1 + r)^{-i} r + (1 + v^-)^i r p_0^M (1 + v^-)^i r}{1 - (1 + r)^{-i} r}; r^* = r$
ZMlogarithmic	$p_0^{M^*} = 1 - (1 - p_0^M) \log(1 + v^-) = \log(1 + v^-); r^* = v^-$

The geometric distribution is not presented as it is a special case of the negative binomial with $r = 1$. For zero-truncated distributions, the above is still used as the distribution for N^* will now be zero-modified. For compound distributions, modify only the secondary distribution. For ETNB secondary distributions the parameter for the primary distribution is multiplied by $1 - p_0^{M^*}$ as obtained above while the secondary distribution remains zero-truncated (however $v^* = v^-$):

There are occasions in which frequency data are collected which provide a model for N^* . There would have to have been a deductible d in place and therefore v is available. It is possible to recover the distribution for N , although there is no guarantee that reversing the process will produce a legitimate probability distribution. The solutions are the same as above only now $v = 1 - F_X(d)$.

Now suppose the current frequency model is N^* which is appropriate for a deductible of d . Now suppose the deductible is to be changed to d^0 . The new frequency for payments is N^0 and is of the same type. Then use the table with $v = [1 - F_X(d^0)] = [1 - F_X(d)]$.

Finally, suppose a truncating limit is imposed. In this case

$$\begin{aligned} Y^* &= \text{undefined, } X \cdot d \\ &= X - d; \quad d < X < u \\ &= \text{undefined, } X \leq u: \end{aligned}$$

Then proceed as above, only now $v = F_X(u) - F_X(d)$.

Appendix E

Limited Expected Value Calculations

The limited expected value is defined as

$$E[X \wedge x] = \int_0^x tf(t)dt + x[1 - F(x)]:$$

It can be used to obtain the expected payment when the loss X or aggregate loss S is modified by a deductible (d), a limit (u), or a coinsurance(α). Suppose all three are present, so that the amount paid, Y , is related to X by

$$\begin{aligned} Y &= 0; & X < d \\ &= \alpha(X - d); & d \leq X \leq u \\ &= \alpha(u - d); & X > u: \end{aligned}$$

Then $E[Y] = \alpha[E[X \wedge u] - E[X \wedge d]]$. If there is inflation so that $X^* = (1 + r)X$ and Y is based on X^* then $E[Y] = \alpha(1 + r)[E[X \wedge u(1 + r)] - E[X \wedge d(1 + r)]]$.

The limited expected squared value is defined as

$$E[(X \wedge x)^2] = \int_0^x t^2f(t)dt + x^2[1 - F(x)]:$$

We then have $E[Y^2] = \alpha^2[E[(X \wedge u)^2] - E[(X \wedge d)^2] - 2dE[X \wedge u] + 2dE[X \wedge d]]$. The variance can be found by subtracting the square of the mean. These calculations can apply to either the individual or the aggregate losses.

Appendix F

The Recursive Formula

The recursive formula is

$$g_i = \frac{[p_{1-i} (a+b)p_0]f_i + \sum_{j=1}^i (a+bj=i)f_j g_{i-j}}{(1-i)af_0}; \quad i = 1; 2; \dots$$

where $g_i = \Pr(S = i)$; $p_i = \Pr(N = i)$, and $f_j = \Pr(X = j)$. The formula must be initialized with the value of g_0 . These values are given below. It should be noted that if N is a member of the $(a; b; 0)$ class, $p_{1-i} (a+b)p_0 = 0$ and so the ...rst term will vanish. If N is a member of the compound class, the recursion must be run twice. The ...rst pass uses the secondary distribution for p_i , a , and b . The second pass uses the output from the ...rst pass as f_j and the primary distribution for p_i , a , and b .

Starting values (g_0) for recursions

Poisson	$\exp[-(f_0 - 1)]$
geometric	$[1 - (1 - f_0)]^{i-1}$
binomial	$[1 + q(f_0 - 1)]^m$
negative binomial	$[1 - (1 - f_0)]^{i-r}$
ZM Poisson	$p_0^M + (1 - p_0^M) \frac{\exp(-f_0) - 1}{f_0}$
ZM geometric	$p_0^M + (1 - p_0^M) \frac{1 - (1 - f_0)^M}{1 - (1 - f_0)}$
ZM binomial	$p_0^M + (1 - p_0^M) \frac{[1 + q(f_0 - 1)]^m - 1}{1 - (1 - q)^m}$
ZM neg. bin.	$p_0^M + (1 - p_0^M) \frac{[1 - (1 - f_0)]^{i-r} - 1}{1 - (1 - f_0)^{i-r}}$
ZM logarithmic	$p_0^M + (1 - p_0^M) \frac{\log[1 - (1 - f_0)]}{\log(1 - f_0)}$

Appendix G

Discretization of the Severity Distribution

There are two ways to discretize the severity distribution that are used in the program. One is the method of rounding and the other is a mean-preserving method.

G.1 The method of rounding

This method has two features, all probabilities are positive and the probabilities add to one. Let h be the span and let Y be the discretized version of X . If there are no modifications then

$$\begin{aligned} f_j &= \Pr(Y = jh) = \Pr[(j-1)h < X < (j+1)h] \\ &= F_X[(j+1)h] - F_X[(j-1)h]: \end{aligned}$$

Suppose a deductible of d , limit of u , and coinsurance of α are to be applied. If the modifications are to be applied before the discretization then

$$\begin{aligned} g_0 &= \frac{F_X(d+h) - F_X(d)}{1 - F_X(d)} \\ g_j &= \frac{F_X[d+(j+1)h] - F_X[d+(j-1)h]}{1 - F_X(d)}; \quad j = 1; \dots; \frac{u-d}{h} \\ g_{(u-d)/h} &= \frac{1 - F_X(u-h)}{1 - F_X(d)}. \end{aligned}$$

where $g_j = \Pr(Z = jh)$ where Z is the modified distribution. This method does not require that the limits be multiples of h but does require that $u-d$ be a multiple of h . This method gives the probabilities of payments per payment.

Finally, if there is truncation from above at u , change all denominators to $F_X(u) - F_X(d)$ and also change the numerator of $g_{(u-d)/h}$ to $F_X(u) - F_X(u-h)$.

G.2 Mean-preserving

This method ensures that the discretized distribution has the same mean as the original severity distribution. With no modifications the discretization is

$$\begin{aligned} f_0 &= 1 - E[X^h] \\ f_j &= (2E[X^{jh}] - E[X^{(j-1)h}] - E[X^{(j+1)h}]) / h; \quad j = 1, 2, \dots \end{aligned}$$

For the modified distribution,

$$\begin{aligned} g_0 &= 1 - \frac{E[X^{d+h}] - E[X^d]}{h[1 - F_X(d)]} \\ g_j &= \frac{2E[X^{d+jh}] - E[X^{d+(j-1)h}] - E[X^{d+(j+1)h}]}{h[1 - F_X(d)]}; \quad j = 1, \dots; \frac{u-d}{h} \leq j < 1 \\ g_{(u-d)/h} &= \frac{E[X^{uh}] - E[X^{(u-1)h}]}{h[1 - F_X(d)]}. \end{aligned}$$

To incorporate truncation from above, change the denominators to $h[F_X(u) - F_X(d)]$. Also, add $h[1 - F_X(u)]$ to the numerator of g_0 and subtract it from the numerator for $g_{(u-d)/h}$.

G.3 Undiscretization of a discretized distribution

Let g_0 be the true probability that the random variable is zero. Let $p_j = \Pr(S^a = jh)$ where S^a is a discretized distribution and h is the span. The following are approximations for the cdf and LEV of S , the true distribution which was discretized as S^a . They are all based on the assumption that S has a uniform distribution over the interval from $(j-1)h$ to $(j+1)h$ for integral j . The first interval is from 0 to $h/2$ and the probability $p_0 \leq g_0$ is assumed to be uniformly distributed over it. Let S^{aa} be the random variable with this approximate continuous distribution. Then the approximate distribution function can be found by interpolation as follows. First, let

$$F_j = F_{S^{aa}}[(j+1/2)h] = \sum_{i=0}^j p_i; \quad j = 0, 1, \dots$$

Then, for x in the interval $(j-1/2)h$ to $(j+1/2)h$,

$$\begin{aligned} F_{S^{aa}}(x) &= F_{j-1/2} + \int_{(j-1/2)h}^x h^{-1} p_j dt = F_{j-1/2} + [x - (j-1/2)h] h^{-1} p_j \\ &= F_{j-1/2} + [x - (j-1/2)h] h^{-1} (F_j - F_{j-1/2}) \\ &= (1-w)F_{j-1/2} + wF_j; \quad w = x - h(j-1/2) \end{aligned}$$

Because the first interval is only half as wide, the formula is, for $0 \leq x < h/2$,

$$F_{S^{aa}}(x) = (1-w)g_0 + wp_0; \quad w = 2x/h$$

It is also possible to express these formulas in terms of the discrete probabilities:

$$\begin{aligned} F_{S^{**}}(x) &= g_0 + \frac{2x}{h}(p_0 - g_0); \quad 0 < x \cdot h=2 \\ &= \sum_{i=0}^{j-1} p_i + \frac{x - (j-1)h}{h} p_j; \quad (j-1)h < x \cdot (j+1)h; \end{aligned}$$

With regard to the limited expected value, expressions for the r st and k th LEV 's are

$$\begin{aligned} E(S^{**} \wedge x) &= x(1 - g_0) + \frac{x^2}{h}(p_0 - g_0); \quad 0 < x \cdot h=2 \\ &= \frac{h}{4}(p_0 - g_0) + \sum_{i=1}^{j-1} i h p_i + \frac{x^2 - [(j-1)h]^2}{2h} p_j \\ &\quad + x[1 - F_{S^{**}}(x)]; \quad (j-1)h < x \cdot (j+1)h; \end{aligned}$$

and

$$\begin{aligned} E[(S^{**} \wedge x)^k] &= \frac{2x^{k+1}}{h(k+1)}(p_0 - g_0) + x^k[1 - F_{S^{**}}(x)]; \quad 0 < x \cdot h=2 \\ &= \frac{(h=2)^k(p_0 - g_0)}{k+1} + \sum_{i=1}^{j-1} \frac{h^k [(i+1)h - (i-1)h]^{k+1}}{k+1} p_i \\ &\quad + \frac{x^{k+1} - [(j-1)h]^{k+1}}{h(k+1)} p_j \\ &\quad + x^k[1 - F_{S^{**}}(x)]; \quad (j-1)h < x \cdot (j+1)h : \end{aligned}$$

Bibliography

- [1] Abramowitz, M. and Stegun, I. (1964), Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, New York: Wiley.
- [2] Nelder, J. and Mead, U. (1965), "A Simplex Method for Function Minimization," The Computer Journal.
- [3] Press, W., Flannery, B., Teukolsky, S., and Vetterling, W. (1988), Numerical Recipes in C, Cambridge: Cambridge University Press.
- [4] Walters, F., Parker, L., Morgan, S. and Deming, S. (1991), Sequential Simplex Optimization, Boca Raton, FL: CRC Press.