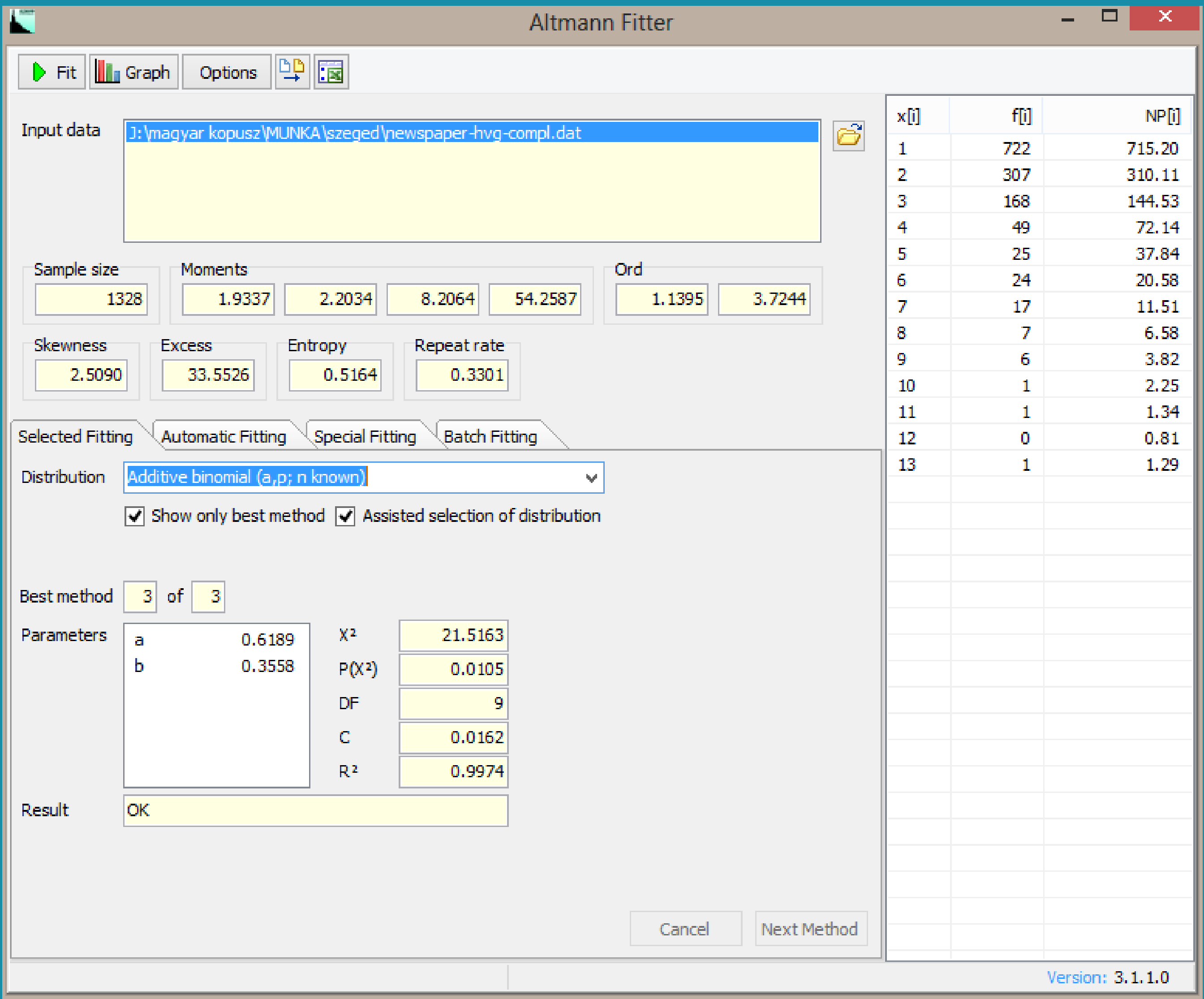
Altmann Fitter

(3.1)



User Guide

Contents

0. Purpose and Method

The Altmann Fitter is an interactive programme for the iterative fitting of univariate discrete probability distributions to frequency data. Its algorithm is based on the Nelder-Mead Simplex method with modifications and refinements. It aims at the analysis of data from all empirical scientific and technical domains and is optimised for application by practitioners.

More than 200 individual probability distributions are defined and implemented and can be used in various ways. The Altmann Fitter contains one of the most voluminous collections of distributions with information about all relevant properties of these distributions. These are automatically used for optimal data analysis.

The mathematical procedures are automated, i.e. no initial estimators or other parameters have to be specified by the user (except a number of explicitly controllable distribution variants for special cases). The programme iteratively improves the goodness-of-fit until no better solution can be found. The goodness-of-fit criterion for the iterative optimisation is based on the chi-square test. Nevertheless, several other criteria are evaluated and presented. A number of options and configurations enables the user to control the optimisation procedures.

1. Requirements and Installation

Altmann Fitter runs under all Microsoft Windows® versions since Windows XP® and including Windows 8®. For best performance, the computer should be equipped with at least 512 MB of RAM.

To install Altmann Fitter, copy the file *Altmann-Fitter v3.1.1 Setup.zip* to your hard disk and extract the files. Then double-click on the file *Altmann-Fitter v3.1.1 Setup.exe* and follow the instructions given during the installation. The Installer will propose a location on your hard disk for installation. You will be asked whether you wish to change the installation path.

2 Data Input

Data file structure

Whatever you want to find out with the help of Altmann Fitter, it will be an analysis of a data set or a number of data sets. We will call a data set a frequency distribution in form of two columns, of which the first specifies the numerical random variable and the second the number of observations. The Altmann Fitter expects data sets as text files with the simple structure shown in Fig. 2.1.

```
1    48
2    29
3    20
4    2
5    2
6    1
7    0
8    0
9    1
```

Fig. 2.1: Structure of a data file

The columns may be separated by one or more spaces or tabs. The lines must be separated by the DOS and Windows style 'end-of-line' codes (CRLF, or hex 0d0a), which is automatically given if you create your files under a windows operating system. If you import them from UNIX or Apple systems, a corresponding transformation may be required.

Classes with zero frequencies such as the classes 7 and 8 in Fig. 2.1 can, but need not be given. The data must, however, be given in increasing order of the x-values.

How to prepare the data

The numerical values of random variable (x) of the data set can begin with any positive number: with a 1 as in the above example, with a 0, or another value such as 4, depending on the nature of the random variable. Altmann Fitter will shift the probability distribution in an appropriate way if needed. You should, however, avoid empty classes, in particular in the beginning of the data.

Additional class as trick for more degrees of freedom

On the other hand, sometimes, an additional class at the end of the data with zero frequency can help to find a fitting distribution because this 'trick' increases the number of degrees of freedom.

The files are expected as text files with the file extension ".dat" as default. You can nevertheless load files with any extension by changing the standard file type from "*.dat" to "all files" in the file open dialog.

Load the data

The empirical data which are to be analysed are given as one or more data files. The Altmann Fitter loads your data when you click on the "open files" button (cf. Fig. 2.2), which shows an opened yellow folder and can be found on the left side of the data area.

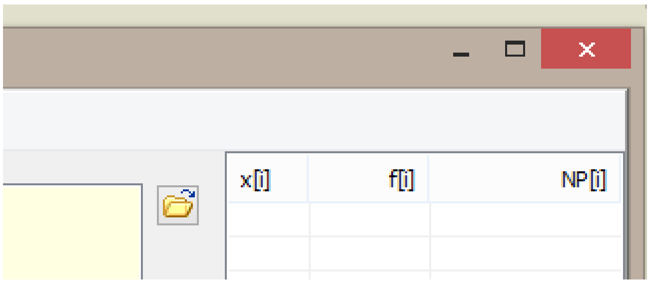


Fig. 2.2: The "open files" button

A click on this button opens the Windows "open dialog", where you can select one or more files. The selected file(s) will be shown in the data file field of the Fitter window (cf. Fig. 2.3).

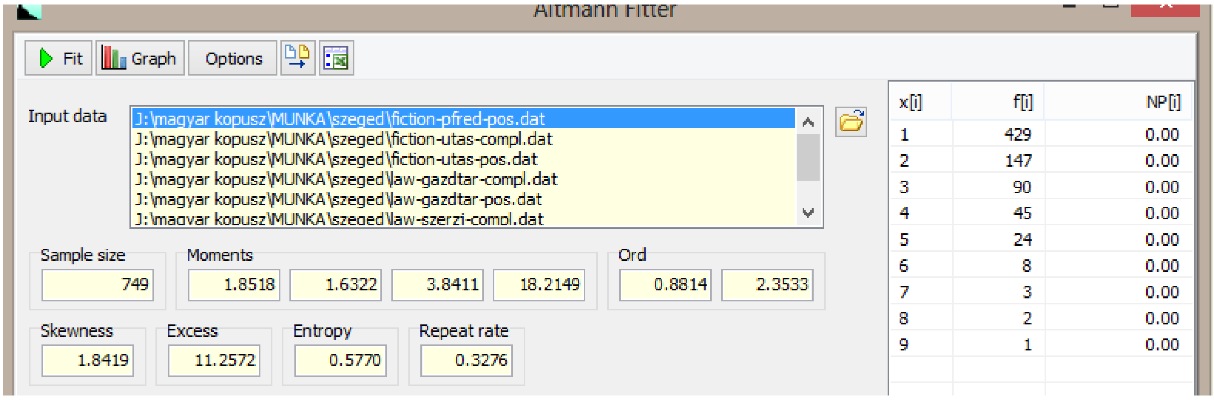


Fig. 2.3: Files are loaded

You can mark a file in this field by moving up and down with the scroll bar and by clicking on a line. The marked file is the one which will be analysed in the "selected fitting" and in the "automatic fitting" modes. In the "batch fitting" mode, all the loaded files will be analysed in one go.

Select from multiple files

On the right hand side, in the data area, the x and the corresponding frequency values in the marked file are shown. The third column (NP(i)) is 0.0 in each row because a calculation of expected frequencies has not yet been performed. The empirical characteristics of the data, however, are determined automatically:

Sample size gives the number of observations in your data set, i.e. the empirical parameter N.

The meaning of the numbers in the cells

Moments: the first four empirical moments m_1 .. m_4:

$$m_1 = \frac{1}{N}\sum_x x f_x$$

– the mean of the frequency distribution;

$$m_2 = \frac{1}{N}\sum_x (x - m_1)^2 f_x$$

– the variance of the frequency distribution;

$$m_3 = \frac{1}{N}\sum_x (x - m_1)^3 f_x$$ – the third central moment of the distribution;

$$m_4 = \frac{1}{N}\sum_x (x - m_1)^4 f_x$$ – the fourth central moment of the distribution;

Ord: Ord's criteria, which are calculated on the basis of the moments:

$$I = \frac{m_2}{m_1} \quad \text{and} \quad S = \frac{m_3}{m_2}$$

(cf. Ord, J. Keith(1972): Families of frequency distributions. London: Griffin).

Skewness and Excess are given to complete the standard set of characteristics of a distribution.

Entropy: The entropy of the frequency distribution.

Repeat rate: The well-know (logarithm-free) alternative for entropy.

You can also inspect your data graphically. By clicking on the "Graph" button in the left upper corner of the Fitter window, on the right hand side of the green "Fit" button, an extra window with a bar graph is opened (Fig. 2.4).

Make a diagram showing your data

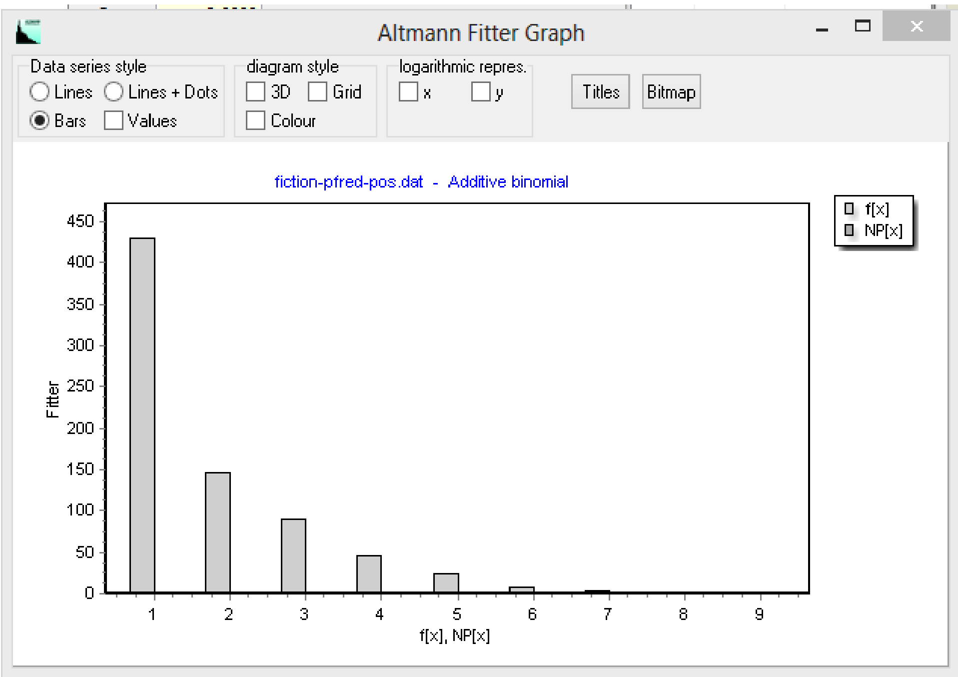


Fig 2.4: The graph window with a visualisation of the empirical data

3. The Modes of Working

There are four modes of working, one of which is recommended only for specially trained experts ("special fitting"). The modes of working are selected by clicking on one of the tabs (cf. Fig. 3.1)

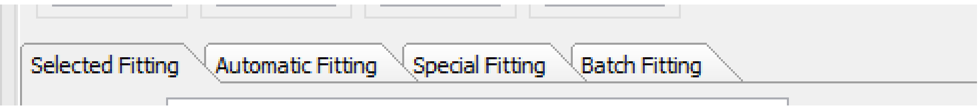


Fig.3.1: The Working Mode Tabs

3.1 Selected Fitting

Find the ditribution you want to test

This mode, the first tab from left, is most appropriate when the user has a specific hypothesis, i.e. when he/she assumes that a specific probability distribution is a good (or a theoretically justified) model. This distribution can be selected by clicking on the down arrow in the dropdown menu just below the tabs (cf. Fig.3.2).

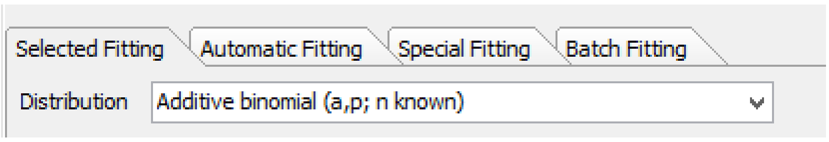


Fig 3.2.: The distribution selection menu

The click opens the menu. You can now either use the scroll bar to find the intended distribution and then click on it. Or, you can type the first character(s) of the distribution's name, and the programme will put the menu's focus at once close to the intended one.

Distributions with "known" parameters

The set of pre-defined distributions which the Altmann Fitter offers for fitting contains a number of probability distributions with "known" parameters. These distribution variants should be used when you want to constrain the estimation of parameters because your model fixes the value of one or more parameters. When you select such a distribution and start the fitting procedure you will be asked to specify the "known" parameters. The pop-up dialog will inform you about admitted ranges of these parameters.

Assisted selection of a distribution

The Altmann Fitter offers assistance for the selection of a distribution in case the user is not sure which distributions would be compatible with the data. Checking the option "Assisted selection of distribution" (Fig. 3.3) will activate an assistant which compares the properties of the data (the data must be loaded at this moment) with the available distributions and keeps only those in the menu which have a chance to fit with your data.

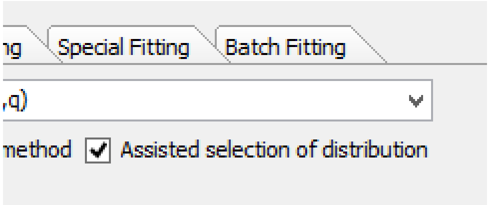


Fig. 3.3: Selection assistant option

When the input data are loaded (see section 2) and a distribution is selected the fitting procedure can be started. Before, the used should check the option (cf. Fig 3.4). Otherwise, the Fitter would present you the results of each individual go with one of its built-in initial parameter estimators. And you would have to confirm each of these steps by clicking on the "next method" button (Fig. 3.5).

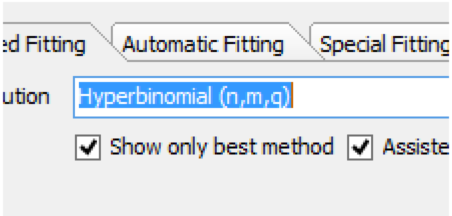


Fig 3.4.: "Best method only" option

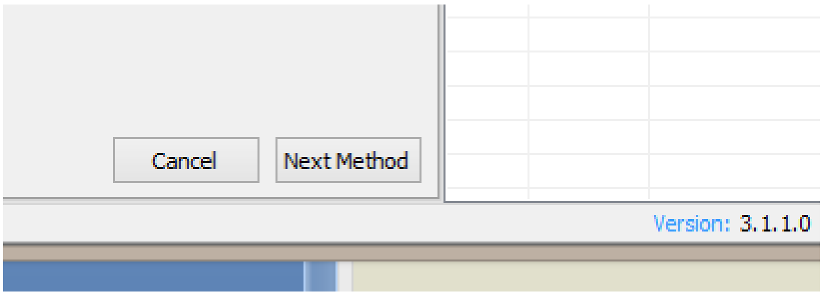


Fig 3.5: The "next method" button

Only very few "insiders", who are familiar with the details of the parameter estimators would profit from the information given in these steps.

Various estimation methods

Finally, the fitting procedure is started by clicking on the "Fit" button near the upper left corner of the Fitter's window (Fig. 3.6).

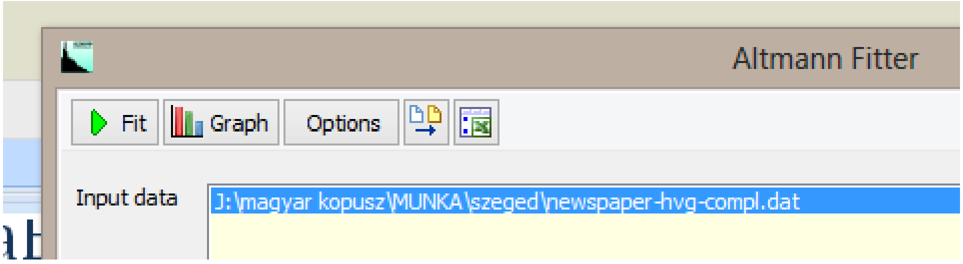


Fig. 3.6: The "Fit" button

Fitting results

The result can be an absolute failure if the selected distribution cannot be fitted to the data. there are several possible reasons for such a failure, among them the case that too few degrees of freedom are left because the data have too few classes with respect to the number of estimated parameters, the case that the distribution is simply inappropriate, or the case that the size of the sample is very large and the chi-square test becomes invalid (see below). The Altmann Fitter tells you what happened in the status field (cf. Fig. 3.7).

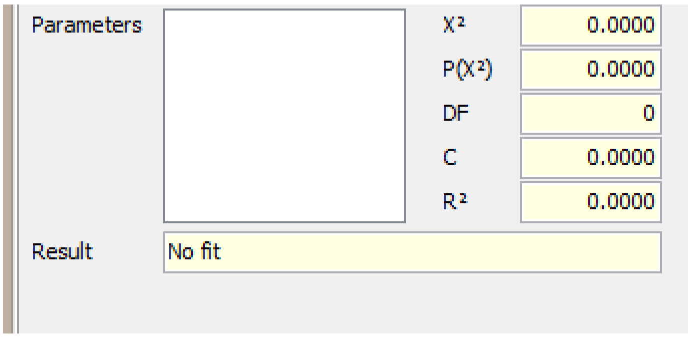


Fig. 3.7: The status field reports "No fit"

In case that no such problem occurred, the programme will show you a screen like Fig. 3.8.

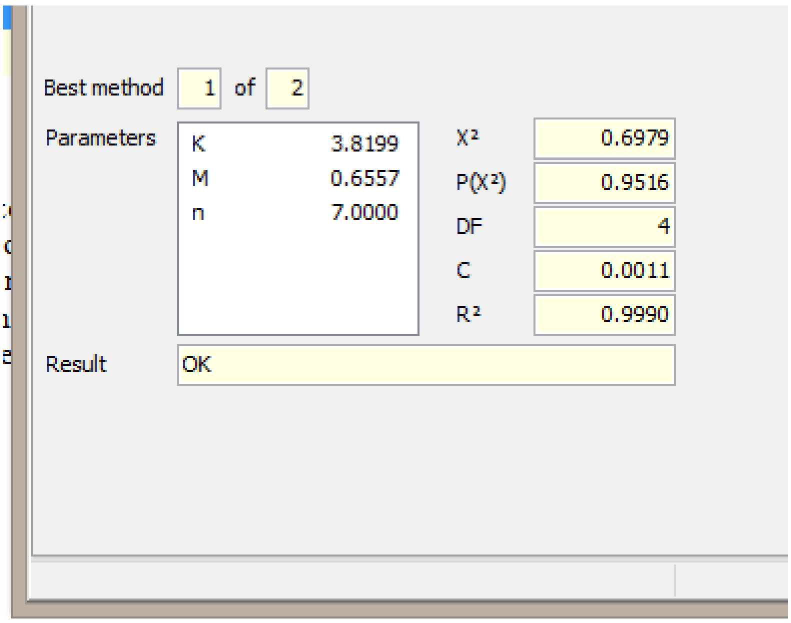


Fig. 3.8: Results of a successful fit

The parameter field gives the resulting parameters of the selected distribution (here, the negative hypergeometric distribution was applied). On the right hand side, the chi-square value, its probability (here, it is much larger than 0.05, i.e. the goodness-of-fit is evaluated as 'very good'). The test was performed with 4 degrees of freedom (the sample has n=8 classes, the distribution has three estimated parameters, whence df = 8 - 1 - 3 = 4). The Coefficient of Discrepancy C, a function of chi-square ($C=X^2/N$), can often be used instead of chi-square when a large sample makes the chi-square test invalid. Additionally, the Coefficient of Determination R^2 is given although it is defined for linear functions only. It may, nevertheless, be interesting in many cases and help to enlarge experience with this coefficient in connection with non-linear functions.

The meaning of the numbers in the results field

The fitting results can be inspected in four ways: A global evaluation is given already in the above-sketched way. Additionally, the individual frequency values can be compared to the theoretically expected values as given in the data field on the right hand side. Furthermore, all the results can be exported to a .txt or an Excel (.xls) file. You find the corresponding buttons in the top button row next to the "Options" button. Clicking on one of these two buttons will open a file save dialog. Finally a graph can be opened by clicking on the "Graph" button (Fig. 3.9).

The data field and the calculated values

The Graph window offers a rich choice of additional options. The user can choose between bar and lines graphs or select a lines plus dots representation. The numerical values of the data points can be superimposed. The contrast of the diagram can be changed by checking the "Colour" option and a grid can be shown. The diagram style can be toggled between a regular and a "3D" variant.

The diagram window and its options

Often, extremely skew distributions must be represented. In such cases, logarithmic transformations of one or both axes may help to form a clearer picture of the data and the theoretical curve. This can be done by checking one or both of the "logarithmic repres." check boxes.

The legends of the diagram can also be changed: click the "Titles" button, and a corresponding window will let you change the pre-defined Titles.

Finally, you can export your diagram in the configured form to the Windows clipboard by clicking on the "Bitmap" button.

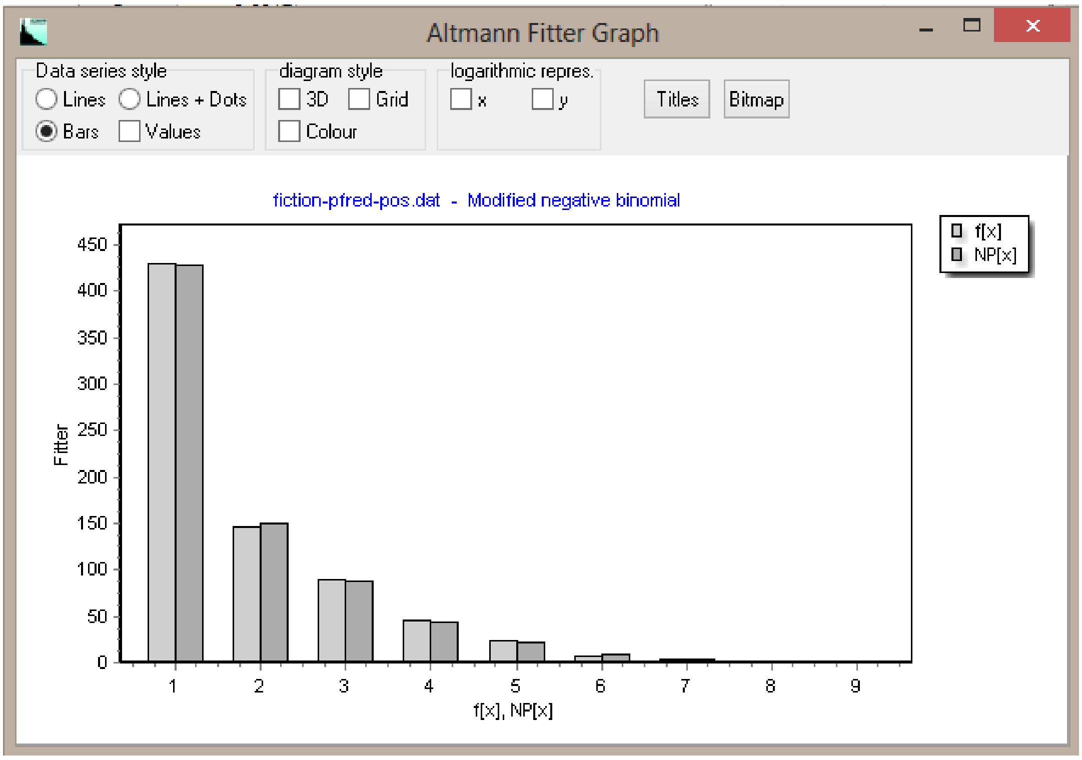


Fig. 3.9: Bar graph with empirical (left hand and light grey) and theoretical (right hand and darker) values.

3.2 Automatic Fitting

When your model is not specific enough to predict a particular distribution or when the user just wants to heuristically find out which distributions would, in principle, match with his/her data, the Automatic fitting mode is appropriate. To select this mode, click on the second tab in the tab row. It is tagged with the descriptor "Automatic fitting".

One of the loaded data sets in the top widow should be selected before the "Fit" button is clicked. The following procedure will apply to the marked data set. Altmann Fitter will run through all the more than 200 distributions in its inventory except the variants with "known" parameters. While testing the distributions, Altmann Fitter displays the individual results in the rows of the table field which appears when the "Automatic fitting" tab is clicked (Fig. 3.10). The columns of this table field contain the following information:

The results of automatic fitting

Distribution: the name of the tested distribution.

X^2: The value of the Chi-square test.

$P(X^2)$: The probability of the found X^2 value.

C: The value of the coefficient of discrepancy. $C = X^2/N$.

DF: The number of degrees of freedom. Altmann Fitter determines the DF number automatically from the number of data classes and the number of estimated parameters.

R^2: The coefficient of determination. R^2 is given although it is defined for linear functions only. It may, nevertheless, be interesting in many cases and help to enlarge experience with this coefficient in connection with non-linear functions.

Selected Fitting | Automatic Fitting | Special Fitting | Batch Fitting

Distribution	X²	P(X²)	C	DF	R²
Mixed Poisson (a,b,α)	2.65	0.4479	0.0035	3	0.9991
Thomas (a,b)	2.24	0.8144	0.0030	5	0.9991
Cernuschi-Castagnetto-Poisson (a,b)	1.81	0.9366	0.0024	6	0.9989
Gegenbauer (a,b,k)	5.95	0.3115	0.0079	5	0.9988
Polya-Aeppli (a,p)	3.33	0.7666	0.0044	6	0.9988
Positive Singh-Poisson (a,α)	4.99	0.2886	0.0067	4	0.9987
Negative hypergeometric (K,M,n)	3.92	0.4163	0.0052	4	0.9986
Polya (s,p,n)	4.10	0.3922	0.0055	4	0.9985
Non-central negative binomial (a,k,p)	4.47	0.4834	0.0060	5	0.9981
Gross-Harris-geometric I (q,α)	7.65	0.2653	0.0102	6	0.9980
Darwin (B; M = x-max)	5.27	0.3833	0.0070	5	0.9976
Consul-Mittal-binomial with 3 parameters (n,p,θ)	6.42	0.1702	0.0086	4	0.9974
Right truncated modified Zipf-Alekseev (a,b; n = x...	10.75	0.0295	0.0144	4	0.9970
Doubly truncated logarithmic (q; L = x-min,R = x-...	13.67	0.0178	0.0183	5	0.9966
Right truncated logarithmic (q; R = x-max)	13.67	0.0335	0.0183	6	0.9966
Bissinger-geometric (p)	12.79	0.0773	0.0171	7	0.9966

Fig. 3.10: The "Automatic fitting" table field

The cells in the $P(X^2)$ and C columns are coloured to indicate the goodness-of fit. A green cell indicates a very good fit, a yellow one an acceptable fit, and a red one a poor fit.

The order in which the rows of the table field are arranged can be changed by clicking on one of the column heads. When you are interested in ranking the distributions according to their $P(X^2)$ values, just click on the column head with the caption " $P(X^2)$". Another click on the same column head will reverse the order.

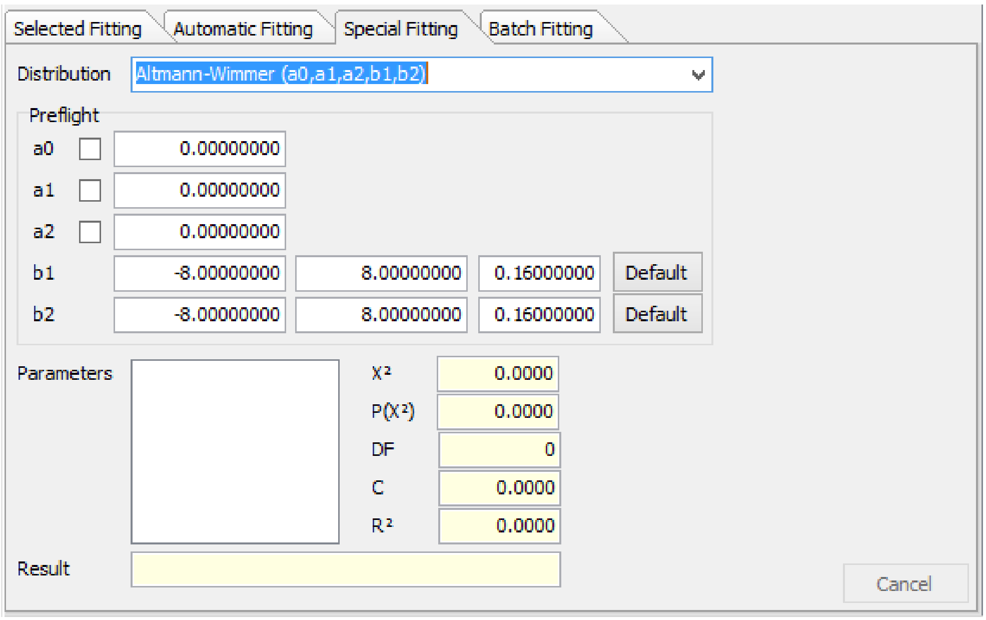
Export your data to txt or Excel files

The content of the table field can be exported to a .txt or to an Excel (.xls) file, where you can evaluate your results using other tools. You find the corresponding buttons in the top button row next to the "Options" button.

3.3 Special Fitting

This mode requires some specific knowledge and will not be described in detail. The control field, which appears when the " Special Fitting " tab is clicked is shown in Fig. 3.11.

An option for experts: the Wimmer-Altmann distribution



The control field offer the use of the Altmann-Wimmer (or Wimmer-Altmann; both name variants are common) distribution – a complex distribution which is derived from Wimmer's and Altmann's "Unified Theory" (cf. Wimmer, Gejza, and Altmann, Gabriel. 2005. „*Unified Derivation of Some Linguistic Laws*." In Quantitative Linguistik. Ein internationales Handbuch. Quantitative Linguistics. An International Handbook, ed. Köhler, Gabriel Altmann, and Rajmund G. Piotrowski, 760-775. Berlin, New York: de Gruyter, 760-775)

The distribution has five parameters, whose values and configuration can be determined such that a large number of individual distributions which are not pre-defined in the inventory of Altmann Fitter can be defined.

Before using this mode, you should be familiar with the *Unified Theory*.

3.4 Batch Fitting

Fit to multiple data sets in one go

The fourth mode, Batch fitting, is extremely useful when a distribution hypothesis is to be tested on a large number of data sets. Previous versions of Altmann Fitter did not provide this mode so that the user had to separately load each file and repeat the steps of the fitting procedure over and over again.

When the Batch mode is activated by clicking on the fourth tab in the tab row, a table field appears in the lower part of the Fitter window, which looks like the table field of the Automatic fitting mode. As opposed to that field, the user will find a horizontal scroll bar at the bottom of this field, which allows moving the window to a large number of columns right to the columns which can also be found in the Automatic fitting table field (Fig. 3.12).

Distribution: Hyperpascal (k,m,q)

k	m	q	N	x-max	NP1	m1	m2	m3	
2.2396	2.1431	0.4951	206	8	100.4629	2.0146	1.6357	3.2926	15.8
1.4429	1.0460	0.4016	206	8	101.3064	1.9126	1.3322	2.4761	11.6
2.5879	1.1335	0.2819	67	5	31.1726	1.8955	1.0488	1.2105	4.1
0.0825	0.0562	0.4926	103	9	42.0404	2.1456	2.4351	7.8748	46.3
4.0259	2.7049	0.4468	40	6	15.5877	2.3000	2.3600	3.5490	15.5
2.2069	1.8702	0.5117	191	9	80.6527	2.2670	2.6879	6.4509	35.1
0.5125	0.3591	0.4128	191	9	92.3466	1.8848	1.8087	5.3490	28.7
143.7690	7.3967	0.0298	103	9	48.8270	1.9320	1.5391	4.6733	29.5
3.2283	2.2723	0.4551	752	10	303.9564	2.2886	2.4713	5.3060	28.6
4.9255	2.8352	0.3741	655	8	270.3574	2.2015	2.2128	4.2967	20.4
12.5366	20.8999	0.7757	1328	13	693.4586	1.9337	2.2034	8.2064	54.2
0.7172	0.6735	0.5867	1328	13	520.5816	2.4752	3.6394	11.3095	75.8
0.0000	0.0000	0.0000	0.0...	0.0...	0.0000	0	0	0.0000	0.0

m2	m3	m4	Ord I	Ord S	Skewness	Excess	Entropy	Rep. rate
1.6357	3.2926	15.8304	1.5739	9.3777	0.8119	2.0130	0.6663	0.3693
1.3322	2.4761	11.6975	1.6104	7.1348	0.6965	1.8587	0.6265	0.3591
1.0488	1.2105	4.1833	1.1270	1.0849	0.5533	1.1542	0.7853	0.3837
2.4351	7.8748	46.3743	2.0723	26.7179	1.1349	3.2339	0.6637	0.3714
2.3600	3.5490	15.5087	0.9789	7.0953	1.0261	1.5038	0.8458	0.4021
2.6879	6.4509	35.1750	1.4639	18.4551	1.1856	2.4000	0.7004	0.3782
1.8087	5.3490	28.7706	2.1990	18.3926	0.9596	2.9574	0.5845	0.3359
1.5391	4.6733	29.5676	2.4476	20.8334	0.7966	3.0365	0.5798	0.3546
2.4713	5.3060	28.6990	1.3658	15.2561	1.0798	2.1471	0.6757	0.3843
2.2128	4.2967	20.4535	1.3053	10.7498	1.0051	1.9417	0.7224	0.3830
2.2034	8.2064	54.2587	2.5090	33.5526	1.1395	3.7244	0.5164	0.3301
3.6394	11.3095	75.8290	1.6289	36.7483	1.4704	3.1075	0.6468	0.3853
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Fig. 3.12: Two cuts from the Batch fitting table field.

These columns contain the complete information which is also given after fitting a distribution to an individual data set. The first columns right to the 'standard' columns, i.e. following "R^2", show the estimated parameter values and are therefore variable as to their size and to their meaning – depending on the selected distribution.

The table can be exported, too, to a .txt or to an Excel (.xls) file, where you can evaluate your results using other tools. You find the corresponding buttons in the top button row next to the "Options" button.

4 Options

The third button in the top button row in the Altmann Fitter window is the "Options" button. Clicking on it opens the menu with optimisation criteria. The six fields of this menu can be edited. The default values are as shown in Fig. 4.1.

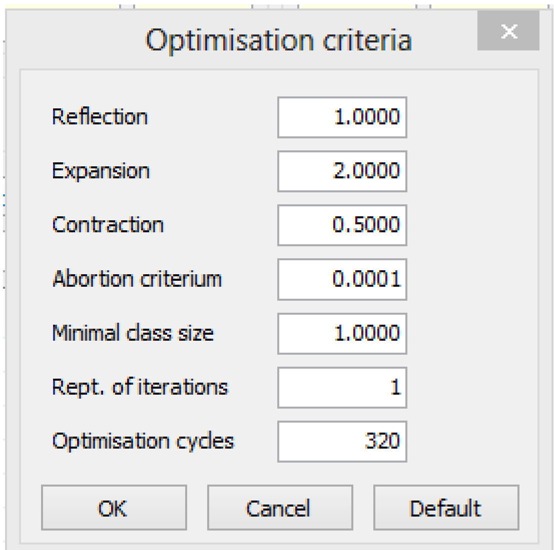


Fig. 4.1: The options menu.

Control the optimisation algorithm

The first four values are the parameters of the Nelder-Mead Simplex algorithm and control the manner in which an optimal estimation for each parameter is searched for.

Minimal class size specifies the theoretical (expected) frequency of a data class which must be reached or exceeded for the goodness-of-fit Chi-square test. Often, this value is set to 5.0 by practitioners, the absolute minimum is 1.0. The value controls the way how Altmann Fitter conducts the test. The program will always try to meet the condition by pooling adjacent classes. If the number of degrees of freedom becomes too small by reducing the number of classes for the test, no fit can be performed.

Repetitions of iterations is set to one. In some cases, a repetition of the fitting procedure may help to improve the results. If you want to do so just set the value to 2.

The number of optimisation cycles should be set at least to 20 or 30. Complicated distributions may – depending on the data under analysis – require more, even 1000 cycles. Unnecessary cycles are avoided by an appropriate abortion criterion.

The default button resets the values to the default values.

5 Register of Distributions

The detailed descriptions of the following distributions can be found in a separate documentation.

additive binomial
additive generalization of the binomial —› additive binomial
adjusted Poisson —› Gokhale-Poisson Type 1
Altham-multiplicative binomial
Arbous-Kerrich-Poisson
Beall-Rescia
Bernoulli(an) —› binomial
beta binomial —› negative hypergeometric
beta-Pascal
Bhattacharya-Holla —› Poisson-uniform binomial
binomial-beta —› negative hypergeometric
binomial-binomial
binomial-geometric
binomial-logarithmic
binomial-negative binomial
binomial-Poisson
Bissinger-binomial
Bissinger-geometric
Bissinger-negative binomial
Bissinger-Poisson
Borel
Borel-Tanner
burnt fingers—› Arbous-Kerrich-Poisson
centrally truncated Poisson
Cemuschi-Castagnetto-Poisson
Chaddha-binomial Type 1
Chaddha-binomial Type 2
Cohen
Cohen-binomial
Cohen-C-Poisson
Cohen-geometric
Cohen-negative binomial
Cohen-Poisson
conditional Poisson —› positive Poisson
confluent hypergeometric
Consul
Consul-Jain-Poisson
Consul-Mittal-binomial with 2 parameters
Consul-Mittal-binomial with 3 parameters
Conway-Maxwell-Poisson

correlated binomial —› additive binomial
Cresswell-Froggatt
Crow-Bardwell —› hyperpoisson Dacey-binomial
Dacey-negative binomia
Dacey-Poisson
Darwin
digamma
d. analogous to Borel-Tanner —› Haight-Borel-Tanner
d. of runs —› Ising-Stevens
doubly truncated binomal
doubly turncated geometric
doubly truncated logarithmic
doubly truncated negative binomial
doubly truncated Poisson
E1CB —› confluent hypergeometric
Engset —› right truncated binomial
Erlang-Poisson
extended Katz —› hyperpascal
extended logarithmic
extended positive binomial
extended positive negative binomial
extended positive Poisson
extended truncated negative binomial —› extended positive negative binomial
extended truncated Poisson —› extended positive Poisson
factorial —› Marlow
Feller-Arley
Ferreri-meta-Poisson
Ferreri-Poisson
Fisher's logarithmic —› logarithmic Fry-Crommelin
Fry-Poisson
Furry geometric
Gegenbauer
generalized geometric —› Consul
generalized Hermite —› Gupta-JainHermite
generalized inflated binomial —› Singh binomial
generalized inflated Poisson —› PandeyPoisson
generalized logarithmic series —› Jain-Consul-logarithmic
generalized negative binomial —› Jain-Consul-negative binomial
generalized non-central binomial —› Ong-Lee-negative binomial
generalized Poisson —› Cohen-Poisson; Consul-Jain-Poisson; ModatPoisson
generalized Pólya-Aeppli —› Poisson-Pascal
generalized Waring —› beta-Pascal geometric
geometric-binomial
geometric-geometric
geometric Gram-Charlier —› Shenton-Skees-geometric

geometric-logarithmic
geometric-negative binomial
geometric-Poisson
Gokhale-Poisson Type 1
Gold-PEBL
Gold-Poisson
Good
Good-Engen
Gross-Harris-geometric I
Gross-Harris-geometric II
grouped Poisson —› Morlat-Poisson
Gupta-Jain-Hermite
Haight-Borel-Tanner
Haight-harmonic
Haight-Poisson-geometric
Haight-zeta
Harris-Poisson
Hermite
Hillier-Conway-Maxwell-Poisson
Hirata-Poisson
hyperbinomial
hypergeometric
hypergeometric waiting time —› inverse hypergeometric
hyper-negative binomial —› hyperpasca1
hyperpascal
hyperpoisson
inflated binomial —› extended positive binomial
inflated generalized Poisson —› Lingappaiah-Poisson
inflated negative binomial —› modified negative binomial
inflated Poisson —› Singh-Poisson
inflated zero-truncated Poisson —› positive Singh-Poisson
inverse hypergeometric
inverse Pólya
Ising-Stevens
Jackson-Nickols Type 1
Jackson-Nickols Type 2
Jain-Consul-logarithmic
Jain-Consul-negative binomial
Jain-Poisson
Jensen
Johnson-Kotz
Katti-Sly
Kemp's binomial convolution —› Ong-Lee-negative binomial
Kendall
Lagrangian Poisson —› Consul-Jain- Poisson

Laguerre series —› non central negative binomial
left truncated binomial
left truncated logarithmic
left truncated negative binomial
left truncated Poisson
Lexis —› mixed binomial
linear function Poisson —› Jain-Poisson
Lingappaiah-Poisson
logarithmic
logarithmic negative mixture —› Shenton-Skees-logarithmic
logarithmic series —› logarithmic
log series with zeroes —› extended logarithmic
lost games —› Haight-Borel-Tanner
MacArthur
Markov —› Pólya
Markov-Pólya —› Pólya
Marlow
Miller
mixed binomial
mixed geometric
mixed geometric-logarithmic
mixed logarithmic
mixed negative binomial
mixed Poisson
mixed Poisson-binomial
mixed positive Poisson
mixture of two Poisson ds. —› mixed positive Poisson
modified beta binomial —› Morrison- Brockway
modified binomial
modified geometric
modified logarithmic —› extended logarithmic
modified negative binomial
modified Poisson —› Singh-Poisson
Morlat-Poisson
Morrison-Brockway
Morse
Naor-Poisson
negative binomial
negative binomial beta —› beta-Pascal
negative binomial-binomial
negative binomial-geometric
negative binomial-logarithmic
negative binomial-negative binomial
negative binomial-Poisson
negative binomial with excess zeroes —› extended positive negative binomial

negative hypergeometric, see also inverse hypergeometric
Neyman Type A
Neyman Type B
Neyman Type C
non central negative binomial
Ong-Lee-negative binomial
Palm
Palm-Poisson
Pandey-Poisson
Pascal —› negative binomial
Pascal beta —› beta-Pascal
Pascal-gamma —› negative binomial-logarithmic
Pascal-Poisson —› negative binomial- Poisson
PEBL
Plunkett-Jain-logarithmic
point binomial —› binomial
Poisson
Poisson-binomial
Poisson-geometric —› Pólya-Aeppli
Poisson-Lindley
Poisson-logarithmic
Poisson mixture —› mixed Poisson
Poisson-negative binomial —› Poisson-Pascal
Poisson-Pascal
Poisson-Poisson —› Neyman Type A
Poisson-reciprocal gamma
Poisson-uniform
Poisson's exponential binomial limit —› PEBL
Poisson type —› Consul-Jain-Poisson
Poisson with excess zeroes —› extended positive Poisson
Poisson with zeroes —› Singh-Poisson Pólya
Pólya-Aeppli
Pólya-Eggenberger —› Pólya
positive binomial
positive Cohen-binomial
positive Cohen-negative binomial
positive Cohen-Poisson
positive modified Poisson —› positive Cohen-Poisson
positive negative binomial
positive Pandey-Poisson
positive Poisson
positive Singh-Poisson
positive Yule
Prasad
pseudo-contagious Poisson —› Singh-Poisson

quasi-binomial —› Consul-Mittal-binomial with 2 parameters; Consul-Mittal-binomial with 3 parameters
right truncated binomial
right truncated Erlang-Poisson
right truncated geometric
right truncated logarithmic
right truncated modified Zipf-Alekseev
right truncated negative binomial
right truncated Poisson
right truncated zeta
Rutherford
Rutherford-binomial
Rutherford-Poisson
second Erlang —› Erlang-Poisson
Shenton-Skees-geometric
Shenton-Skees-logarithmic
shifted positive Poisson
shifted zero-truncated Poisson
short —› Cresswell-Froggatt
Singh-binomial
Singh-Poisson
STER-binomial —› Bissinger-binomial
STER-geometric —› Bissinger-geometric
STER-negative binomial —› Bissinger-negative binomial
STER-Poisson —› Bissinger-Poisson
Stirling Type 1
Stirling Type 2
stuttering Poisson —› Hirata-Poisson
Suzuki-binomial
Suzuki-Poisson
Swensson
synchronous counting —› Morlat-Poisson
Syski
Syski-binomial
Takács
Thomas
Toft-Boothroyd-Poisson Type 1
Toft-Boothroyd-Poisson Type 2
trigamma
truncated Poisson —› positive Poisson
Waring
Yule
zero-truncated negative binomial —› positive negative binomial
zeta
Zipf-Mandelbrot

6 Supplement

New in Altmann Fitter Version 3.x

binomial-arc-sine
Chung-Feller
Dacey3
discrete Tuldava
discrete Zipf
Estoup
Harris
Kelly
Kemp2
Macutek-geometric
Martin-Lof-Sverdrup reversed
right truncated Good
right truncated Kemp2
right truncated Salvia-Bolinger
right truncated Waring
right truncated Yule
Salvia-Bolinger
Wimmer-Altmann family 1

6 Formulae and Details of the New Distributions

Notation

$\mathbb{N}$ = set of natural numbers = {1,2,3,...}

$\mathbb{R}$ =set of real numbers

Factorial expressions: $\begin{cases} x_{(k)} = x(x-1)(x-2)...(x-k+1) \\ x^{(k)} = x(x+1)(x+2)...(x+k-1) \end{cases}$

Probability generating function

$$G(t) = \sum_{x=0}^{\infty} P_x t^x$$

Hypergeometric functions

$$_1F_1(a;b;t) = \sum_{j=0}^{\infty} \frac{a^{(j)}}{b^{(j)}} \frac{t^j}{t!} = 1 + \frac{a}{b}\frac{t}{1} + \frac{a(a+1)}{b(b+1)}\frac{t^2}{2!} + ...$$

$$_2F_1(a,b;c;t) = \sum_{j=0}^{\infty} \frac{a^{(j)} b^{(j)}}{c^{(j)}} \frac{t^j}{t!} = 1 + \frac{ab}{c}\frac{t}{1} + \frac{a(a+1)b(b+1)}{c(c+1)}\frac{t^2}{2!} + ...$$

Gamma function

$$\Gamma(a) = \int_0^{\infty} x^{a-1} e^{-x} dx$$

Digamma function

$$\Psi(z) = \frac{\Gamma'(z)}{\Gamma(z)} = \frac{\partial \ln \Gamma(z)}{\partial z}$$

Trigamma function

$$\Psi'(z) = \frac{\partial^2 \ln \Gamma(z)}{\partial z^2}$$

Beta function

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx$$

Beta distribution

$$f(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}, \qquad 0 < x < 1$$

Laguerre polynomial

$$L_r^a(x) = \sum_{j=0}^{r} (-1)^j \binom{r+a}{r-j} \frac{x^j}{j!}$$

Stirling numbers of the first kind: S(n,x)

$$S(n+1,x) = S(n,x-1) - nS(n,x)$$

Stirling numbers of the second kind: Z(n,x)

$$Z(n+1,x) = xZ(n,x) + Z(n,x-1)$$

Modified Bessel function of the first kind

$$I_k(z) = \left(\frac{z}{2}\right)^k \sum_{j=0}^{\infty} \frac{\left(\frac{z}{2}\right)^{2j}}{j!\,\Gamma(k+j+1)}$$

Modified Bessel function of the second kind

$$K_k(z) = \frac{\pi}{2} \frac{I_{-k}(z) - I_k(z)}{\sin k\pi}$$

binomial – arc-sine distribution

$$P_x = \binom{\alpha+n-x-1}{n-x}\binom{x-\alpha}{x}, \qquad x = 0,1,2,\ldots,n$$

Parameter limits:

n is a positive integer or zero

$0 \le \alpha \le 1$

Recurrence formula

$$P_x = \frac{(x-\alpha)(n-x+1)}{x(n-x+\alpha)} P_{x-1}$$

$$P_0 = \binom{n+\alpha-1}{n}$$

Chung-Feller distribution

$$P_x = \binom{2n-2x}{n-x}\binom{2x}{x}2^{-2n}, \quad x = 0,1,\ldots,n$$

Parameter limits:

n is a positive integer or zero

Recurrence formula

$$P_x = \frac{\left(x-\frac{1}{2}\right)(n-x+1)}{x\left(n-x+\frac{1}{2}\right)} P_{x-1}$$

$$P_0 = \binom{2n}{n}2^{-2n}$$

Dacey 3 distribution

$$P_x = \frac{\binom{m+n-x-2}{n-x}}{\binom{m+n-1}{n}}, \quad x = 0,1,2,\ldots,n$$

Parameter limits:

n is a positive integer or zero

$m \geq 1$

Recurrence formula

$$P_x = \frac{n-x+1}{m+n-x-1} P_{x-1}$$

$$P_0 = \frac{m-1}{m+n-1}$$

---

discrete Tuldava (= generalized Whitworth) distribution

$$P_x = c\left(a + b\sum_{j=x}^{n} \frac{1}{j}\right), \quad x = 1,2,\ldots,n$$

Parameter limits:

$b > 0$

$a \geq -\frac{b}{n}$

n is a positive integer

Normalization constant

$$c = \frac{1}{n(a+b)}$$

Recurrence formula

$$P_x = P_{x-1} - \frac{b}{n(a+b)(x-1)}, \quad x = 2,3,\dots,n$$

$$P_1 = \frac{1}{n(a+b)}\left(a + b\sum_{j=x}^{n}\frac{1}{j}\right)$$

---

discrete Zipf distribution

$$P_x = \frac{\binom{x-b-1}{x-1}}{\binom{n-b}{n-1}}, \quad x = 1,2,\dots n$$

Parameter limits:

$b < 1$

n is a positive integer

Recurrence formula

$$P_x = \left(1 - \frac{b}{x-1}\right)P_{x-1}$$

$$P_1 = \frac{1}{\binom{n-b}{n-1}}$$

Estoup distribution

$$P_x = \frac{c}{x}, \; x = 1, 2, \ldots, n$$

Parameter limits:

n is a positive integer

Normalization constant

$$c = \frac{1}{\Psi(n+1)+\gamma}, \text{ hence } P_x = \frac{1}{x\left(\Psi(n+1)+\gamma\right)}$$

Recurrence formula

$$P_x = \left(1-\frac{1}{x}\right)P_{x-1}, \quad P_1 = \frac{1}{\Psi(n+1)+\gamma}$$

Harris distribution

$$P_x = \frac{N}{N+n}\frac{\binom{n}{x}}{\binom{N+n-1}{x}}, \quad x=0,1,\ldots,n$$

Parameter limits:

n is a positive integer or zero

$N > 0$

Recurrence formula

$$P_x = \frac{n-x+1}{N+n-x}P_{x-1}$$

$$P_0 = \frac{N}{N+n}$$

Kelly distribution

$$P_x = \frac{\frac{\theta}{n}\binom{n}{x}}{\binom{n+\theta-1}{x}}, \qquad x = 1, 2, \ldots, n$$

Parameter limits:

n is a positive integer

$\theta > 0$

Recurrence formula

$$P_x = \frac{n-x+1}{n-x+\theta} P_{x-1}$$
$$P_1 = \frac{\theta}{n+\theta-1}$$

Kemp 2 distribution

$$P_x = \frac{a}{a+2x}\binom{a+2x}{x}\left(\frac{1}{2}\right)^{a+2x}, \quad x=0,1,2,\ldots$$

Parameter limits:

$a>0$

Recurrence formula

$$P_0 = \left(\frac{1}{2}\right)^a, \quad P_x = \frac{(a+2x-1)(a+2x-2)}{4x(a+x)} P_{x-1}, \quad x = 1, 2, \ldots$$

Mačutek-geometric distribution

$$P_x = cp^{x-1}\left(1+\frac{a}{n-x+1}\right), \quad x = 1,2,\dots,n$$

Normalization constant

$$c = \left(\frac{1-p^n}{1-p} + ap^{n-1}\left(\Phi\left(\frac{1}{p},1,1\right) - \frac{1}{p^{n+1}}\Phi\left(\frac{1}{p},1,n+1\right)\right)\right)^{-1}$$

Parameter limits:

n is a positive integer

$a \geq -1$

p>0

Recurrence formula

$$P_x = \left(p - \frac{\frac{ap}{a+1}}{x-n-1} + \frac{\frac{ap}{a+1}}{x-n-a-2}\right)P_{x-1} \text{ for } a \neq -1$$

$$P_x = \left(p - \frac{p}{(x-n-1)^2}\right)P_{x-1} \text{ for } a = -1$$

Martin-Löf -Sverdrup distribution (reversed form)

$$P_x = \frac{\binom{n-x}{n-m}}{\binom{n+1}{n-m+1}}, \qquad x = 0,1,\dots,m$$

Parameter limits:

n is a positive integer or zero

m is a positive integer or zero

$n \geq m$

Recurrence formula

$$P_0 = \frac{n-m+1}{n+1}$$

$$P_x = \frac{m-x+1}{n-x+1} P_{x-1}$$

---

right truncated Good distribution

$$P_x = \frac{cp^x}{x^a}, \quad x = 1, 2, \ldots, n$$

Normalization constant

$$c = \frac{1}{p\left(\Phi\left(p, a, 1\right) - p^n \Phi\left(p, a, n+1\right)\right)}$$

Parameter limits:

p>0

a - no constraints

n is a positive integer

Recurrence formula

$$P_x = p\left(1 - \frac{1}{x}\right)^a P_{x-1}$$

$$P_1 = cp$$

---

right truncated Kemp 2 distribution

$$P_x = c\frac{a}{a+2x}\binom{a+2x}{x}\left(\frac{1}{2}\right)^{a+2x}, \quad x=0,1,2,\ldots,n$$

Parameter limits:

$a>0$

n is a positive integer or zero

c - normalization constant

Recurrence formula

$$P_0 = c\left(\frac{1}{2}\right)^a, \quad P_x = \frac{(a+2x-1)(a+2x-2)}{4x(a+x)}P_{x-1}, \quad x = 1, 2, ..., n$$

---

right truncated Salvia-Bolinger distribution

$$P_x = \begin{cases} c\alpha & x = 0 \\ c\frac{\alpha(1-\alpha)\ldots(x-\alpha)}{(x+1)!} & x = 1, 2, \ldots, n \end{cases}$$

Parameter limits:

$0 < \alpha \le 1$

Recurrence formula

$$P_x = \frac{x-\alpha}{x+1}P_{x-1}$$
$$P_0 = c\alpha$$

---

right truncated Waring distribution

$$P_x = c\frac{a^{(x)}}{(a+b+1)^{(x)}}, \quad x = 0,1,2,\ldots,n$$

Parameter limits:

a>=0

b>0

n is a positive integer or zero

c - normalization constant

Recurrence formula

$$P_x = \frac{a+x-1}{a+b+x}P_{x-1}$$

$$P_1 = c$$

---

right truncated Yule distribution

$$P_x = c\frac{bx!}{(b+1)^{(x+1)}}, \quad x = 0,1,2,\ldots,n$$

Parameter limits:

b>0

n is a positive integer or zero

c - normalization constant

Recurrence formula

$$P_x = \frac{x}{b+x+1}P_{x-1}$$

$$P_1 = cb$$

Salvia-Bolinger distribution

$$P_x = \alpha \qquad x = 0$$

$$P_x = \frac{\alpha(1-\alpha)\ldots(x-\alpha)}{(x+1)!} \qquad x = 1, 2, 3, \ldots$$

Parameter limits:

$0 < \alpha \le 1$

Recurrence formula

$$P_x = \frac{x-\alpha}{x+1} P_{x-1}$$

$$P_0 = \alpha$$