



Tutorials and worked examples for simulation,
curve fitting, statistical analysis, and plotting.
<http://www.simfit.org.uk>

The trinomial distribution is encountered in the analysis of trichotomous data, where N observations are made in a situation where there are only three possible disjoint categories, say x , y , or z . Data triples of counts in categories x , y , z can be any partitions, such as the number of male, female or dead hatchlings from a batch of eggs where it is hoped to determine a shift from equi-probable sexes.

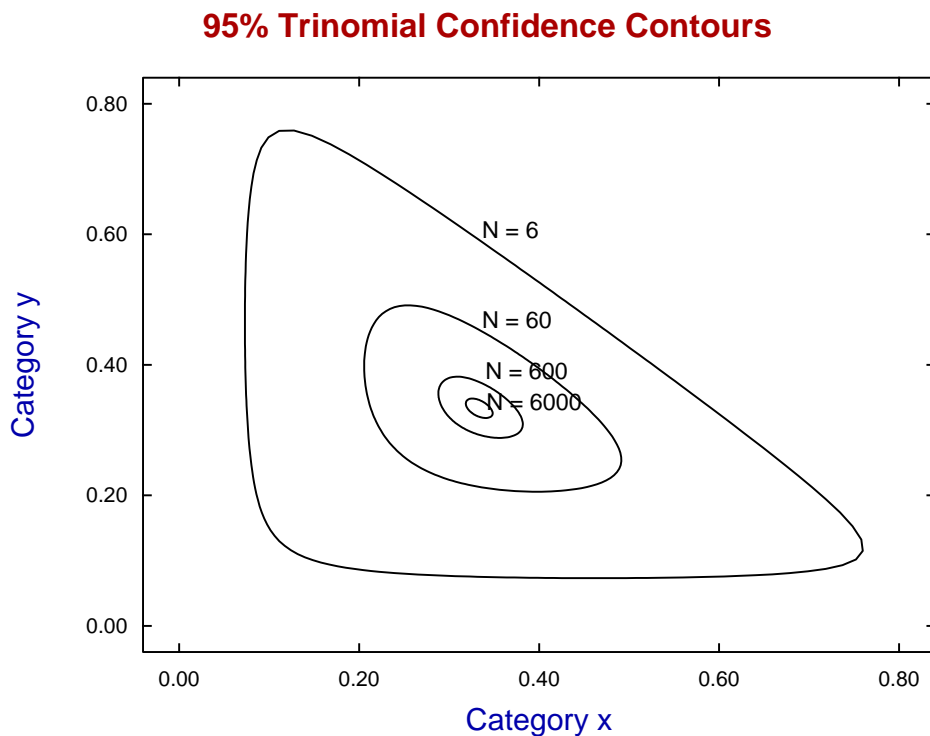
From the main SIMFIT menu select [A/Z], open program **binomial**, choose to plot trinomial confidence limits, then analyze the default data set `trinom.tf1` which has the following format.

N_x	N_y	N
2	2	6
20	20	60
200	200	600
2000	2000	6000

The format for trinomial analysis must be as in this data set as now summarized.

- Column 1: $N_x > 0 \dots$ The number of times category x was observed
- Column 2: $N_y > 0 \dots$ The number of times category y was observed
- Column 3: $N = N_x + N_y + N_z \dots$ The total number of observations

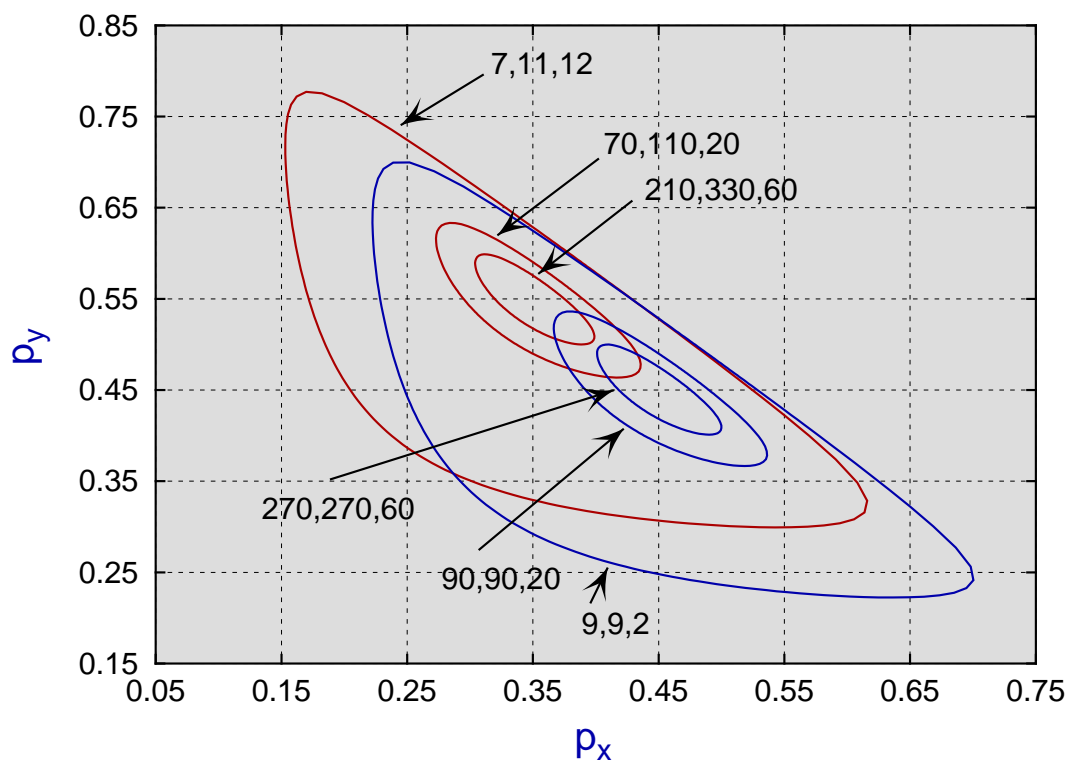
Clearly, the values in such a data matrix must all be non-negative integers subject to the constraint that column 1 plus column 2 cannot exceed column 3 in any row. After performing a chi-square test the following plot is displayed which powerfully demonstrates the contraction of the confidence regions as the sample size increases



The main value of this graphical technique is to examine confidence regions for overlap in order to better interpret the results from the chi-square test that SIMFIT always performs on such data matrices. Also, the plot can provide insight into the comparison of several different experiments using the following argument.

A useful rule of thumb to see if parameter estimates differ significantly is to check their approximate central 95% confidence regions. If the regions are disjoint it indicates that the parameters differ significantly and, in fact, parameters can differ significantly even with limited overlap. If two or more parameters are estimated, it is valuable to inspect the joint confidence regions defined by the estimated covariance matrix and appropriate chi-square critical value. Consider, for example, this figure generated by the contour plotting function of program **binomial**.

Trinomial Parameter 95% Confidence Contours



The contours are defined by

$$((\hat{p}_x - p_x), (\hat{p}_y - p_y)) \begin{bmatrix} p_x(1-p_x)/N & -p_x p_y / N \\ -p_x p_y / N & p_y(1-p_y)/N \end{bmatrix}^{-1} \begin{pmatrix} \hat{p}_x - p_x \\ \hat{p}_y - p_y \end{pmatrix} = \chi_{2;0.05}^2$$

where

$$\begin{aligned} N &= N_x + N_y + N_z \\ \hat{p}_x &= N_x / N \\ \text{and } \hat{p}_y &= N_y / N \end{aligned}$$

When $N = 20$ the triples 9,9,2 and 7,11,2 cannot be distinguished, but when $N = 200$ the orbits are becoming elliptical and converging to asymptotic values. By the time $N = 600$ the triples 210,330,60 and 270,270,60 can be seen to differ significantly.

Theory

If, in a trinomial distribution, the probability of category i is p_i for $i = 1, 2, 3$, then the probability P of observing n_i in category i in a sample of size $N = n_1 + n_2 + n_3$ from a homogeneous population is given by

$$P = \frac{N!}{n_1!n_2!n_3!} p_1^{n_1} p_2^{n_2} p_3^{n_3}$$

and the maximum likelihood estimates, of which only two are independent, are

$$\begin{aligned} \hat{p}_1 &= n_1/N, \\ \hat{p}_2 &= n_2/N, \\ \text{and } \hat{p}_3 &= 1 - \hat{p}_1 - \hat{p}_2. \end{aligned}$$

The bivariate estimator is approximately normally distributed, when N is large, so that

$$\begin{bmatrix} \hat{p}_1 \\ \hat{p}_2 \end{bmatrix} \sim MN_2 \left(\begin{bmatrix} p_1 \\ p_2 \end{bmatrix}, \begin{bmatrix} p_1(1-p_1)/N & -p_1p_2/N \\ -p_1p_2/N & p_2(1-p_2)/N \end{bmatrix} \right)$$

where MN_2 signifies the bivariate normal distribution. Consequently

$$((\hat{p}_1 - p_1), (\hat{p}_2 - p_2)) \begin{bmatrix} p_1(1-p_1)/N & -p_1p_2/N \\ -p_1p_2/N & p_2(1-p_2)/N \end{bmatrix}^{-1} \begin{pmatrix} \hat{p}_1 - p_1 \\ \hat{p}_2 - p_2 \end{pmatrix} \sim \chi_2^2$$

and hence, with probability 95%,

$$\frac{(\hat{p}_1 - p_1)^2}{p_1(1-p_1)} + \frac{(\hat{p}_2 - p_2)^2}{p_2(1-p_2)} + \frac{2(\hat{p}_1 - p_1)(\hat{p}_2 - p_2)}{(1-p_1)(1-p_2)} \leq \frac{(1-p_1-p_2)}{N(1-p_1)(1-p_2)} \chi_{2;0.05}^2.$$

Such inequalities define regions in the (p_1, p_2) parameter space which can be examined for statistically significant differences between $p_{i(j)}$ in samples from populations subjected to treatment j .

Hence, where regions are clearly disjoint for groups treated differently or for different samples, it can be concluded that parameters have been significantly affected by the treatments, as illustrated previously.