



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

Just as the normal distribution plays an important part in the analysis of measurements, the multinomial distribution is central to the analysis of counts, e.g. frequencies, contingency tables, etc. For instance, suppose an experiment leads to just one of m possible and exclusive events E_1, E_2, \dots, E_m each with fixed probability $p_i \geq 0$, then the probability that in n successive experiments there will be x_i events E_i , for $i = 1, 2, \dots, m$ is

$$f(x_1, x_2, \dots, x_m) = \frac{n!}{x_1! x_2! \dots x_m!} p_1^{x_1} p_2^{x_2} \dots p_m^{x_m}$$

where

$$p_1 + p_2 + \dots + p_m = 1.$$

1 Definitions

A special case is the binomial distribution where each experiment has only one of two possible outcomes, e.g. heads or tails, success or failure, positive or negative, black or white, improvement or deterioration, and where each experiment is independent of all previous experiments. A succession of N such Bernoulli trials where the probability of success is p , so that the possibility of failure is $1 - p$, can be described by a random variable X , which is just the number of successes in the N trials. Then the probability of k successes being recorded $P(X = k)$ is

$$f(k) = \binom{N}{k} p^k (1 - p)^{N-k}$$

where the binomial coefficient, i.e. the number of ways of selecting k items from N , is

$$\binom{N}{k} = \frac{N!}{k!(N - k)!},$$

and the expected value and variance of such a binomial variable are

$$E(X) = Np$$
$$V(X) = Np(1 - p).$$

2 Simfit program binomial

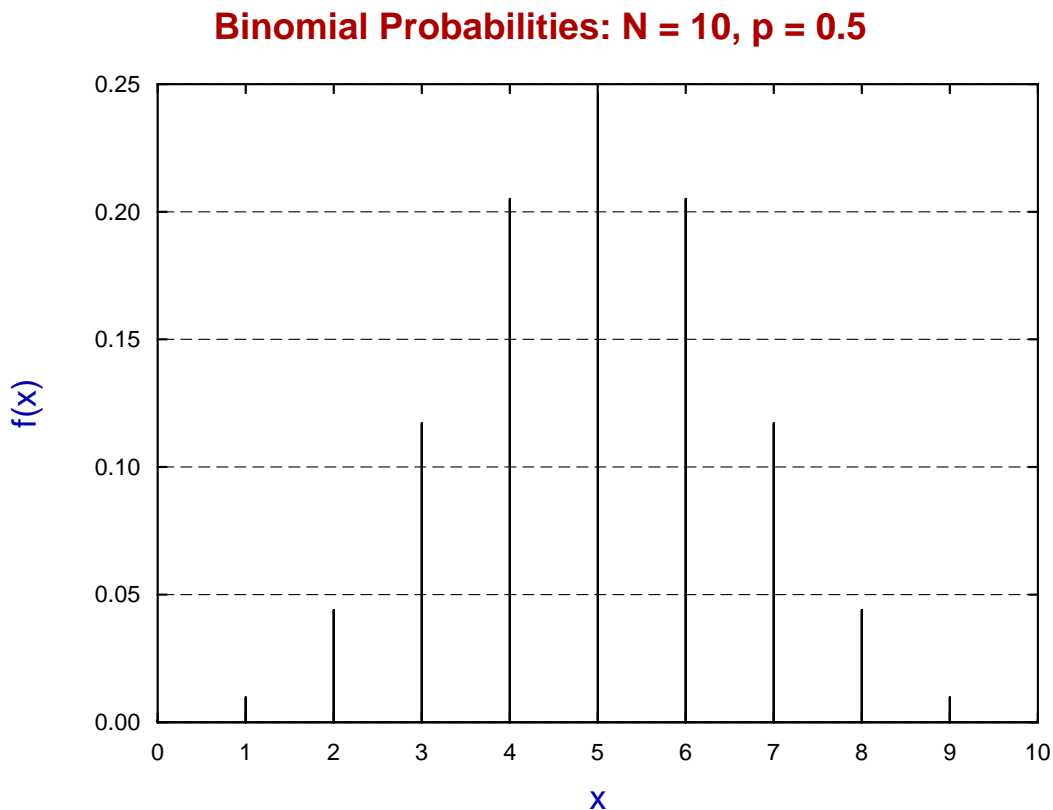
Choose [A / Z] from the main SIMFIT menu and open program **binomial** when the following options will be available.

```

Input: b(N,p), and P(lambda) parameters
Input: binomial x ... calculate pmf(x)
Input: binomial x ... calculate cdf(x)
Input: binomial % ... calculate x-critical
Input: binomial N,x, calculate NCX(x)
Input: binomial N,x, estimate p, con. lim.
Input: a sample, test if distributed b(N,p)
Input: binomial x,N,t, analysis of proportions
Input: trinomial x,y,N, plot conf. reg.
Input: Poisson x ... calculate pmf(x)
Input: Poisson x ... calculate cdf(x)
Input: Poisson % ... calculate x-critical
Input: Poisson x, estimate lambda and con.lim.
Input: a sample, test if distributed P(lambda)
Calculate: power and sample size
Calculate: change confidence limits (now ,i3,%)
Calculate: using the non-central beta distribution

```

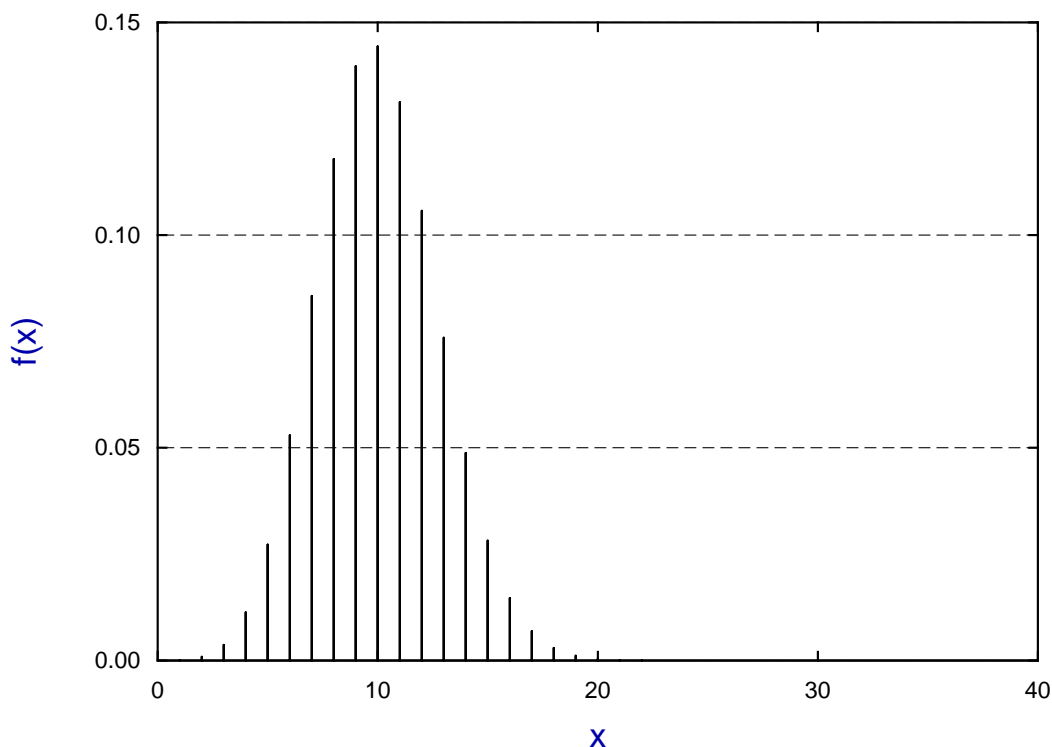
Choosing the first option allows you to change the binomial parameters as required for the subsequent binomial options. For instance, the plot with default parameters $N = 10$ and $p = 0.5$ is as follows



which demonstrates the symmetry when $p = 0.5$.

Changing the p value away from 0.5 results in skewing the probabilities and removing the symmetry as will be seen from the next plot.

Binomial Probabilities: $N = 40, p = 0.25$



3 Critical values

It should be noted that calculating critical points for a discrete distribution, like the binomial distribution, is problematical as the cumulative distribution is a step function, and so only ranges can be given, as in the following example for the 99%, 95%, 90%, 75%, and 50% points.

```

Current binomial parameters: N = 100, p = 0.75
P(X =< 64) = 0.00941      P(X =< 65) = 0.01643
P(X > 64) = 0.99059 *** P(X > 65) = 0.98357 (99%, 64 or 65)
P(X =< 67) = 0.04460      P(X =< 68) = 0.06935
P(X > 67) = 0.95540 *** P(X > 68) = 0.93065 (95%, 67 or 68)
P(X =< 68) = 0.06935      P(X =< 69) = 0.10379
P(X > 68) = 0.93065 *** P(X > 69) = 0.89621 (90%, 68 or 69)
P(X =< 71) = 0.20754      P(X =< 72) = 0.27762
P(X > 71) = 0.79246 *** P(X > 72) = 0.72238 (75%, 71 or 72)
P(X =< 74) = 0.44653      P(X =< 75) = 0.53833
P(X > 74) = 0.55347 *** P(X > 75) = 0.46167 (50%, 74 or 75)

```

4 Binomial parameter confidence limits

If there are X successes in N trials then the best estimate for the population parameter is the sample estimate \hat{p} given by

$$\hat{p} = X/N.$$

However, to get some idea of the reliability of such estimates it is necessary to calculate confidence limits, and it often surprises experimentalists that these limits are not symmetrical, as illustrated by the next set of results.

Binomial p and 95% limits given X successes in N trials.

$N = 10$ $X = 5$:Lower 95% = 0.18709, $p = 0.5$, Upper 95% = 0.81291

$N = 100$ $X = 50$:Lower 95% = 0.39832, $p = 0.5$, Upper 95% = 0.60168

$N = 1000$ $X = 500$:Lower 95% = 0.46855, $p = 0.5$, Upper 95% = 0.53145

$N = 10$ $X = 3$:Lower 95% = 0.06674, $p = 0.3$, Upper 95% = 0.65245

$N = 100$ $X = 30$:Lower 95% = 0.21241, $p = 0.3$, Upper 95% = 0.39981

$N = 1000$ $X = 300$:Lower 95% = 0.27172, $p = 0.3$, Upper 95% = 0.32946

For large N and p close to 0.5 limits are approximately central, but the asymmetry increases and the confidence range becomes increasingly skewed as p moves away from 0.5 since, of course, confidence limits must remain between 0 and 1. When comparing sequential estimates for a binomial parameter, say as a function of some variable t , it is useful to plot \hat{p} values with individual confidence limits along with the overall estimate from the combined sample to detect trends, as shown next.

p-estimated as a function of t

