

The Effects of an Arcsin Square Root Transform on a Binomial Distributed Quantity.

P.A. Bromiley and N.A. Thacker

Last updated
13 / 6 / 2002

This document forms part of the **Statistics and Segmentation Series (2008-001)** available from www.tina-vision.net.

- 2007-008 Tutorial: Defining Probability for Science.
- 2001-007 Performance Characterisation in Computer Vision:
The Role of Statistics in Testing and Design.
- 2002-007 The Effects of an Arcsin Square Root Transform on a Binomial Distributed Quantity.
- 2001-010 The Effects of a Square Root Transform on a Poisson Distributed Quantity.
- 2004-004 Shannon Entropy, Renyi Entropy, and Information.
- 2002-002 Validating MRI Field Homogeneity Correction Using Image Information Measures.
- 2004-001 Empirical Validation of Covariance Estimates for Mutual Information Coregistration.
- 2004-005 The Equal Variance Domain: Issues Surrounding the Use of Probability Densities in Algorithm Design.
- 2009-008 Avoiding Zero and Infinity in Sample Based Algorithms.
- 2001-008 Derivation of the Renormalisation Formula for the Product of Uniform Probability Distributions and Extension to Non-Integer Dimensionality.
- 2001-005 Model Selection and Convergence of the EM Algorithm.
- 2003-007 Noise Filtering and Testing for MR Using a Multi-Dimensional Partial Volume Model.
- 2002-004 A Novel Method for Non-Parametric Image Subtraction:
Identification of Enhancing Lesions in Multiple Sclerosis from MR Images.
- 2001-014 Bayesian and Non-Bayesian Probabilistic Models for Image Analysis.
- 1997-001 The Bhattacharyya Metric as an Absolute Similarity Measure for Frequency Coded Data.
- 1999-001 The Bhattacharyya Measure requires no Bias Correction.
- 1999-004 B-Fitting: An Estimation Technique With Automatic Parameter Selection.
- 2005-008 Tutorial: Beyond Likelihood.



Imaging Science and Biomedical Engineering Division,
Medical School, University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT.

The Effects of an Arcsin Square Root Transform on a Binomial Distributed Quantity

P.A. Bromiley and N.A. Thacker
Imaging Science and Biomedical Engineering Division
Medical School, University of Manchester
Manchester, M13 9PT, UK
paul.bromiley@man.ac.uk

Abstract

This document provides proofs of the following:

- The binomial distribution can be approximated with a Gaussian distribution at large values of N .
- The arcsin square-root transform is the variance stabilising transform for the binomial distribution.
- The Gaussian approximation for the binomial distribution is more accurate at smaller values of N after the variance stabilising transform has been applied.

The conclusion contains some comments concerning the relationship between the variance stabilising transform and the improved accuracy of the Gaussian approximation, which holds for both the binomial and Poisson distributions.

1 Gaussian Approximation to the Binomial Distribution

The binomial distribution

$$P(n|N) = \frac{N!}{n!(N-n)!} p^n q^{(N-n)} \quad (1)$$

gives the probability of obtaining n successes out of N Bernoulli trials, where p is the probability of success and $q = 1 - p$ is the probability of failure. The shape of the distribution approaches that of a Gaussian distribution at large N as a consequence of the central limit theorem. In order to demonstrate this, it is necessary to expand the distribution as a Taylor series around the maximum.

First, let n' be the position of the maximum of $P(n)$, giving $n = n' + \eta$, and rather than expanding the distribution itself, expand the natural logarithm of the distribution. Expanding as a Taylor series

$$f(a+h) = f(a) + hf'(a) + \frac{h^2}{2!} f''(a) + \dots + \frac{h^{(n-1)}}{(n-1)!} f^{(n-1)}(a) + \dots \quad (2)$$

gives

$$\ln[P(n'+\eta)] = \ln[P(n')] + \eta B_1 + \frac{\eta^2}{2!} B_2 + \frac{\eta^3}{3!} B_3 \dots \quad (3)$$

where

$$B_k = \left. \frac{d^k \ln[P(n)]}{dn^k} \right|_{n=n'} \quad (4)$$

Since this is an expansion around the position of the maximum, $B_1 = 0$ and B_2 is negative, so put $B_2 = -|B_2|$.

The next step is to find the position of the maximum. Taking the log of the distribution and applying Stirling's approximation

$$\ln(n!) \approx n \ln n - n \quad (5)$$

gives

$$\begin{aligned} \frac{d \ln[P(n)]}{dn} &\approx \frac{d}{dn} [\ln N! - n \ln n + n - (N-n) \ln(N-n) + (N-n) + n \ln p + (N-n) \ln q] \\ &= -\ln n + \ln(N-n) + \ln p + \ln q \end{aligned} \quad (6)$$

At the maximum this is equal to zero, giving

$$0 = \ln\left[\frac{N-n}{n} \frac{p}{q}\right] \quad (7)$$

or

$$1 = \frac{N-n}{n} \frac{p}{q} \quad (8)$$

Substituting $n = n'$

$$1 = \frac{N-n'}{n'} \frac{p}{q} \quad (9)$$

which reduces to

$$n' = Np \quad (10)$$

Now the B_k of the Taylor expansion can be found. We know that $B_1 = 0$. B_2 is given by

$$B_2 = \left| \frac{d^2 \ln[P(n)]}{dn^2} \right|_{n=n'} \approx \left[-\frac{1}{n} - \frac{1}{N-n} \right]_{n=n'} = -\frac{1}{Np(1-p)} \quad (11)$$

Similarly

$$B_3 = \left| \frac{d^3 \ln[P(n)]}{dn^3} \right|_{n=n'} \approx \frac{1-2p}{N^2 p^2 (1-p)^2} \quad (12)$$

and

$$B_4 = \left| \frac{d^4 \ln[P(n)]}{dn^4} \right|_{n=n'} \approx \frac{2(3p^2 - 3p + 1)}{N^3 p^3 (1-p)^3} \quad (13)$$

The terms are getting smaller by $O(1/N)$ each time, so truncate the Taylor series at the B_2 term:

$$P(n) \approx P(n') e^{-|B_2| \frac{n^2}{2}} \quad (14)$$

In order to find $P(n')$, impose the requirement of normalisation. Further, assume that the distribution can be treated as continuous i.e.

$$\lim_{N \rightarrow \infty} \sum_{n=0}^N P(n) \approx \int P(n) = \int_{-\infty}^{\infty} P(n' + \eta) d\eta = 1 \quad (15)$$

so, using the standard Gaussian integral

$$\int_{-\infty}^{\infty} P(n') e^{-|B_2| \frac{\eta^2}{2}} d\eta = P(n') \sqrt{\frac{2\pi}{|B_2|}} = 1 \quad (16)$$

so

$$P(n) = \frac{1}{\sqrt{2\pi Npq}} e^{-\frac{(n-Np)^2}{2Npq}} \quad (17)$$

Comparing this to the Gaussian distribution

$$P(n) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(n-\mu)^2}{2\sigma^2}} \quad (18)$$

it can be seen that, at values of the parameters where the assumptions made in the derivation hold (i.e. at large N), the binomial distribution can be approximated by a Gaussian distribution with mean

$$\mu = Np$$

and standard deviation

$$\sigma = \sqrt{Npq}$$

2 A Variance Stabilising Transform for the Binomial Distribution

The most commonly used variance stabilising transform for the binomial distribution is

$$y = \arcsin \sqrt{\frac{n}{N}} \quad (19)$$

Applying the change-of-variables formula,

$$g(y) = f[r^{-1}(y)] \frac{dr^{-1}y}{dy} \quad \text{if } y = r(x) \quad (20)$$

gives the distribution in the new space

$$P(y) = \frac{N!}{(N \sin^2 y)!(N - N \sin^2 y)!} 2N \sin y \cos y p^{N \sin^2 y} q^{(N - N \sin^2 y)} \quad (21)$$

It is again possible to demonstrate that a Gaussian distribution provides a good approximation to this distribution at large N . As before, the aim is to expand the distribution as a Taylor series around the maximum

$$\ln[P(y' + z)] = \ln[P(y')] + zB_1 + \frac{z^2}{2!}B_2 + \frac{z^3}{3!}B_3 \dots \quad (22)$$

$$B_k = \left. \frac{d^k \ln[P(y)]}{dy^k} \right|_{y=y'} \quad (23)$$

where y' is the position of the mean and z is the distance in y from the mean. Again, since this is at the maximum B_1 must be zero and B_2 must be negative.

Taking the log of the distribution in y space and applying Stirling's approximation

$$\alpha! \approx \alpha^\alpha e^{-\alpha} \sqrt{2\pi\alpha} \quad (24)$$

gives, after simplification

$$\begin{aligned} \ln[P(y)] \approx & N \ln N + \frac{1}{2} \ln N + N \sin^2 y \ln p - N \sin^2 y \ln(N \sin^2 y) + N \ln q \\ & - N \sin^2 y \ln q - N \ln(N \cos^2 y) + N \sin^2 y \ln(N \cos^2 y) + \ln \sqrt{\frac{2}{\pi}} \end{aligned} \quad (25)$$

Differentiating gives

$$\begin{aligned} \frac{d \ln[P(y)]}{dy} \approx & 2N \sin y \cos y \ln p - 2N \sin y \cos y \ln(N \sin^2 y) - 2N \sin y \cos y \\ & - 2N \sin y \cos y \ln q + \frac{2N \sin y}{\cos y} - \frac{2N \sin^3 y}{\cos y} + 2N \sin y \cos y \ln(N \cos^2 y) \end{aligned} \quad (26)$$

which reduces to

$$\frac{d \ln[P(y)]}{dy} \approx 2N \sin y \cos y \ln \frac{p}{q} + 2N \sin y \cos y \ln \frac{\cos^2 y}{\sin^2 y} \quad (27)$$

Setting this equal to zero gives the position of the maximum: dividing through by $2N \sin y \cos y$ gives

$$\ln\left(\frac{p}{q}\right) = \ln\left(\frac{\sin^2 y}{\cos^2 y}\right) \quad (28)$$

Since $\cos^2 y = 1 - \sin^2 y$ and $q = 1 - p$, this gives the position of the maximum as

$$p = \sin^2 y = \frac{n}{N} \quad \text{so} \quad y = \arcsin \sqrt{p} \quad (29)$$

Armed with the position of the maximum, we can go on to find the higher differentials and thus the B_k :

$$\frac{d^2 \ln[P(y)]}{dy^2} \approx 2N \ln\left(\frac{p}{q}\right) \cos^2 y - 2N \ln\left(\frac{p}{q}\right) \sin^2 y + 2N \ln\left(\frac{\cos^2 y}{\sin^2 y}\right) \cos^2 y - 2N \ln\left(\frac{\cos^2 y}{\sin^2 y}\right) \sin^2 y - 4N \quad (30)$$

so

$$B_2 = -4N \quad (31)$$

Similarly

$$\frac{d^3 \ln[P(y)]}{dy^3} \approx -8N \ln\left(\frac{p}{q}\right) \cos y \sin y - 8N \cos y \sin y \ln\left(\frac{\cos^2 y}{\sin^2 y}\right) + 4N \frac{\sin y}{\cos y} - 4N \frac{\cos y}{\sin y} \quad (32)$$

giving

$$B_3 = 4N \left[\sqrt{\frac{p}{q}} - \sqrt{\frac{q}{p}} \right] \quad (33)$$

and

$$\frac{d^4 \ln[P(y)]}{dy^4} \approx 8N \ln\left(\frac{p}{q}\right) (\sin^2 y - \cos^2 y) + 8N \ln\left(\frac{\cos^2 y}{\sin^2 y}\right) (\sin^2 y - \cos^2 y) + 4N \left(\frac{1}{\sin^2 y} - \frac{1}{\cos^2 y} \right) + 16N \quad (34)$$

giving

$$B_4 = 4N \left[4 + \frac{1}{p} - \frac{1}{q} \right] \quad (35)$$

Each term of the Taylor expansion is again $O(N)$ smaller than its predecessor, (although this time the factors of N are hidden in the variable itself) implying that the approximation to the lower order terms becomes exact for large samples. Again, truncating the series at the B_2 term, plugging the B 's back into the expansion, and exponentiating gives

$$P(y) = P(y') e^{-|B_2| \frac{z^2}{2}} \quad (36)$$

In order for this to be normalized

$$P(y') = \sqrt{\frac{|B_2|}{2\pi}} \quad (37)$$

So, using $B_2 = 4N$, $z = y - y'$ and $y' = p$

$$P(y) = \sqrt{\frac{2N}{\pi}} e^{-2N(y - \arcsin \sqrt{p})^2} \quad (38)$$

and so the distribution is a Gaussian with mean

$$\mu = \arcsin \sqrt{p}$$

and standard deviation

$$\sigma = \sqrt{\frac{1}{4N}}$$

The variance is now independent of the mean, and so the transform to $\arcsin \sqrt{x}$ space has been shown to be the variance stabilising for the binomial distribution. Further, the approximation of the probability density distribution to a Gaussian becomes exact for large samples. The fact that the variance depends on N is a consequence of the space being scaled by N .

3 Accuracy of the Gaussian Approximation following the Transform

In order to demonstrate that the Gaussian approximation to the binomial distribution is more accurate after the variance stabilising transform than before it, first notice that in both cases the approximation is exact at the position of the maximum of the distribution. This can be seen from the Taylor expansions (Eqn. 3 and Eqn. 22): every term except the first term contains some power of the distance away from the mean, and so will fall to zero at the mean, and so at the position of the mean the Gaussian approximations and the original binomial distributions are exactly equal. Secondly, notice that each term in the expansion is $O(1/N)$ smaller than its predecessor in both cases, and so the cubic term in the expansions dominates the error. The Gaussian approximations in both cases required normalisation of the expansions, and so it is not sufficient to look at the cubic terms in isolation. However, the linear (B_1) terms were zero in both cases, and the quadratic (B_2) terms were the only terms retained. Therefore, we can examine the ratios of the cubic to the quadratic terms to give measures of the proportional errors in each case. Furthermore, since the errors are zero at the position of the maxima of the distributions, it is sufficient to find which of these ratios has the highest first differential at the position of the maxima i.e. which proportional error grows faster as we move away from that position.

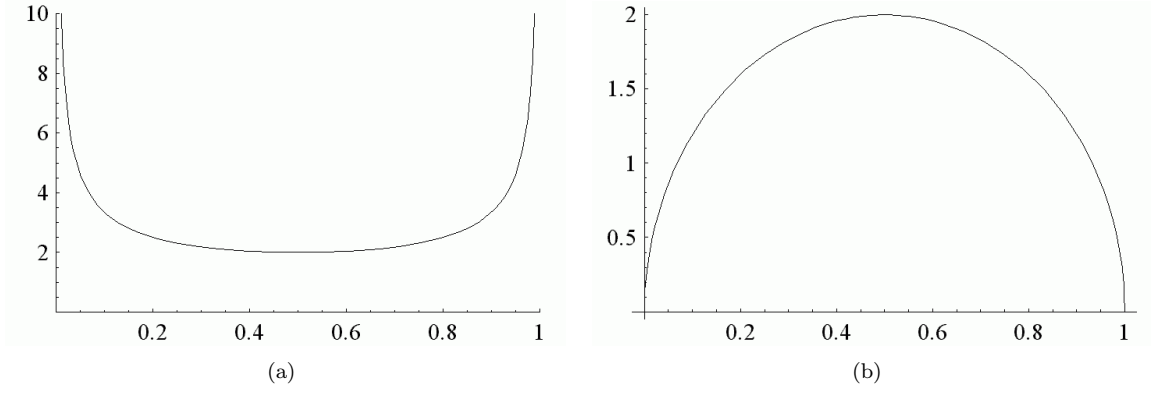


Figure 1: Ratios of the first derivatives of the proportional errors on the B_3 term to the B_2 term before to after the variance stabilising transform. The expression is plotted against p for $x = 0.5$ (a) and against x for $p = 0.5$ (b).

Before the variance stabilising transform, the quadratic term of the expansion is

$$\frac{1}{Np(1-p)} \frac{(n-Np)^2}{2!} \quad (39)$$

and the cubic term is

$$-\frac{(1-2p)}{N^2p^2(1-p)^2} \frac{(n-Np)^3}{3!} \quad (40)$$

so the ratio is

$$R_B = -\frac{1-2p}{Np(1-p)} \frac{(n-Np)}{3} \quad (41)$$

For ease of comparison with the ratio after the variance stabilising transform, write this in terms of the proportion $x = n/N$,

$$R_B = -\frac{1-2p}{p(1-p)} \frac{(x-p)}{3} \quad (42)$$

Similarly, after the variance stabilising transform the quadratic and cubic terms of the Taylor expansion are

$$-4N \frac{(\arcsin\sqrt{x} - \arcsin\sqrt{p})^2}{2!} \quad (43)$$

and

$$4N \left(\sqrt{\frac{p}{q}} - \sqrt{\frac{q}{p}} \right) \frac{(\arcsin\sqrt{x} - \arcsin\sqrt{p})^3}{3!} \quad (44)$$

and so the ratio is

$$R_A = -\left(\sqrt{\frac{p}{q}} - \sqrt{\frac{q}{p}} \right) \frac{(\arcsin\sqrt{x} - \arcsin\sqrt{p})}{3} \quad (45)$$

Using

$$\frac{d}{dx} \arcsin\sqrt{x} = \frac{1}{2\sqrt{x(1-x)}} \quad (46)$$

the first differentials of the ratios are

$$\frac{dR_B}{dx} = -\frac{(1-2p)}{3p(1-p)} \quad (47)$$

and

$$\frac{dR_A}{dx} = -\left(\sqrt{\frac{p}{q}} - \sqrt{\frac{q}{p}} \right) \frac{1}{6\sqrt{x(1-x)}} = \frac{1}{6\sqrt{x(1-x)}} \frac{(1-2p)}{\sqrt{p(1-p)}} \quad (48)$$

One additional comment should be made here: notice that neither of these functions go to zero anywhere in the range $0 \leq x \leq 1$. Therefore, the proportional error functions have no extrema i.e. they cannot change signs. The ratio of the differentials is

$$\frac{dR_B}{dx} \bigg/ \frac{dR_A}{dx} = 2 \frac{\sqrt{x(1-x)}}{\sqrt{p(1-p)}} \quad (49)$$

At the position of the maximum $x = p$ this reduces to

$$\frac{dR_B}{dx} \bigg/ \frac{dR_A}{dx} = 2 \quad (50)$$

Therefore, the errors on the Gaussian approximation grow twice as fast around the mean before the variance stabilising transform as they do after it, and so the transform makes the Gaussian approximation approximately twice as accurate. Fig. 1 shows plots of Eqn. 49 against p for $x = 0.5$ and against x for $p = 0.5$. It should be noted that the choice of fixed x or p in these plots only affects the scaling, not the shape of the curve, and that the scaling is always fixed so that the expression is equal to 2 at $x = p$. Examining the plot against p , it can be seen that the curve is almost flat across most of the range, but rises steeply below around $p = 0.2$ and above $p = 0.8$. This is as expected: it corresponds to the ends of the range, where the binomial distribution is highly skewed and so is not well approximated by a Gaussian, and also agrees with the literature on the variance stabilising transform for the binomial distribution, which states that it is only necessary to use this transform when the data contains proportions in the extremes of the range. Examining the plot against x , it can be seen that the curve is greater than 1 over a large range around the mean. The curve drops below 1 at the ends of the range, indicating that the errors on the Gaussian approximation after the variance stabilising transform are beginning to increase faster than those before the transform. However, this curve shows the ratio of the first differentials. The area between the curve and the line $y = 1$ therefore indicates the absolute errors involved, and since that area is greater between $p = x$ and the point where $y = 1$ than between $y = 1$ and $p = 0$, the errors after the transform can never catch up with the errors before it. The variance stabilising transform therefore makes the Gaussian approximation more accurate across the entire range.

4 Conclusion

This report has demonstrated that the binomial distribution may be approximated with a Gaussian distribution at large N ; that the transformation to $\arcsin\sqrt{x}$ space is the variance stabilising transform for the binomial distribution; and that the Gaussian approximation is approximately twice as accurate after the variance stabilising transform for small values of N . The same behavior was seen with the variance stabilising transform of the Poisson distribution: the Gaussian approximation is more accurate after the application of the transform. This effect is often seen with normalising or variance stabilising transforms: that a transform intended for one purpose (variance stabilisation) also aids another (normalisation). However, it is not generally true, and furthermore the transformations that approximately achieve both ends tend to be compromises, and there may exist other transformations that achieve, for example, better normalisation at the expense of no variance stabilisation. A fuller discussion of these matters with examples is given in [1]. This is a reminder of the fact that such transformations result in only better *approximations* of the data to Gaussian distributions so that we can then more reliably use established methods which make such an assumption. Of course it would be possible though more awkward to have an exact likelihood calculation for the data, based directly upon the binomial distribution, in the same way that the ‘Fisher exact test’ is available for Poisson distributed data. The use of transform techniques must therefore be applied with the an understanding of the practical consequences of this approximation.

There are several practical implications of the above results. The first is in the construction of chi-square similarity measures for data and a ratio predicted from theory (h_i and t_i) or data and data (h_i and k_i), which rely on the approximation of a binomial by a Gaussian for the construction of a “least-squares” measure. These functions are better estimated (particularly for small values of h_i and k_i) as

$$\chi^2 = 4 \sum_i N_i (\arcsin\sqrt{h_i/N_i} - \arcsin\sqrt{t_i})^2 \text{ and } \chi^2 = 2 \sum_i N_i (\arcsin\sqrt{h_i/N_i} - \arcsin\sqrt{k_i/N_i})^2$$

In addition, the binomial can be better approximated by a Gaussian of form given in Eqn. 38 than by the standard Gaussian of form Eqn. 17.

The \arcsin transform and *square root* transforms [2] are also necessary for the construction of (frequentist) probability similarity measures [3].

Finally, a Gaussian random variate x with variance¹ $1/4N$ and mean $\arcsin\sqrt{n/N}$ can be used to approximate a binomial variate y for use in Monte-Carlo simulations by applying the transform $y = \sin^2(x)$.

¹Thanks to Iain Buchan for pointing out a long-standing typographical error in previous versions of this document.

References

- [1] A. Stuart and K. Ord, "Kendall's Advanced Theory of Statistics Vol 2a: Classical Inference and the Linear Model", pp. 753-759, Arnold, 1994.
- [2] N.A. Thacker and P.A. Bromiley, The Effects of a Square Root Transform on a Poisson Distributed Quantity. Tina memo, 2001-010, 2001.
- [3] N.A.Thacker, P.Bromiley, The Equal Variance Domain: Issues Surrounding the use of Probability Densities for Algorithm Construction. Tina memo, 2004-005, 2004.