On Finding Geodesic Equation of Two Parameters Binomial Distribution

William W.S. Chen

Department of Statistics, The George Washington University
Washington D.C. 20013
williamwschen@gmail.com

ABSTRACT

The purpose of this paper is to find a general form of the geodesic equation of the binomial distribution. Using Darboux's theory we will set up a second order partial differential equation. Then we will apply the chain rule to transform the variable and rotate the axis to remove the interaction term, which will lead us to find the geodesic equation of binomial distribution. To illustrate how we can find such a geodesic equation in practice, we demonstrate by an example.

SOME KEY WORDS AND PHRASES: Bernoulli Trial, binomial distribution, Darboux Theory, differential geometry, geodesic equation, digamma function, Rotation axis, second order partial differential equation, trigamma function.

1   Introduction

A Bernoulli experiment is a random experiment, the outcome of which can be classified in but one of two mutually exclusive and exhaustive ways: success or failure. A sequence of Bernoulli trials occurs when a Bernoulli experiment is performed several independent times so that the probability of success, p, remains the same from trial to trial. We may let p denote the probability of success on each trial and let

q=1-p denote the probability of failure. In a sequence of Bernoulli trials, we are often interested in the total number of successes and are not interested in the order of this occurrence. If we let the random variable x equal the number of observed success in n Bernoulli trial, then n-x failure occur. The number of

ways of selecting x positions for the x successes in the n trials is $\binom{n}{x} = \dfrac{n!}{x!(n-x)!}$, and the probability of

each of these ways is $p^x (1-p)^{n-x}$ Thus, the probability density function of x is the sum of the probability

of these $\binom{n}{x}$ mutually exclusive events; $f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \ x=0,1,...n$ The random variable x

is said to have binomial distribution. The constants n and p are called the parameters of the binomial distribution. "Are we alone in the galaxy?" The existence of life on Earth does not mean that the probability of life emerging once is equal to 1. When we roll a die and get a 6, the probability of getting

the 6 does not change to 1, it remains 1/6. However, we do know that the probability of life emerging on a habitable planet is at least greater than zero. The binomial distribution tells us the probability of x=k successes in a series of Bernoulli trials. So, for n habitable planets, we can calculate the probability that x=k of these planets will host life, $f(x) = \binom{n}{x} p^x (1\text{-}p)^{n\text{-}x}$ for k=0, we get $f(k=0) = (1-p)^n$. The

probability that life exists on at least one planet, $P(k \geq 0)$ or $P(k \neq 0)$, is equal to 1 minus the

probability that life exists on zero planets, $P(k \neq 0) = 1\text{-}(1\text{-}p)^n$.

To answer the question "are we alone?" What we really want to know is the probability that life

exists on at least two planets; $P(k \geq 2) = 1\text{-}P(k=0)\text{-}P(k=1) = 1\text{-}(1\text{-}p)^n - np(1-p)^{n-1}$. The probability

tells us, given n and p, the probability that two or more planets will host life. In this paper, we give out another useful result: geodesic equation.

2 List the Fundamental Tensor

The standard form of the two parameters binomial distribution has the probability density function given by,

$$f(x) = \binom{n}{x} p^x (1\text{-}p)^{n\text{-}x} = \frac{n!}{x!(n\text{-}x)!} p^x (1\text{-}p)^{n\text{-}x}$$

$$= \frac{\Gamma(n+1)}{\Gamma(x+1)\Gamma(n-x+1)} p^x (1\text{-}p)^{n\text{-}x} \qquad\qquad 0 < p < 1, \quad x = 0,1,2,...n$$

$$\ln f(x) = \ln\Gamma(n+1) - \ln\Gamma(x+1) - \ln\Gamma(n-x+1) + x\ln p + (n\text{-}x)\ln(1\text{-}p) \qquad (2.1)$$

From the equation (2.1) above, we derive the basic metric tensor components for this distribution as follows,

$$\frac{\partial \ln f(x)}{\partial n} = \frac{\Gamma'(n+1)}{\Gamma(n+1)} - \frac{\Gamma'(n-x+1)}{\Gamma(n-x+1)} + \ln(1\text{-}p) = \Psi(n+1)\text{-}\Psi(n\text{-}x+1) + \ln(1\text{-}p)$$

$$\frac{\partial^2 \ln f(x)}{\partial n^2} = \Psi'(n+1)\text{-}\Psi'(n\text{-}x+1) \qquad\qquad\qquad (2.2)$$

$$\frac{\partial^2 \ln f(x)}{\partial p \partial n} = \frac{-1}{1-p}; \qquad\qquad \frac{\partial \ln f(x)}{\partial p} = \frac{x}{p} - \frac{n-x}{1-p} \qquad\qquad (2.3)$$

$$\frac{\partial^2 \ln f(x)}{\partial p^2} = \frac{-x}{p^2} - \frac{n-x}{(1-p)^2} \qquad\qquad\qquad (2.4)$$

Taking the negative expectation of equations (2.2),(2.3) and (2.4), we define the coefficient of the first fundamental

---

form E, F, and G as follows:

$$E = -E(\frac{\partial^2 \ln f(x)}{\partial n^2}) = -\Psi'(n+1) + E(\Psi'(n-x+1)) = \lambda_x(n,p) \qquad (2.5)$$

$$\text{where } \lambda_x(n,p) = -\Psi'(n+1) + \sum_{x=0}^{n} \Psi'(n-x+1)\binom{n}{x}p^x(1-p)^{n-x}$$

$$F = -E(\frac{\partial^2 \ln f(x)}{\partial p \partial n}) = \frac{1}{1-p}; \qquad (2.6)$$

$$G = -E(\frac{\partial^2 \ln f(x)}{\partial p^2}) = \frac{np}{p^2} + \frac{n-np}{(1-p)^2} = \frac{n}{p(1-p)} \qquad (27)$$

Equations (2.5), (2.6) and (2.7) will be used to set up the partial differential equation in the next section.

3   The Geodesic Equation

In this section, we will find the geodesic equation of the binomial distribution by solving a second order partial differential equation. This idea originated from Darboux's, theory. We can set up $\nabla Z = 1$ as follows:

$$\frac{EZ_p^2 - 2FZ_nZ_p + GZ_n^2}{EG - F^2} = 1$$

$$\lambda_x(n,p) Z_p^2 - \frac{2}{1-p}Z_nZ_p + \frac{n}{p(1-p)}Z_n^2 = \frac{n\lambda_x(n,p)(1-p)-p}{p(1-p)^2} \qquad (3.1)$$

To solve the partial differential equation (3.1) above, we would consider the polar coordinate transformation. Let $n = r\cos\theta$, *and*

$p = r\sin\theta$, we should be aware of the fact that Z is a function (n,p) while both (n,p) are also functions of (r, θ). In calculus, we learn that the chain rule will give us the following results:

$$Z_r = \frac{\partial Z}{\partial n}\frac{\partial n}{\partial r} + \frac{\partial Z}{\partial p}\frac{\partial p}{\partial r} = Z_n\cos\theta + Z_p\sin\theta,$$

$$Z_\theta = \frac{\partial Z}{\partial n}\frac{\partial n}{\partial \theta} + \frac{\partial Z}{\partial p}\frac{\partial p}{\partial \theta} = Z_n(-r\sin\theta) + Z_p r\cos\theta, \qquad (3.2)$$

Using the Cramer Rule in equation (3.2), we can solve reversely for Z_n *and* Z_p as a function of Z_r *and* Z_θ as follows:

$$Z_n = \frac{\begin{vmatrix} Z_r & \sin\theta \\ Z_\theta & r\cos\theta \end{vmatrix}}{\begin{vmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{vmatrix}} = \frac{r\cos\theta Z_r - \sin\theta Z_\theta}{r\cos^2\theta + r\sin^2\theta} = \cos\theta Z_r - \frac{1}{r}\sin\theta Z_\theta$$

$$Z_p = \frac{\begin{vmatrix} \cos\theta & Z_r \\ -r\sin\theta & Z_\theta \end{vmatrix}}{\begin{vmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{vmatrix}} = \frac{\cos\theta Z_\theta + r\sin\theta Z_r}{r\cos^2\theta + r\sin^2\theta} = \sin\theta Z_r + \frac{1}{r}\cos\theta Z_\theta \tag{3.3}$$

Substitute (3.3) into (3.1), we get:

$$\lambda_x(n,p)(\sin\theta\, Z_r + \frac{1}{r}\cos\theta\, Z_\theta)^2 - \frac{2}{1-p}(\sin\theta\, Z_r + \frac{1}{r}\cos\theta\, Z_\theta)(\cos\theta\, Z_r - \frac{1}{r}\sin\theta\, Z_\theta)$$

$$+\frac{n}{p(1-p)}(\cos\theta Z_r - \frac{1}{r}\sin\theta Z_\theta)^2 = \frac{n\lambda_x(n,p)(1-p)-p}{p(1-p)^2} \tag{3.4}$$

To calculate the coefficient of $Z_r Z_\theta$:

$$\frac{\lambda_x(n,p)}{r}2\sin\theta\cos\theta - \frac{2}{1-p}(\frac{\cos^2\theta}{r} - \frac{\sin^2\theta}{r}) - \frac{n}{p(1-p)}\frac{2}{r}\sin\theta\cos\theta = 0$$

$$(\lambda_x(n,p)p(1-p)-n)\tan 2\theta = 2p$$

$$2\theta = \tan^{-1}\frac{2p}{\lambda_x(n,p)p(1-p)-n} = \tan^{-1}\lambda_2(n,p) \tag{3.5}$$

To calculate the coefficient of Z_r^2:

$$\lambda_3(n,p) = \lambda_x(n,p)\sin^2\theta - \frac{1}{1-p}\sin 2\theta + \frac{n\cos^2\theta}{p(1-p)} \tag{3.6},$$

To calculate the coefficient of Z_θ^2:

$$r^2\lambda_4(n,p) = \lambda_x(n,p)\cos^2\theta + \frac{\sin 2\theta}{1-p} + \frac{n\sin^2\theta}{p(1-p)} \tag{3.7},$$

and constant term $C_0 = \dfrac{n\lambda_x(n,p)(1-p)-p}{p(1-p)^2}$.

After rotating to a proper angle, the new partial differential equation (3.4) becomes

$$\lambda_3(n,p)Z_r^2 + r^2\lambda_4(n,p)Z_\theta^2 = C_0$$

$$\lambda_3(n,p)Z_r^2 = C_0 - r^2\lambda_4(n,p)Z_\theta^2 = A^2 \tag{3.8}$$

where $\lambda_3(n,p)$, and $\lambda_4(n,p)$ defined the same as equation (3.6) and (3.7).

Now, we can break the above equation (3.8) into two separate parts, and let them equal the same constant, say A^2,

part 1,

$$Z_r^2 = \frac{A^2}{\lambda_3(n,p)}; \quad Z_r = \pm\frac{A}{\sqrt{\lambda_3(n,p)}} \quad Z = \pm\frac{Ar}{\sqrt{\lambda_3(n,p)}} \tag{3.9}$$

part 2,

$$C_0 - r^2\lambda_4(n,p)Z_\theta^2 = A^2; \quad Z_\theta^2 = \frac{C_0 - A^2}{r^2\lambda_4(n,p)}; \quad Z = \pm\sqrt{\frac{C_0 - A^2}{r^2\lambda_4(n,p)}} \ \theta; \tag{3.10}$$

Put equations (3.9) and (3.10) together to arrive at the general solution of equation (3.8),

$$Z = \pm\frac{Ar}{\sqrt{\lambda_3(n,p)}} \ \pm\sqrt{\frac{C_0 - A^2}{r^2\lambda_4(n,p)}} \ \theta; \tag{3.11}$$

Applying the Darboux Theory, we find that the geodesic equation of binomial distribution is given by,

$$\frac{\partial Z}{\partial A} = B \ ;$$

$$\pm\frac{r}{\sqrt{\lambda_3(n,p)}} \ \pm\frac{A\theta}{\sqrt{(C_0 - A^2)\lambda_4(n,p)r^2}} = B \tag{3.12}$$

From previously defined relations, we know that (r,θ) and (n,p) are related to $n = r\cos\theta, \ p = r\sin\theta,$

$$r^2 = n^2 + p^2 \quad \text{and} \quad \tan\theta = \frac{p}{n};$$

$$or \ \ r = \pm\sqrt{n^2 + p^2} \quad \text{and} \quad \theta = \tan^{-1}\frac{p}{n} \tag{3.13}$$

hence after substituting the relation (3.13) into equation (3.12) we find our geodesic equation of binomial distribution as:

$$\pm\frac{\sqrt{n^2 + p^2}}{\sqrt{\lambda_3(n,p)}} \ \pm\frac{A\tan^{-1}\frac{p}{n}}{\sqrt{(C_0 - A^2)(n^2 + p^2)\lambda_4(n,p)}} = B \tag{3.14}$$

where C_0, $\lambda_3(n,p)$ and $\lambda_4(n,p)$ are defined the same as before, and A, B are arbitrary constants.

4   Example

In this section we choose two values for the parameters $n = 10$ *and* $p = 0.5$. Later we will calculate five constants to specify a unique fixed binomial distribution. We list these five constants as follows:

$$\lambda_x(n, p) = -\Psi'(n+1) + \sum_{x=0}^{n} \Psi'(n\text{-}x + 1) \binom{n}{x} p^x (1 - p)^{n-x}$$

$$\lambda_2(n, p) = \frac{2p}{\lambda_x(n, p)p(1 - p) - n};$$

$$\tan 2\theta = \lambda_2(n, p); \quad \theta = \frac{1}{2}\tan^{-1}\lambda_2(n, p)$$

$$\lambda_3(n, p) = \lambda_x(n, p)\sin^2\theta - \frac{\sin 2\theta}{1 - p} + \frac{n\cos^2\theta}{p(1 - p)};$$

$$r^2\lambda_4(n, p) = \lambda_x(n, p)\cos^2\theta + \frac{\sin 2\theta}{1 - p} + \frac{n\sin^2\theta}{p(1 - p)};$$

$$C_0 = \frac{n\lambda_x(n, p)(1 - p) - p}{p(1 - p)^2}$$

We translate the equation of $\lambda_x(n, p)$ directly from R-package language as follows:

$\lambda_x(10, 0.5)$ **<-** -trigamma(11)+sum(trigamma(10+1-(0:10))*

   choose(10,0:10)*0.5^10) (4.1)

We tabulate our computation results as follows:

$\lambda_x(10, 0.5) = 0.10674$, $\lambda_2(10, 0.5) = -0.1002676$, $\theta = -0.04996678$,

$\lambda_3(10, 0.5) = 40.10002$, $r^2\lambda_4(10, 0.5) = 0.006723169$, $C0 = 0.2695992$

So the geodesic equation for $n = 10$ and $p = 0.5$ is given by

$$\pm\frac{\sqrt{n^2 + p^2}}{\sqrt{40.10002}} \pm \frac{A\tan^{-1}\dfrac{p}{n}}{\sqrt{0.006723169(0.2695992 - A^2)}} = B$$

where A and B are arbitrary constants, α and p are some parameters

5   Concluding Remarks

Among the five constants, we realize that the first constant, $\lambda_x(n, p)$, is the most critically important one, because the remaining four constants are dependent on its value. We suggested two methods to compute this constant. Method one as we demonstrated in section 4 by the R-Package. Alternatively, one can use Chen (1982) coded in Fortran IV computer program that can compute the digamma psi functions. It is important to aware of the fact that a finite sum of binomial series converge quickly. However, negative binomial distribution is an infinite series sum, and has no promise of converging. This finite sum may be viewed as a weighted mean of binomial distribution plus some constant. After we find the first constant, $\lambda_x(n, p)$, it would be a straightforward matter to compute the remaining four parameters. Give us a proper angle to rotate, and an interaction term will disappear. Then a specified binomial distribution geodesic equation can be defined.

REFERENCES

[1]   Apostol T.M.(1974) Mathematical Analysis. Addison-Wesley Publishing Company. Second Edition.

[2]   Balakrishnan N. and Nevzorov V.B.(2003) A Primer on Statistical Distributions. John Wiley & Sons, Inc.

[3]   Chen W.W.S (1982) Evaluation of the first 12 derivatives of the digamma psi functions with applications . Proceeding of Statistical Computing Section, 1982, pp293-298.

[4]   Chen W.W.S. (2017) On Finding Geodesic Equation of Student T Distribution. Journal of Mathematics Research. Vol. 9. No. 2, April 2017, pp32-37.

[5]   Crawley M.J.(2007) The R Book. John Wiley & Sons Ltd. The Atrium, Southern Gate, Chichester, West Sussex PO198SQ, England.

[6]   Darboux, G.(2nd ed, 1914) Lecons sur la theorie generale des surfaces. 4 vols, Gauthier-Villars, Paris. I, 1887, 513pp. ;II,1889,522pp.; III,1894,512pp.; IV,1896, 548pp.

[7]   Grey A.(1993) Modern differential geometry of curves andsurfaces. CRC Press, Inc. Boca Raton.

[8]   Kass, R.E, and Vos, P.W.(1997) Geometrical foundations of asymptotic inference. John Wiley & Sons, Inc. New York. https://doi.org/10.1002/9781118165980

[9]   Struik, D.J.(1961) Lectures on classical differential geometry. Second Edition. Dover Publications, Inc.