ICASTOR Journal of Mathematical Sciences
Vol. 4, No. 2 (2010) 197 – 207

THE NEGATIVE BINOMIAL PARAMETER *k* AS A MEASURE OF DISPERSION

I. C. Kipchirchir

School of Mathematics
University of Nairobi
Nairobi, Kenya

Correspondence: School of Mathematics, University of Nairobi, P.O. Box 30197-00100, Nairobi. Kenya. Email: kipchirchir@uonbi.ac.ke

ICASTOR Journal of Mathematical Sciences
Vol. 4, No. 2 (2010) 197 – 207

THE NEGATIVE BINOMIAL PARAMETER k AS A MEASURE OF DISPERSION

I. C. Kipchirchir

School of Mathematics
University of Nairobi
Nairobi, Kenya

ABSTRACT

The role of the negative binomial parameter k and index of patchiness as measures of dispersion in relation to random removal of individuals are discussed. The aim of this paper is to demonstrate analytically that the negative binomial parameter k is a measure of dispersion. The negative binomial parameter k and index of patchiness are inversely related. Equicorrelation matrix is analysed in relation to coefficient of determination, partial correlation and principal components with respect to the negative binomial parameter k. The analysis demonstrates that small values of k are associated with overdispersion whereas large values are associated with randomness.

KEYWORDS: Dispersion, overdispersion, randomness, negative binomial, measure of dispersion.

INTRODUCTION

Dispersion is the description of the pattern of distribution of organisms in space (Southwood, 1966) and is often referred to as spatial distribution[1]. It is a characteristic ecological property. Probability distribution models are used to quantify and classify the dispersion of organisms. They have been widely used in entomological research to describe the dispersion of insects.

In order to describe dispersion it is assumed that organisms are confined to discrete habitable sites called units (sampling units). In reference to sampling of pests on plants as a prototype, for instance larvae of a pest species that attack the shoots of a plant, the following assumptions are made:

(i) Each shoot constitute a habitable site and is a natural sampling unit.
(ii) The larvae will not be found elsewhere than on shoots and therefore the space available to them is discontinuous.
(iii) Migrations from one shoot to another, even if possible, are assumed to be uncommon enough to be ignored.

[1] Spatial distribution is a visual description and not a probability distribution.

Correspondence: School of Mathematics, University of Nairobi, P.O. Box 30197-00100, Nairobi. Kenya.
Email: Kipchirchir@uonbi.ac.ke

Thus, if we count the number of organisms per unit for a large sample of the units, the observations clearly convey something about the spatial pattern of the species.

Now, suppose every unit contains a large number of locations and each of these can be occupied by a single individual. We suppose that η is the maximum number of individuals that a unit may contain and the probability that every location in every unit is occupied is ξ (a constant). We further suppose that a random variable X represents the number of individuals a unit may contain. Assuming independence, the probability that exactly x locations in any one unit will be occupied is given by the binomial probability

$$P(X = x) = \binom{\eta}{x} \xi^x (1-\xi)^{\eta - x}, \quad x = 0,1,2,\ldots,\eta; \quad 0 < \xi < 1. \tag{1}$$

Assuming that η is very large, ξ is very small, so that $\eta\xi = \lambda$ (constant), then

$$\lim_{\eta \to \infty} P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0,1,2,\ldots; \quad \lambda > 0. \tag{2}$$

which is the Poisson probability.

If the individuals had been assigned independently and at random to the available units (each organism has the same chance of occupying any unit), the spatial distribution is said to be random and the population pattern is also random. Random spatial distribution is described by Poisson distribution. For Poisson distribution, the mean (m) and the variance (v) are equal.

In ecological studies, the variance is usually found to be larger than the mean and the Poisson distribution rarely fits the observed frequency distribution of the number of individuals per unit.

Departure from randomness is centred on the heterogeneity of the environment and pest behaviour (Young and Young, 1990). The behavioural response to their environment (food-gathering traits) and other organisms (mating behaviour) results in the presence of one organism in a unit, increasing the chances of occurrence of another organism in the unit. Thus, units are not equally receptive or attractive to organisms and are depicted by a larger variance than the mean. When the variance is larger than the mean, the spatial distribution is said to be contagious and the population pattern in this case is clumped or patchy or aggregated or clustered or overdispersed.

If the organisms are distributed in a uniform or regular fashion, the variance is less than the mean. The spatial distribution is said to be regular and the population pattern is said to be underdispersed (uniform). The regular distribution[2] is best described by the binomial distribution $(v < m)$.

Many overdispersed pest populations that have been studied can adequately be described by the negative binomial distribution

[2] Regular distribution is seldom observed unless during presence-absence sampling.

$$P(X=x)=\binom{k+x-1}{x}\left(\frac{p}{1+p}\right)^{x}\left(\frac{1}{1+p}\right)^{k},\ x=0,1,2,\ldots;\ k>0,\ p>0 \tag{3}$$

so that $v = m(1 + p) > m = kp$.

Anscombe (1949) gave statistical analysis of insect counts based on the negative binomial distribution. The negative binomial parameter k is considered as a dispersion parameter. Small values of $k(k \to 0)$ are associated with overdispersion, whereas large values of $k(k \to \infty)$ are associated with randomness. The three spatial distributions are illustrated in Figure 1.

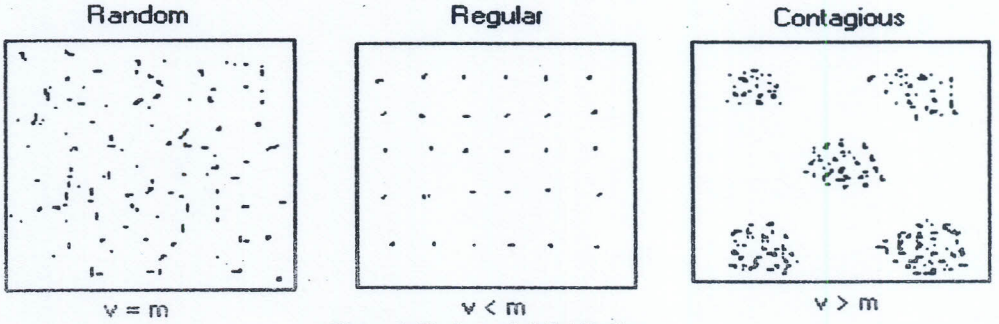

Figure 1: Basic spatial distributions.

1. MEASURING DISPERSION

A measurable property of a population spatial pattern, which is equivalent to dispersion but independent of population density, can be derived from observed frequency distribution of the number of individuals per unit. Measures of dispersion are often used to draw inferences relating to the spatial pattern of a population. Measures of dispersion, which are not affected by small changes in population density, are thought of as representing some intrinsic property of a spatial pattern, whatever the density.

It is assumed that if some individuals are randomly removed from the population, the remaining individuals still remain at their original locations. Thus, random removal of individuals would only affect the population density while leaving the pattern unchanged. However, if a large number of individuals were removed from the originally densely populated units, we expect the pattern to change, resulting in reduced aggregation.

One of the measures of dispersion is the index of patchiness

$$\tilde{P}=1+\frac{1}{m}\left(\frac{v}{m}-1\right) \tag{4}$$

so that $\tilde{P} = 1, \tilde{P} > 1$ and $\tilde{P} < 1$ for random, contagious and regular spatial distributions respectively. The index of patchiness is not affected by random removal of individuals (Pielou, 1978). Intuitively, populations with more or less different densities can exhibit the same degree of patchiness.

Let X be a random variable having the negative binomial distribution with parameters k and $m_0 = kp$, representing the number of individuals in a unit before removal, and let Y be a random variable representing the number of individuals remaining in a unit after removal. The probability that a unit will contain y individuals, given that it formally contained x individuals before removal, is

$$P(Y = y/x, \theta) = \binom{x}{y} \theta^y (1-\theta)^{x-y}, \quad y \leq x, \quad 0 < \theta < 1 \tag{5}$$

where θ is the probability of an individual remaining in the population and

$$\begin{aligned} P(Y = y/\theta) &= \sum_{x=y}^{\infty} P(Y = y, X = x/\theta) \\ &= \sum_{x=y}^{\infty} P(Y = y/x, \theta) P(X = x) \\ &= \sum_{x=y}^{\infty} \binom{x}{y} \theta^y (1-\theta)^{x-y} \binom{k+x-1}{x} \left(\frac{p}{1+p}\right)^x \left(\frac{1}{1+p}\right)^k \\ &= \binom{k+y-1}{y} \left(\frac{\theta p}{1+\theta p}\right)^y \left(\frac{1}{1+\theta p}\right)^k, \quad y = 0,1,2,\ldots \end{aligned} \tag{6}$$

which is negative binomial distribution with parameters k and $m_1 = k\theta p = \theta m_0$. Whilst the population density has changed, k has remained the same. Thus, k is unaltered by random removal of individuals and it can be used as a measure of dispersion in patterns that yield negative binomial distribution when sampled. The variances before and after random removal are $v_0 = m_0(1+p)$ and $v_1 = m_1(1+\theta p)$ respectively.

The negative binomial distribution has index of patchiness

$$\tilde{P} = 1 + \frac{1}{m}\left(\frac{v}{m} - 1\right) = 1 + \frac{1}{k} > 1 \tag{7}$$

implying that negative binomial distribution describes a contagious distribution. Furthermore,

$$\tilde{P}_0 = 1 + \frac{1}{m_0}\left(\frac{v_0}{m_0} - 1\right) = 1 + \frac{1}{k} = 1 + \frac{1}{m_1}\left(\frac{v_1}{m_1} - 1\right) = \tilde{P}_1 \tag{8}$$

implying $\tilde{P}$ is not affected by random removal of individuals. We observe that

$$\lim_{k\to\infty} \tilde{P} = \lim_{k\to\infty} \left(1+\frac{1}{k}\right) = 1 \tag{9}$$

implying as $k \to \infty$, negative binomial distribution describes a random distribution.

From Equation (7),

$$k = \frac{1}{\tilde{P}-1}, \quad \tilde{P} \geq 1, \tag{10}$$

implying, $k > 0$ and $\tilde{P} \geq 1$ are inversely related and consequently

$$k \to \begin{cases} 0, & as \quad \tilde{P} \to \infty \\ \infty & as \quad \tilde{P} \to 1 \end{cases} \tag{11}$$

that is, small values of k are associated with overdispersion whereas large values of k are associated with randomness. Underdispersion ($\tilde{P} < 1$) is not accounted for by the negative binomial distribution.

For fixed $m = kp$ and reparameterizing $\beta = \frac{p}{1+p}$, then $k \to \infty$ or $p \to 0$ or $\beta \to 0$ imply a random distribution while $k \to 0$ or $p \to \infty$ or $\beta \to 1$ imply a contagious distribution.

2. EQUICORRELATION MATRIX ANALYSIS

We suppose that there are n sampling units and to obtain $Corr(X_i, X_j), i \neq j$, we have to consider the multivariate negative binomial distribution

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \frac{\Gamma(k+\sum_{i=1}^{n} x_i)}{q^k \Gamma(k) \prod_{i=1}^{n} x_i!} \prod_{i=1}^{n} \left(\frac{p}{q}\right)^{x_i}, \tag{12}$$

$$x_i = 0,1,2,\dots; \quad q = np+1, \quad p > 0$$

having probability generating function

$$g(s_1, s_2, \dots, s_n) = \left(q - p\sum_{i=1}^{n} s_i\right)^{-k} \tag{13}$$

so that

$$E(X_i) = \left.\frac{\partial g(s_1, s_2, \dots, s_n)}{\partial s_i}\right|_{s_i=1} = kp,$$
$$E\big(X_i(X_i-1)\big) = \left.\frac{\partial^2 g(s_1, s_2, \dots, s_n)}{\partial s_i^2}\right|_{s_i=1} = k(k+1)p^2,$$
$$E\big(X_iX_j\big) = \left.\frac{\partial^2 g(s_1, s_2, \dots, s_n)}{\partial s_i \partial s_j}\right|_{s_i=s_j=1} = k(k+1)p^2, \ i \neq j, \tag{14}$$
$$Var(X_i) = E\big(X_i(X_i-1)\big) + E(X_i) - \big(E(X_i)\big)^2 = kp(1+p),$$
$$Cov\big(X_i, X_j\big) = E\big(X_iX_j\big) - E(X_i)E\big(X_j\big) = kp^2, \ i \neq j$$

and

$$Corr\big(X_i, X_j\big) = \frac{Cov\big(X_i, X_j\big)}{\sqrt{Var(X_i)Var\big(X_j\big)}} = \frac{p}{1+p} = \beta, \ i \neq j. \tag{15}$$

From (15), we observe that

$$Corr\big(X_i, X_j\big) \to \begin{cases} 0, & random \\ \\ 1, & contagious \end{cases} \tag{16}$$

The equicorrelation matrix is

$$P_n = (1-\beta)I_n + \beta J_n \tag{17}$$

where I_n is an n-dimensional identity matrix, J_n is an $n \times n$ matrix of ones (unit matrix) and β is the equicorrelation. We observe that

$$P_n \to \begin{cases} I_n, & random \\ \\ J_n, & contagious \end{cases} \tag{18}$$

The unit matrix J_n can be expressed as

$$J_n = \mathbf{1}\mathbf{1}' \tag{19}$$

where $\mathbf{1}'$ is an n-dimensional vector of ones (unit vector). Since J_n is symmetric and has rank one, it has the only non-zero eigenvalue given by

$$tr(J_n) = tr(\mathbf{1}\mathbf{1}') = \mathbf{1}'\mathbf{1} = n \tag{20}$$

and the eigenspace is generated by $\mathbf{1}$. Since I_n has all its eigenvalues equal to one, from Equation (17) the eigenvalues of P_n are

$$\lambda_1 = (1-\beta) + n\beta = 1 + (n-1)\beta \tag{21}$$

and

$$\lambda_2 = \lambda_3 = \cdots = \lambda_n = 1 - \beta \tag{22}$$

so that

$$|P_n| = \prod_{i=1}^{n} \lambda_i = (1 + (n-1)\beta)(1-\beta)^{n-1} \tag{23}$$

and

$$P_n^{-1} = \frac{1}{1-\beta}\left(I_n - \frac{\beta J_n}{1 + (n-1)\beta}\right) \tag{24}$$

which exists if and only if $\beta \neq 1$ and $\beta \neq -\frac{1}{n-1}$. Further, P_n is positive definite (all eigenvalues positive) if and only if $\beta \in \left(-\frac{1}{n-1}, 1\right)$. The diagonal elements of P_n^{-1} are all equal to

$$\frac{1}{1-\beta}\left(\frac{1 + (n-2)\beta}{1 + (n-1)\beta}\right) \tag{25}$$

Each diagonal element of P_n^{-1} is related to the proportion of variation in the corresponding variable explained by regressing on the remaining variables (Whittaker, 1990). More precisely, each diagonal element equals

$$\frac{1}{1-R^2} \tag{26}$$

where $R^2 (0 \leq R^2 \leq 1)$ is the coefficient of multiple determination and R is the multiple correlation coefficient between the variable and the rest. In other words, R^2 is the proportion of variation explained by the regression. The upper bound of R^2 is achieved when the fit is perfect (all residuals zero). As R^2 approaches one, the variable is most predictable whereas as it approaches zero, the variable is least predictable given the rest.

Equating (25) and (26), we obtain

$$R^2 = \left(\frac{(n-1)\beta^2}{1 + (n-2)\beta}\right) \tag{27}$$

so that

$$R^2 \rightarrow \begin{cases} 0, & random \\ 1, & contagious \end{cases} \tag{28}$$

Thus, for a random distribution, a variable given the rest is least predictable whereas for a contagious distribution, a variable given the rest is most predictable. Intuitively, for a random population, individuals are independent of each other, hence most unlikely to locate one given the rest whereas for an overdispersed population, individuals are clustered (there is interaction or interdependence between individuals), hence most likely to locate one given the rest.

The inverse correlation matrix is now scaled to have unit entries on the diagonal in the same way as a covariance matrix is scaled to give a correlation matrix and we obtain

$$Scaled\ P_n^{-1} = I_n - \frac{\beta}{1+(n-2)\beta}(J_n - I_n). \tag{29}$$

The off diagonal elements of this scaled inverse correlation matrix are the negatives of the partial correlation coefficients between the corresponding pair of variables given the remaining variables (Whittaker, 1990).

From Equation (29),

$$Scaled\ P_n^{-1} \to \begin{cases} I_n, & random \\ I_n - \dfrac{1}{n-1}(J_n - I_n), & contagious \end{cases} \tag{30}$$

so that the partial correlation coefficients between the corresponding pair of variables given the remaining variables is

$$Partial\ Corr(X_i, X_j) \to \begin{cases} 0, & random \\ \dfrac{1}{n-1}, & contagious \end{cases} \tag{31}$$

and together with Equation (16), supports the result in Equation (28). Further, for a contagious distribution the partial correlation coefficient between a pair of variables given the remaining variables diminishes to zero as the number of sampling units increases.

FromEquation (23), we observe that P_n is singular when $\beta = 1$ and $\beta = -\frac{1}{n-1}$, that is, when

$$\beta = \begin{cases} Corr(X_i, X_j), & contagious \\ -\ Partial\ Corr(X_i, X_j), & contagious \end{cases} \tag{32}$$

and hence P_n^{-1} does not exist for a contagious distribution. Further, P_n is positive semi-definite (all eigenvalues non-negative) if and only if $\beta \in \left[-\frac{1}{n-1}, 1\right]$.

3. PRINCIPAL COMPONENT ANALYSIS OF P_n

The central idea of principal component analysis is to reduce the dimensionality of a data set in which there are a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. This reduction is achieved by transforming to a new set of variables, the principal components, which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all the original variables.

If a set of variables $X = (X_1, X_2, \ldots, X_n)', n > 2$ has substantial correlations among them, then the first principal component accounts for as much of the variability in the variables as possible, and each succeeding component accounts for as much of the remaining variability as possible. If q $(< n)$ principal components account for most of the variation in X in the n-dimensional space (one axis per variable), the principal component analysis gives us a lower-dimensional picture of this object described by the principal components in a $q-$dimensional subspace.

Principal component analysis can be used to divide variables into groups. The connection with principal component analysis is that when the variables fall into well-defined clusters, there will then be one high-variance principal component, and, except in the case of 'single-variable' clusters, one or more low variance principal component(s) associated with each cluster of variables (Jolliffe, 1986).

Computation of principal components is reduced to the solution of an eigenvalue-eigenvector[3] problem for a positive semi-definite symmetric matrix so that eigenvectors are normalized and orthogonal. For a group of r equal eigenvalues, the corresponding r eigenvectors span a certain unique r-dimensional space, but, within this space, they are, a part from being orthogonal to one another, arbitrary.

Principal components of standardized variables $Z_i = \dfrac{X_i - m}{\sqrt{m/(1-\beta)}}, i = i = 1, 2, 3, \ldots, n$ are obtained from the eigenvectors of P_n. P_n is positive since β is positive and the largest eigenvalue is given by Equation (21) with associated eigenvector

$$e_1' = \frac{1}{\sqrt{n}}(1,1,\ldots,1). \tag{33}$$

The largest eigenvalue is unique, since Perron-Frobenius theorem states that a positive matrix has a unique largest real and positive eigenvalue and the corresponding eigenvector has strictly positive components (Gantmacher, 1959). The remaining $n - 1$ eigenvalues are given by Equation (22) and one choice of their eigenvectors is

---

[3] The eigenvalue-eigenvector pair derived from the covariance matrix is, in general, not the same as the ones derived from the correlation matrix.

$$
\begin{aligned}
e_2' &= \frac{1}{\sqrt{2}}(1,-1,0,\dots,0),\\
e_3' &= \frac{1}{\sqrt{6}}(1,1,-2,0,\dots,0),\\
&\vdots\\
e_i' &= \frac{1}{\sqrt{(i-1)i}}(1,\dots,1,-(i-1),0,\dots,0),\\
&\vdots\\
e_n' &= \frac{1}{\sqrt{(n-1)n}}(1,\dots,1,-(n-1)).
\end{aligned}
\tag{34}
$$

The first principal component

$$Y_1 = e_1'Z = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i \tag{35}$$

is proportional to the sum of the n standardized variables and is unique since principal components are necessarily orthogonal. This principal component explains a proportion

$$\frac{\lambda_1}{n} = \beta + \frac{1-\beta}{n} \tag{36}$$

of the total variation. For an equicorrelation matrix, the first principal component for the original variables[4], X, is the same (Johnson and Wichern, 2002), that is,

$$Y_1 = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i, \tag{37}$$

a measure of total size and it explains the same proportion (36) of total variance.

From Equations (21) and (22), we have for a random distribution

$$\lim_{\beta\to 0}\frac{\lambda_i}{n} = \frac{1}{n}, \quad i = 1,2,3,\dots,n \tag{38}$$

and for a contagious distribution

---

[4] The eigenvalues are of course different since the covariance matrix of X is $\Sigma_n = \frac{m}{1-\beta}P_n$ and in this case $\lambda_1 = \frac{m}{1-\beta}(1+(n-1)\beta)$ and $\lambda_2 = \lambda_3 = \cdots = \lambda_n = m$.

$$\lim_{\beta \to 0} \frac{\lambda_1}{n} = 1 \tag{39}$$

that is, for a random distribution, each component explains $(100/n)\%$, whereas for a contagious distribution, the first component explains 100%.

Thus, for a contagious distribution, the first principal component accounts for all the variation in Z (and in X) in the n-dimensional space and hence gives us a lower-dimensional picture of this object described by the first principal component in a 1-dimensional subspace resulting in a 'single variable' cluster. Intuitively, instead of all individuals occupying the n sampling units, they occupy one sample unit, reminiscent of overdispersion. On the other hand, for a random distribution, the n-dimensional space cannot be described by principal components in a lower-dimensional subspace, that is, no clustering, reminiscent of randomness.

CONCLUSIONS

The negative binomial parameter k is a measure of dispersion so that $k \to 0$ $(\beta \to 1)$ describes a contagious distribution and $k \to \infty$ $(\beta \to 0)$ describes a random distribution, that is, small values of k are associated with overdispersion, whereas large values of k are associated with randomness.

REFERENCES

1. Anscombe, F. J. The statistical analysis of insect counts based on the Negative Binomial Distribution; Biometrics; 5, 1949, pp. 165-174.
2. Gantmacher, F. R. Application of the Theory of Matrices; Interscience Publishers, Inc; New York, 1959.
3. Johnson, R. A. and Wichern, D. W. Applied Multivariate Statistical Analysis; Pearson Education, Inc, 2002.
4. Jolliffe, I. T. Principal Component Analysis; Springer-Verlag New York Inc, 1986.
5. Pielou, E. C. Mathematical Ecology; A Wiley Interscience Publication, 1978.
6. Southwood, T. R. E. Ecological Methods; Methuen and Co. Ltd., 1966.
7. Whittaker, J. Graphical Models in Applied Multivariate Statistics; John Willey and Sons, 1990.
8. Young, L. J. and Young, J. H. A spatial view of the Negative Binomial parameter k when describing insect populations; Proceedings of the 1990 Kansas State University Conference on Applied Statistics in Agriculture, 1990.