BETA-NORMAL DISTRIBUTION: BIMODALITY PROPERTIES AND APPLICATION

Felix Famoye, Department of Mathematics, Central Michigan University, Mt. Pleasant, Michigan 48859
Carl Lee, Department of Mathematics, Central Michigan University, Mt. Pleasant, Michigan 48859
Nicholas Eugene, Department of Computer Science & Math, Coppin State College, Baltimore, MD 21216

Key Words: Logit; parameter space; curve estimation; likelihood ratio test.

ABSTRACT

The beta-normal distribution is characterized by four parameters that jointly describe the location, the scale and the shape properties. The beta-normal distribution can be unimodal or bimodal. This paper studies the bimodality properties of the beta-normal distribution. The region of bimodality in the parameter space is obtained. A test for bimodality is proposed for the distribution. The beta-normal distribution is applied to fit a bimodal numerical data set.

1. Introduction

Bimodal distributions occur in many areas of science. Withington et al. (2000), in their study of cardiopulmonary bypass in infants showed that plasma vecuronium and vecuronium clearance requirements have bimodal distributions. They concluded that their findings on bimodal distributions for plasma vecuronium and vecuronium clearance requirements highlight the need for individual monitoring of neuromuscular blockade. Espinoza et al. (2001) discussed the importance of bimodal distributions in the study of size distribution of metals in aerosols. Bimodal distributions also occur in the study of genetic diversity (Freeland et al., 2000), in the study of agricultural farm size distribution (Wolf and Sumner, 2001), in the study of atmospheric pressure (Zangvil et al., 2001), and in the study of anabolic steroids on animals (Isaacson, 2000).

Let $F(x)$ be the cumulative distribution function (CDF) of a random variable *X*. The cumulative distribution function for a generalized class of distributions for the random variable *X* can be defined as the logit of the beta random variable given by

$$G(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^{F(x)} t^{\alpha-1}(1-t)^{\beta-1}\,dt, \qquad (1.1)$$
$$0 < \alpha,\ \beta < \infty.$$

Eugene et al. (2002) considered $F(x)$ as the CDF of the normal distribution with parameters μ and σ. Thus, the random variable *X* has the beta-normal distribution with probability density function (pdf)

$$BN(\alpha,\beta,\mu,\sigma) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\left[\Phi(\frac{x-\mu}{\sigma})\right]^{\alpha-1} \cdot\left[1-\Phi(\frac{x-\mu}{\sigma})\right]^{\beta-1}\sigma^{-1}\phi(\frac{x-\mu}{\sigma}), \qquad (1.2)$$

where $\phi\left(\frac{x-\mu}{\sigma}\right)$ is the normal pdf and $\Phi\left(\frac{x-\mu}{\sigma}\right)$ is the normal CDF.

The distribution in (1.2) may be symmetric, skewed to the left, or skewed to the right. The distribution may be unimodal or bimodal. Eugene et al. (2002) discussed the shape properties of the unimodal beta-normal distribution. Furthermore, they considered the estimation of its parameters by the method of maximum likelihood.

In the analysis of bimodal data, a mixture of two normal densities is usually used as a model (e.g., Cobb et al., 1983). The mixture of normal distribution is used as a model to analyze bimodal data because the mixture of normal densities can take on bimodal shapes depending on the parameters of the distribution. Eisenberger (1964) showed how the parameters of a mixture of normal distributions determine its shape. When a mixture assumption is not required or justified the beta-normal distribution can serve as a model to analyze data since only one distribution has to be used and one less parameter to estimate.

In section 2, we provide some bimodality properties of beta-normal distribution. In section 3, we obtain the region of bimodality in the parameter space. A likelihood ratio test is developed to test for bimodality in section 4. In section 5, we illustrate the application of beta-normal distribution to a numerical data set that exhibit two modes.

2. Bimodality Properties

In this section, we provide some results on the bimodality properties of beta-normal distribution.

Theorem: The mode(s) of $BN(\alpha,\ \beta,\ \mu,\ \sigma)$ is at the point $x_0 = x_0(\alpha,\ \beta)$ that satisfies

$$x_0 = \frac{\sigma\phi(\frac{x_0-\mu}{\sigma})}{1-\Phi(\frac{x_0-\mu}{\sigma})}\{2-\alpha-\beta\} + \frac{(\alpha-1)\phi(\frac{x_0-\mu}{\sigma})\sigma}{\Phi(\frac{x_0-\mu}{\sigma})[1-\Phi(\frac{x_0-\mu}{\sigma})]} + \mu\,. \qquad (2.1)$$

Proof: Differentiating $BN(\alpha,\ \beta,\ \mu,\ \sigma)$ in (1.2) with respect to *x*, setting it equal to zero, and solving for *x* gives the result in (2.1).

Corollary 1: If $\alpha = \beta$ and one mode of $BN(\alpha,\ \beta,\ \mu,\ \sigma)$ is at x_0, then the other mode is at the point $2\mu - x_0$.

Proof: If $BN(\alpha,\ \beta,\ \mu,\ \sigma)$ is unimodal, then the only mode occurs at the point $x_0 = \mu$. For bimodal case, we need to show that if we replace x_0 with $2\mu - x_0$, then equation (2.1) remains the same. When $\alpha = \beta$, equation (2.1) becomes

$$x_0 = \frac{\sigma\phi(\frac{x_0-\mu}{\sigma})(\alpha-1)}{\Phi(\frac{x_0-\mu}{\sigma})[1-\Phi(\frac{x_0-\mu}{\sigma})]} .\left\{1-2\Phi(\frac{x_0-\mu}{\sigma})\right\} + \mu\,. \qquad (2.2)$$

If x_0 in (2.2) is replaced with $2\mu - x_0$, we obtain

$$2\mu - x_0 = \frac{\sigma\phi(\frac{\mu-x_0}{\sigma})(\alpha-1)}{\Phi(\frac{\mu-x_0}{\sigma})[1-\Phi(\frac{\mu-x_0}{\sigma})]} .\left\{1-2\Phi(\frac{\mu-x_0}{\sigma})\right\} + \mu\,. \qquad (2.3)$$

By using $\Phi\left(\frac{x_0-\mu}{\sigma}\right) = 1-\Phi\left(\frac{\mu-x_0}{\sigma}\right)$ and $\phi\left(\frac{x_0-\mu}{\sigma}\right) = \phi\left(\frac{\mu-x_0}{\sigma}\right)$ in (2.3) and on simplification, we get the result in (2.2).

Corollary 2: If $BN(\alpha,\ \beta,\ \mu,\ \sigma)$ has a mode at x_0, then $BN(\alpha,\ \beta,\ \mu,\ \sigma)$ has a mode at $2\mu - x_0$.

Proof: We need to show that if we replace α with β, and $2\mu - x_0$ with x_0, equation (2.1) remains the same. Equation (2.1) can be written as

$$x_0 = \frac{\sigma\phi(\frac{x_0-\mu}{\sigma})}{\Phi(\frac{x_0-\mu}{\sigma})[1-\Phi(\frac{x_0-\mu}{\sigma})]} .\left\{(2-\alpha-\beta)\Phi(\frac{x_0-\mu}{\sigma}) + (\alpha-1)\right\} + \mu\,. \qquad (2.4)$$

If x_0 is replaced with $2\mu - x_0$ and α is replaced with β in (2.4), using $\Phi\left(\frac{x_0-\mu}{\sigma}\right) = 1-\Phi\left(\frac{\mu-x_0}{\sigma}\right)$ and $\phi\left(\frac{x_0-\mu}{\sigma}\right) = \phi\left(\frac{\mu-x_0}{\sigma}\right)$, and on simplification, we get back the result in (2.4).

Corollary 3: The modal point $x_0(\alpha,\ \beta)$ is an increasing function of α and a decreasing function of β.

Proof: Differentiating the result in (2.1) with respect to α and β gives

$$\frac{\partial x_0(\alpha,\beta)}{\partial\alpha} = \frac{\sigma\phi(\frac{x_0-\mu}{\sigma})}{\Phi(\frac{x_0-\mu}{\sigma})} > 0$$

and

$$\frac{\partial x_0(\alpha,\beta)}{\partial\beta} = \frac{-\sigma\phi(\frac{x_0-\mu}{\sigma})}{1-\Phi(\frac{x_0-\mu}{\sigma})} < 0\,.$$

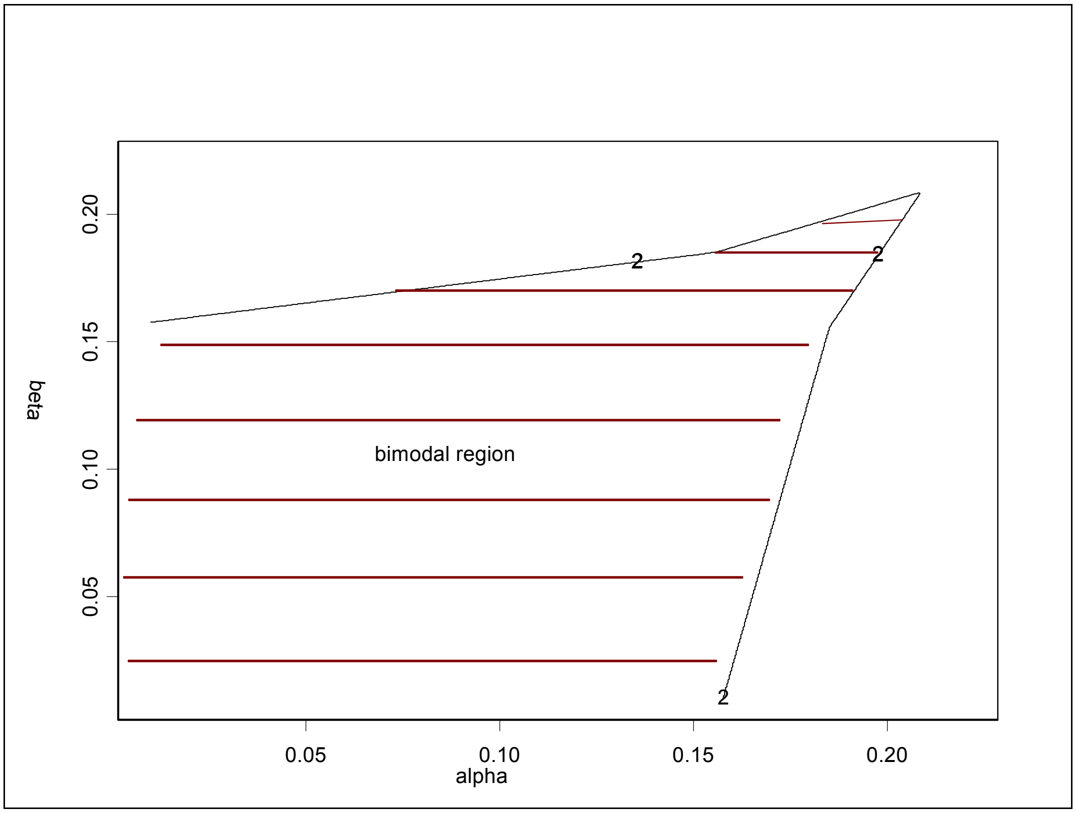
Hence $x_0(\alpha,\ \beta)$ is an increasing function of α and a decreasing function of β.

Eugene et al. (2002) showed that the beta-normal distribution is symmetric about μ when $\alpha = \beta$. From this result and corollary 3, the modal value is greater than μ if $\alpha > \beta$. Also, the modal value is less than μ if $\alpha < \beta$. The beta-normal distribution has a very distinct property in that it can be used to describe both bimodal and unimodal data.

3. Region of Bimodality

The beta-normal distribution becomes bimodal for certain values of the parameters α and β, and the analytical solution of α and β where the distribution becomes bimodal cannot be solved algebraically. A numerical solution is obtained, however, by solving the number of roots of the derivative of $BN(\alpha, \beta, \mu, \sigma)$. Numerically, the largest value of α or β that gives bimodal property is approximately 0.214. Figure 1 shows a plot of the boundary region of α and β values where $BN(\alpha, \beta, 0,1)$ is bimodal.

Figure 1. Plot of bimodal region for beta-normal distribution $BN(\alpha, \beta, 0, 1)$



Corollary 4: The bimodal property of $BN(\alpha, \beta, \mu, \sigma)$ is independent of the parameters μ and σ.
Proof: The mode(s) of $BN(\alpha, \beta, \mu, \sigma)$ is at the point $x_0 = x_0(\alpha, \beta)$ given in (2.1). On re-writing (2.1), one obtains (2.4). On taking the μ on the right hand side of (2.4) to the left hand side, dividing through by σ, and replacing $(x_0 - \mu)/\sigma$ by z_0, one obtains

$$z_0 = \frac{\phi(z_0)}{\Phi(z_0)[1-\Phi(z_0)]}\left\{(2-\alpha-\beta)\Phi(z_0)+(\alpha-1)\right\}, \tag{2.5}$$

which is independent of parameters μ and σ.

In corollary 4, we showed that the bimodal property of $BN(\alpha, \beta, \mu, \sigma)$ is robust against the parameters μ and σ. In other words, regardless of the values of μ and σ, the α and β range for the bimodality of $BN(\alpha, \beta, \mu, \sigma)$ remains the same. To get more accurate values of the pairs of (α, β) values that lie on the boundary of the region where the beta-normal distribution becomes bimodal, regression lines were drawn to estimate each boundary. The regression line that traced the boundaries of Figure 1 was approximated using curve estimation. For the values of α in the interval [0.01, 0.16), the values of β at the upper boundary in Figure 1 were estimated by $\hat{\beta} = 0.8591\alpha^2 + 0.0453\alpha + 0.1603$. For α in the interval [0.16, 0.214] we estimated β values by $\hat{\beta} = 4.4113\alpha^2 - 1.1966\alpha + 0.2675$. For the values of α in the interval [0.16, 0.189), the values of β at the lower boundary in Figure 1 were estimated by $\hat{\beta} = -116.15\alpha^2 + 45.4657\alpha - 4.2908$. For α in the interval [0.189, 0.214] we use equation $\hat{\beta} = -41.972\alpha^2 + 18.9913\alpha - 1.9281$ to estimate the value of β.

If $BN(\alpha, \beta, \mu, \sigma)$ is unimodal, the distribution is skewed to the right whenever $\alpha > \beta$ and it is skewed to the left whenever $\alpha < \beta$. If $BN(\alpha, \beta, \mu, \sigma)$ is bimodal, the distribution is skewed to the right when $\alpha < \beta$ and it is skewed to the left when $\alpha > \beta$. Thus, the beta-normal distribution provides great flexibility in modeling symmetric, skewed and bimodal distributions.

4. Test for Bimodality

A likelihood ratio test statistic is developed to test the bimodal beta-normal distribution against a unimodal alternative. The beta-normal distribution in (1.2) is bimodal if the parameters α and β fall in the bimodal region (which is shaded) in Figure 1. Thus, to test for bimodality we test the hypothesis

H_0 : Values of α and β are outside the bimodal region
$H_a : H_0$ is not true.

$$\tag{4.1}$$

Let $\omega = (\alpha,\ \beta,\ \mu,\ \sigma)$ be the set of beta-normal distribution parameters where both α and β lie inside the bimodal region in Figure 1. Under H_0, the set of beta-normal distribution parameters will be given by the complement ω'. Let the whole parameter space for beta-normal density be given by $\Omega = \omega \cup \omega'$. We define the likelihood ratio statistic

$$\lambda = \frac{\max\limits_{(\alpha,\beta,\mu,\sigma)\in\omega'} L_x}{\max\limits_{(\alpha,\beta,\mu,\sigma)\in\Omega} L_x} \tag{4.2}$$

where L_x is the likelihood function for beta-normal distribution. The numerator in (4.2) is maximized over α, β, μ, and σ outside the shaded region in Figure 1. In this maximization, two restrictions (one for α and the other for β) are imposed on the parameter space. The denominator in (4.2) is maximized over α, β, μ, and σ by using the whole parameter space Ω. Under very general conditions, the quantity –2log(λ) has an approximate chi-square distribution with 2 degrees of freedom. This enables us to test the hypothesis in (4.1) at any given level of significance.

5. Application to Bimodal Data

Sewell and Young (1997) studied the egg size distributions of echinoderm. In marine invertebrates, a species produces either many small eggs with planktotrophic development or fewer larger eggs with lecithotrophic development, Thorson (1950). The models developed by Vance (1973a, 1973b) viewed planktotrophy and lecithotrophy as extreme forms of larvae development. Subsequent modifications of these models [see the references in Sewell and Young (1997)] predict that eggs of marine invertebrates have bimodal distributions. Christiansen and Fenchel (1979) reported a bimodal distribution of egg sizes within prosobranchs. Emlet et al. (1987) described bimodal distributions in asteroid and echinoid echinoderms. For echinoids and asteroids [see Tables 2 and 7 of Emlet et al. (1987)], the egg diameters for species with planktotrophic larvae have less variation than species with lecithotrophic larvae. Because of this variation, the egg diameters appear to have one mode. However, with logarithmic transformation, the effect of large eggs in lecithotrophic species is reduced and the distribution of eggs becomes bimodal for both echinoids and asteroids.

Sewell and Young (1997) reported that many of the early studies used data sets that were not appropriate for a valid test of the egg size distribution patterns. They defined three criteria for appropriate data sets. The most widely cited example of bimodality in egg sizes is the data set compiled by Emlet et al. (1987). This data set satisfied the three criteria defined by Sewell and Young.

Sewell and Young (1997) reexamined the asteroid and echinoid egg size data in Emlet et al. (1987) with some additional data from more recent study. The additional data used by Sewell and Young were not available in their published paper. In this paper, we have applied the beta-normal distribution to fit the logarithm of the egg diameters of the asteroids data in Emlet et al. (1987). The valid data consists of 88 asteroid species divided into three types consisting of 35 planktotrophic larvae, 36 lecithotrophic larvae, and 17 brooding larvae. These species are from a variety of habitats.

The maximum likelihood estimation method is used for parameter estimation. Eugene et al. (2002) gave the detailed discussion of this estimation technique. The parameter estimates for beta-normal distribution are $\hat{\alpha}$ = 0.013, $\hat{\beta}$ = 0.007, $\hat{\mu}$ = 5.747, and $\hat{\sigma}$ = 0.068. The estimates for α and β fall in the bimodal region in Figure 1. The log-likelihood value is –109.48. A histogram of the data with the beta-normal distribution super imposed is presented in Figure 2. We checked the goodness of fit of beta-normal distribution to the data by using the Kolmogorov-Smirnov test [see the book by DeGroot and Schervish (2002), page 568]. In Figure 3, we present both the empirical CDF and the beta-normal CDF for the data. The absolute maximum difference between the empirical cumulative distribution function and the beta-normal cumulative distribution function is D_n^* = 0.1244. This provides a test statistic $\sqrt{n}D_n^*$ = 1.167 with a significance probability of 0.1310. Thus, the beta-normal distribution provides an adequate fit to the data.

We tested the hypothesis in (4.1) about the bimodality of the asteroids species data. The log-likelihood statistic under the null hypothesis H_0 in (4.1) is –118.61. The likelihood ratio statistic in (4.2) is 18.26 with 2 degrees of freedom. The corresponding p-value for this test is 0.0001. Thus, the null hypothesis is rejected and hence the data is bimodal.

Figure 2. Histogram of log(egg diameter) with beta-normal super imposed

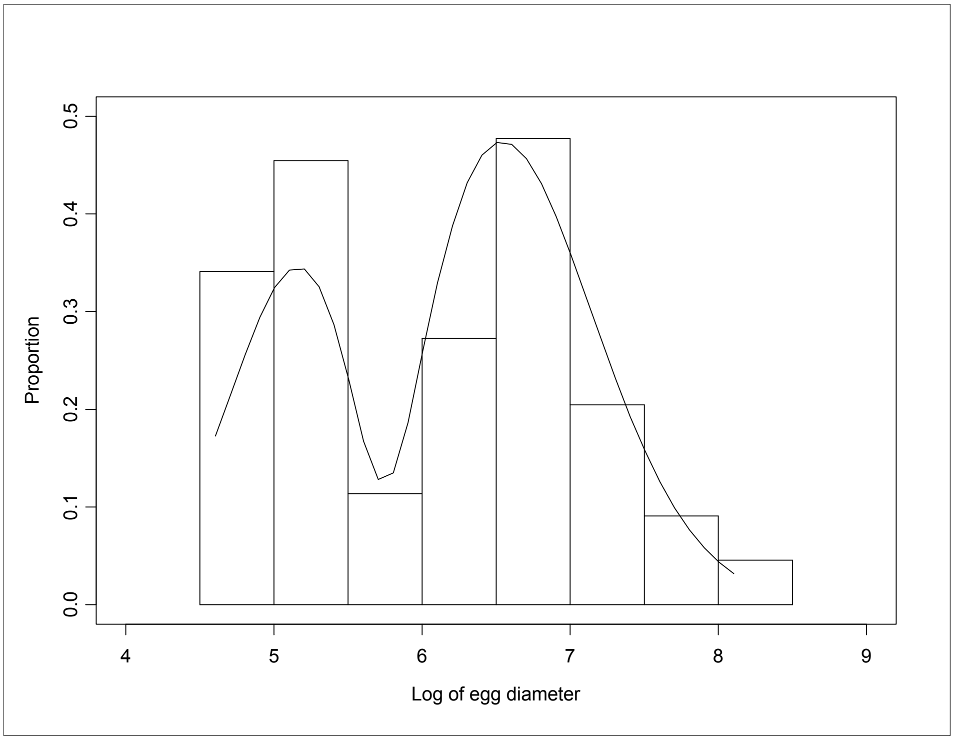
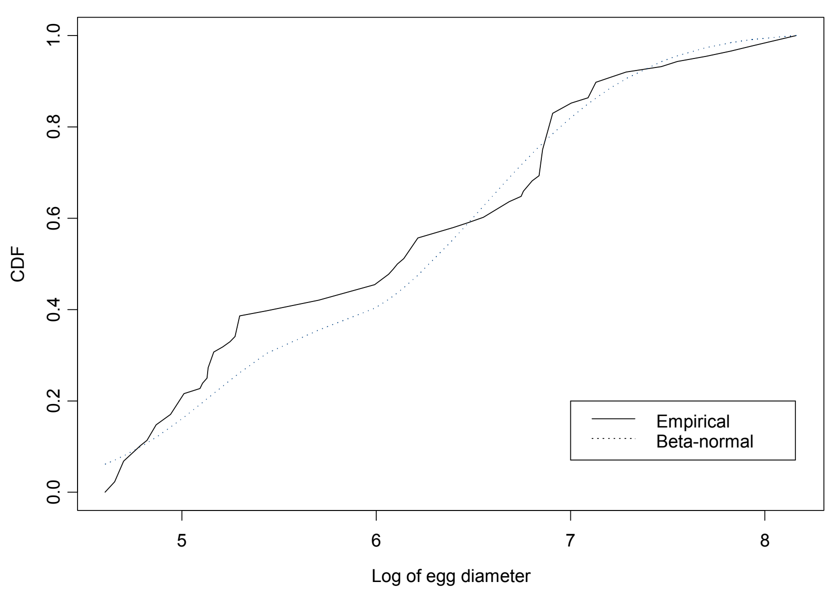


Figure 3. Empirical CDF and beta-normal CDF for log(egg diameter)



REFERENCES

Christiansen, F.B. and Fenchel, T.M., 1979: Evolution of marine invertebrate reproductive patterns. *Theor. Popul. Biol*. **16**, 267-282.

Cobb, L., Koppstein, P. and Chen, N.H., 1983: Estimation and Moment Recursion for Multimodal Distributions of the Exponential Family, *JASA*, Vol. **78**, 124-130.

DeGroot, M.H. and Schervish, M.J., 2002: *Probability and Statistics*, 3rd ed. Addison Wesley.

Eisenberger, I., 1964: Genesis of bimodal distributions, *Technometrics*, Vol. **6**, 357-363.

Emlet, R.B., McEdward, L.R. and Strathmann, R.R., 1987: Echinoderm larval ecology viewed from the egg. pp. 55-136 in Echinoderm Studies, Vol. **2**, M. Jangoux and J.M. Lawrence, eds. A.A. Balkema, Rotterdam.

Espinosa, A., Rodriguez, M., de la Rosa, F. and Sanchez, J., 2001: Size of metals in urban aerosol in Serville, *Atmospheric Environment*, Vol. **35**, 2595-2601.

Eugene, N. Lee, C. and Famoye, F., 2002: Beta-normal distribution and its applications. *Communications in Statistics: Theory and methods*, 31(4), 497-512.

Freeland, J.R., Romualdi, C. and Okamura, B., 2000: Gene flow and genetic diversity; a comparison of freshwater bryozoan populations in Europe and North America, *Heredity*, Vol. **85**, 498-508.

Isaacson, R.L., 2000: Non linear effects in the retention of an avoidance task induced by anabolic steroids, *European Journal of Pharmacology*, Vol. **405**, 177-186.

Sewell, M.A. and Young, C.M., 1997: Are echinoderm egg size distributions bimodal? *Biol. Bull*. **193**, 297-305.

Thorson, G., 1950: Reproductive and larval ecology of marine bottom invertebrates. *Biol. Rev.* **25**, 1-45.

Vance, R.R., 1973a: On reproductive strategies in marine benthic invertebrates. *Am. Nat.* **107**, 339-352.

Vance, R.R., 1973b: More on reproductive strategies in marine benthic invertebrates. *Am. Nat.* **107**, 353-361.

Withington, D., Menard, G., Harris, J., Kulkarni, P., Donati, F. and Varin, F., 2000: Vecuronium pharmaco-kinetics and pharmacodynamics during hypothermic cardiopulmonary bypass in infants and children, *Canadian Journal of Anesthesia*, Vol. **47**, 1188-1195.

Wolf, C.A. and Sumner, D.A., 2001: Are farm size distributions bimodal? Evidence from kernel density estimates of dairy farm size distributions, *American Journal of Agricultural Economics*, Vol. **83**, 77-88.

Zangvil, A., Portis, D.H. and Lamb, P.J., 2001: Investigation of the large-scale atmospheric moisture field over the Midwestern United States in relation to summer precipitation. Part I: Relationships between moisture budget components on different timescales, *Journal of Climate*, Vol. **14**, 582-597.