STATISTICAL INFERENCE PROCEDURE BY THE INFORMATION-BASED TEST AND ITS APPLICATION IN MARINE CLIMATOLOGY

SEZER, A. – ASMA, S. – OZDEMIR, O.*

Department of Statistics, Faculty of Science, Anadolu University, 26470 Eskisehir, Turkey (phone: +90- 222-335-0580; fax: +90-222-320-4910)

**Corresponding author*
e-mail: ozerozdemir@anadolu.edu.tr

(Received 9th Oct 2017; accepted 13th Mar 2018)

Abstract. Objectives of ecological studies should drive all aspects of design. These objectives must include hypothesis testing and appropriate sample size. High level of unexplained variation is typical in many ecological studies and may lead to incorrect inference about the population. Choosing appropriate sample size is one key strategy to cope with unexplained variation. In this study, we aim to determine sample size which depends upon the information-based test and show the superiority of this approach over the likelihood ratio test. Particularly, we focused on finding appropriate sample size for testing the variance of the normal distribution and Rayleigh distribution. The power curves are obtained both for information-based test and the likelihood ratio test by the Monte Carlo simulations. We used wave height data to show how the inference procedure should follow both for the likelihood test and information-based test procedure. In agreement with the theoretical results of Janssen (2014) it is found that wave height obeys the Rayleigh distribution. Sample size determination for testing the variance of the normal distribution and Rayleigh distribution with different parameters is demonstrated for the fixed effect sizes.
Keywords: *likelihood ratio test, effect size, sample size, Monte Carlo simulation, power curves*

Introduction

Sample size determination is one of the most essential parts of the statistical design in the ecological studies and it is usually a difficult one. The ecologists, first want to know how many subjects should be included in their study (sample size) and how these subjects should be selected (sampling methods). Second, they desire to attribute a p-value to their results to claim significance of results. In ecological point of view, choosing appropriate sample size is essential to develop statistical distributions, used for predicting the occurrence of a species at a particular location (Wood et al., 2015; Kass et al., 2016). The availability of novel and sophisticated statistical techniques means we are better equipped than ever to extract signal from noisy ecological data, but it remains challenging to know how to apply these tools, and which statistical technique(s) might be best suited to answering specific questions (Low-Decarie et al., 2014; Zuur and Ieno, 2016).

Lynch (2017) has shown the principle of minimizing the cost-plus-loss is used to simultaneously find optimal plot size and sample size for forest sampling using the Fairfield Smith relationship between plot size and the variance among cubic metre volumes per hectare. Indeed, reducing the number of animal subjects used in biomedical experiments is desirable for ethical and practical reasons. Kramer and Font (2017) discussed how the number of current control animals can be reduced, without loss of statistical power, by incorporating information from historical controls. Dziak et al. (2014) empirically addressed the question of how large a sample size is needed to avoid

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1602_19731982

underextraction when using the bootstrap likelihood ratio test to choose a number of classes in latent class analysis.

Once the basic study design and planned analysis methods are defined, there are three parameters under the control which define the power of the study. These are the sample size, the effect size defined by the alternative hypothesis and significance level. Considering these parameters and the power of the study together, fixing any of three will allow calculating the fourth. However there are several obstacles that we need to handle. For example how do we know the population parameter, given that we are only planning the experiment and so no data have been collected yet? In those cases, sometimes, there are historical data that can be used to estimate parameter(s) in the power function. However most of the time we may not have preliminary data from which to estimate population parameter, in this case a pilot study is needed.

In the frequentist inference, sample size calculations are made by the power fuction. Lenth (2001) made criticism for some ill-advised shortcuts relating to power and sample size. Actually, there are several approaches to determine sample size. One can specify the desired confidence interval and then determine sample size that achieves that goal. Another approach involves studying the power of a test of hypothesis. The power of a test is defined as the probability that we correctly reject the null hypothesis, given that alternative hypothesis is true.

Most of the literature about determination of sample size depends on likelihood ratio test. Recently, Kullback-Leibler information-based test have become popular as complementary approaches to classical hypothesis testing (Burnham and Anderson, 2002). Esteban et al. (2001) also gave the power comparisons of four different entropy estimators. Gupta (2007) studied testing equality of variances of observations in the different treatment groups assuming treatment effects are fixed. Akaike's information criterion based on information theory, is a common metric used by ecologist to evaluate and select among alternative ecological models (Burnham and Anderson, 2002; Richards, 2005). Applications of entropy principles to evolution and ecology are of tantamount importance given the central role spatiotemporal structuring plays in both evolution and ecological succession. Recently, Schepsmeier (2015) tested normality and introduced a new goodness-of-fit test for regular vine copula models, a flexible class of multivariate copulas based on a pair-copula construction. Vranken et al. (2015) made a review on the use of entropy in landscape ecology for heterogeneity, unpredictability, scale dependence and their links with thermodynamics. Roach et al. (2017) obtained a qualitative interpretation of the role of entropy in evolving ecological systems. Martin and Poleto (2018) showed how to apply the theory of entropy to determine sediment concentration.

It is common to test hypothesis about the population standard deviation using the likelihood ratio test (LRT). In particular, the likelihood ratio test statistic is assumed to follow a χ^2 with degrees of freedom equal to number of parameters that are tested. For the purpose of the power and sample calculation in the generalized models, two major tests have been proposed. These are the score test and the likelihood ratio statistics developed by Self and Mauritsen (1988). Lehmann (2006) pointed that the LRT agrees with tests obtained from other principles (for example it is UMP unbiased or UMP invariant). Generally LRT seems to lead to satisfactory tests. However, counter-examples are also known in which the test is quite unsatisfactory; see for example Perlman and Wu (1999) and Men'endez et al. (1992).

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1602_19731982

Self et al. (1992) described an approach for sample size/power calculations that is based on non-central chi-square approximation to the distribution of likelihood ratio statistic. Shieh (2000) has given a direct extension of the approach described in Self et al. (1992) for power and sample calculations in generalized linear models. There is also a detailed review paper of Adcock (1997) which presents key techniques on frequentist approach and several Bayesian methods.

Information-theoretic principles are commonly, used in the test literature for different purposes such as testing normality, homogeneity of variances for several populations. Unfortunately, there is not enough study about the sample size determination related with entropy based statistic. In this study, we goal to obtain appropriate sample size which depends upon the information-based test (IT) and show the superiority of this approach over the likelihood ratio test by a simulation study and real data set.

Materials and methods

Likelihood ratio test and information-based test

Likelihood ratio test

Let $X_1, X_2, \mathrm{K}, X_n$ be a random sample of n observations from a population with probability density function (pdf) $f(x|\theta)$. Let null hypothesis (H_0) specify that $\theta \in \omega_0$, where ω_0 is the subset of the set of all possible values of θ and the alternative hypothesis H_a specify that $\theta \in \omega_1$, where ω_1 is disjoint from ω_0 and parameter space $\Omega = \{\omega_0 \cup \omega_1\}$.

Assume that hypothesis are composite, each likelihood value is evaluated at that value of θ that maximizes it, yielding the generalized likelihood ratio test statistic (*Eq. 1*):

$$\Lambda = \frac{\sup_{\theta \in \omega_0} [L(\theta)]}{\sup_{\theta \in \Omega} [L(\theta)]}, \qquad \text{(Eq.1)}$$

where $L(\theta)$ denote the likelihood function of the probability distribution.

The ratio of this two maxima is small if there are parameter points in the alternative hypothesis is more likely than for any parameter point in the null hypothesis. In this case null hypothesis should be rejected and alternative hypothesis accepted as true.

Information-based test

Shannon's entropy associated with a probability density function f with respect to sigma finite measure (ν) is defined by (*Eq. 2*)

$$H(f) = -\int_{\chi} f(x) \log f(x) d\nu(x) \qquad \text{(Eq.2)}$$

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1602_19731982

where $\chi = \{x : f(x) > 0\}$ is the support set for f. Entropy is used as an index of diversity in a population.

In the case when $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right)$ is the pdf associated with a normal distribution with mean μ and variance σ^2, the entropy of the normal probability distribution is

$$H\left(N\left(\mu,\sigma^2\right)\right) = \log\sqrt{\sigma^2 2\pi e}\,. \qquad \text{(Eq.3)}$$

The entropy of normal distribution does not depend on the mean value, so the information-based statistic defined by *Eq. 3* cannot be applied to a test concerned with the mean of the population. In the present study, we define the information-based statistic as the difference of the entropy of the distribution under the null hypothesis and alternative hypothesis (*Eq. 4*),

$$I = \log\sqrt{\sigma^2 2\pi e} - \log\sqrt{\hat{\sigma}^2 2\pi e}\,, \qquad \text{(Eq.4)}$$

where $\hat{\sigma}^2$ is maximum likelihood estimator of the variance of $N\left(\mu,\sigma^2\right)$.

According to central limit theorem information-based statistic *I* approaches 0 as the sample size increases. It should be noted that presented statistic can take both negative and positive values. This means that if this statistic takes values close to zero the null hypothesis will be accepted otherwise test will be rejected. However, it is not possible to find the asymptotic distribution of this test statistic. Bootstrap sampling method will be used in order to calculate the critical values of this distribution.

Results

Example 1: a simulation study with normal distribution

In this section, first we conduct Monte Carlo simulations to show the superiority of IT over the LRT test and after that we consider the sample size determination with respect to the power level, effect sizes and significance level. We generated 10,000 random sample of size $n \geq 5$ from several normal densities. Then the information based test values are ordered $I_{n,1}, I_{n,2}\,\mathrm{K}\,, I_{n,10000}$ and since $\alpha = 0.05$ is used, the required critical values for the two sided hypothesis are found as the 2500^{th} and 7500^{th} order statistics.

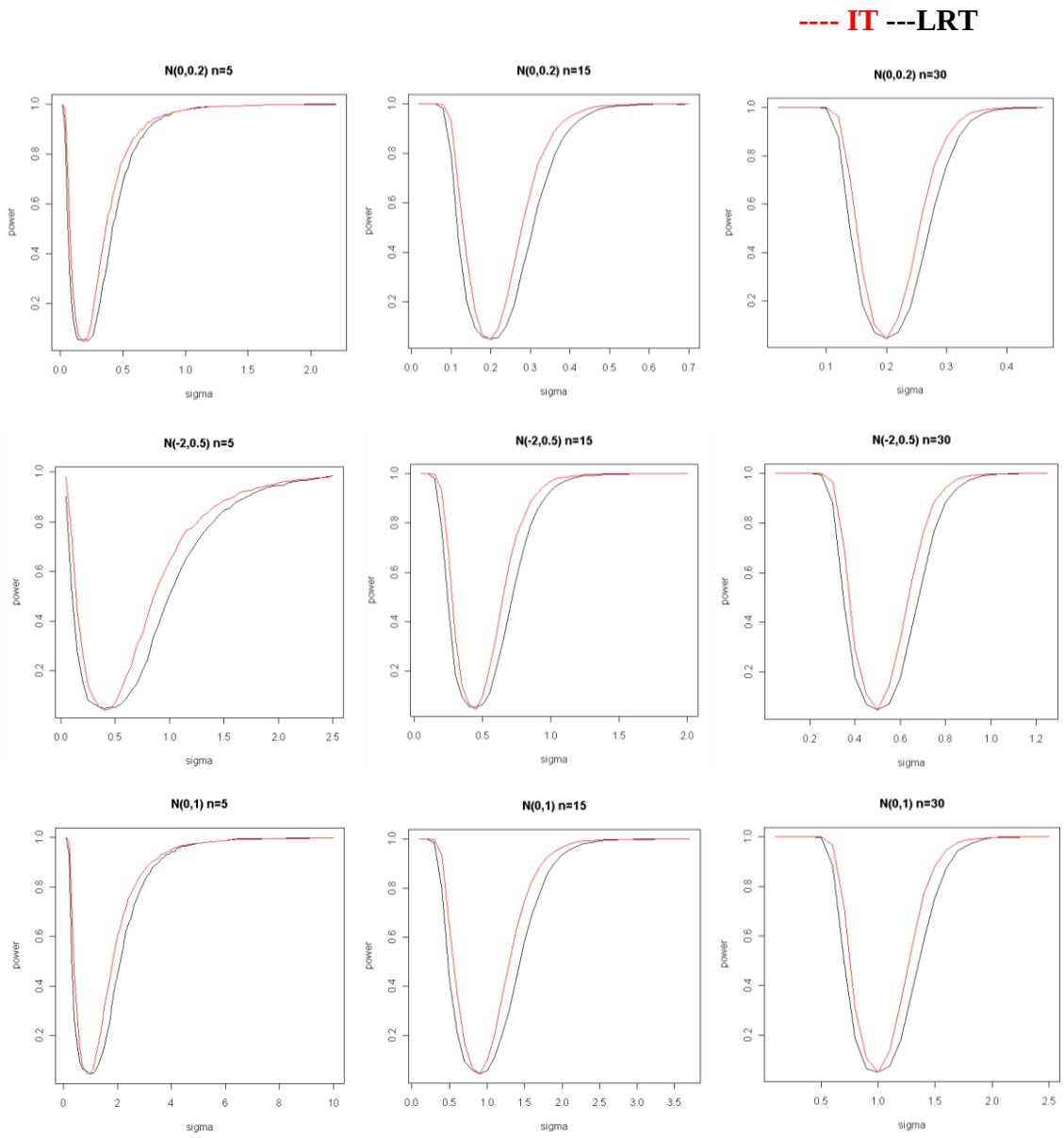
Since IT for the normal distribution depends only on the population standard deviation, we should state alternative hypothesis for the IT by considering effect size on standard deviation of the population. Suppose that we wish to test the hypothesis that the variance of a normal population σ^2 equals a specified value, say $\sigma_0{}^2$. We state the null hypothesis $H_0 : \sigma^2 = \sigma_0{}^2$, and the alternative hypothesis $H_1 : \sigma^2 = \sigma_0{}^2 + \delta$, where δ denote the effect size. The effect size δ refers to the magnitude of the effect under the alternative hypothesis. The effect size should represent the smallest difference that would be of clinical or biological significance and is usually determined by the

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1602_19731982

scientific considerations, rather than statistical ones. Obtaining an effect size of scientific importance requires meaningful input from the researcher. For example, a treatment effect that reduces failure of a disease or a treatment by 1% might be clinically important, while a treatment effect that reduces asthma symptoms by 20% may be of little clinical interest.

Although we can obtain the asymptotic distribution of LRT for testing the standard deviation of a normal distribution, the distribution of the information based test under the normality cannot be found analytically. In particular, the likelihood ratio test statistic is assumed to follow a χ^2 with degrees of freedom equal to number of parameters that are tested. However, the critical values for IT with common significance level $\alpha = 0.05$ are calculated by the Monte Carlo method.

First, we obtained the power curves for the normal distributions with the sample sizes n = 5, 15, 30. These power curves are illustrated in *Figure 1*.



http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1602_19731982

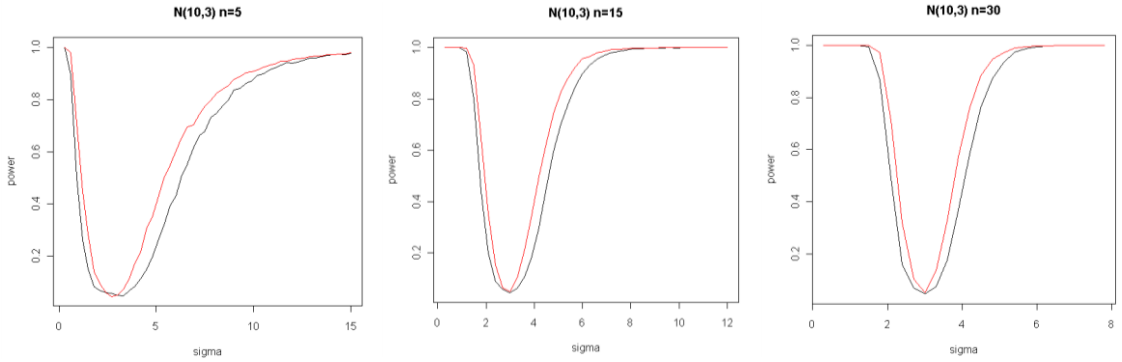
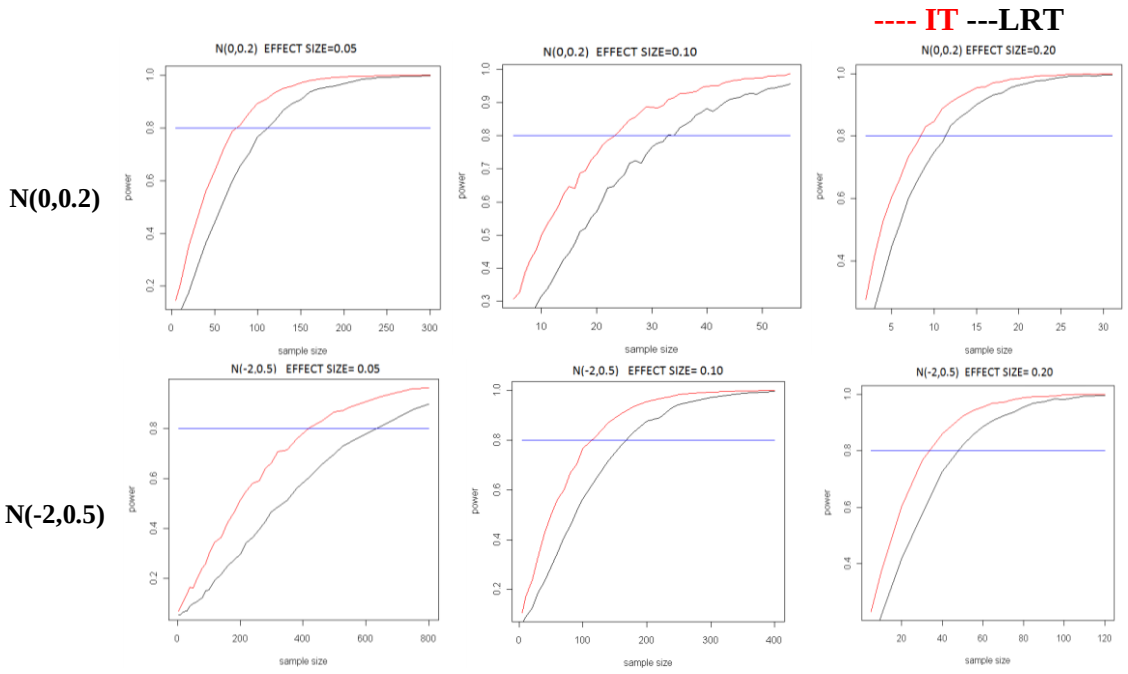


Figure 1. *Power curves for the normal distributions with the sample sizes n = 5, 15, 30*

We used these graphs to decide the appropriate effect size to obtain the required power for the design of the study. Since, it is common to choose the power between .80 and .98 in the literature, the domain of the power function corresponding to these points provide us an idea for the range of δ values. From this range, one may specify the δ values according to the ecological importance. For instance, for N(0,1) with sample size 30, the δ interval that corresponds to power between .80 and .98 is [1.6-2.0]. This means that δ can be chosen from the interval [0.6-1.0] to attain the power between .80 and .98. Indeed, we take the effect size δ as 0.6, 0.7, 0.8 to demonstrate the change in the power with respect to sample size (*Fig. 2*).

In a similar way, we choose the effect size δ as 0.05, 0.10, 0.20 both for N(0,0.2) and N(-2,0.5) and as 2.6, 2.7, 2.8 for N(10,3).

When we graphed the power versus sample size with different effect sizes, it can be seen that both IT and LRT curves are monotonically increasing function of the sample size. *Figure 2* is also helpful for the researchers to decide the appropriate sample size for fixed power and effect size.



APPLIED ECOLOGY AND ENVIRONMENTAL RESEARCH 16(2):1973-1982.
http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1602_19731982

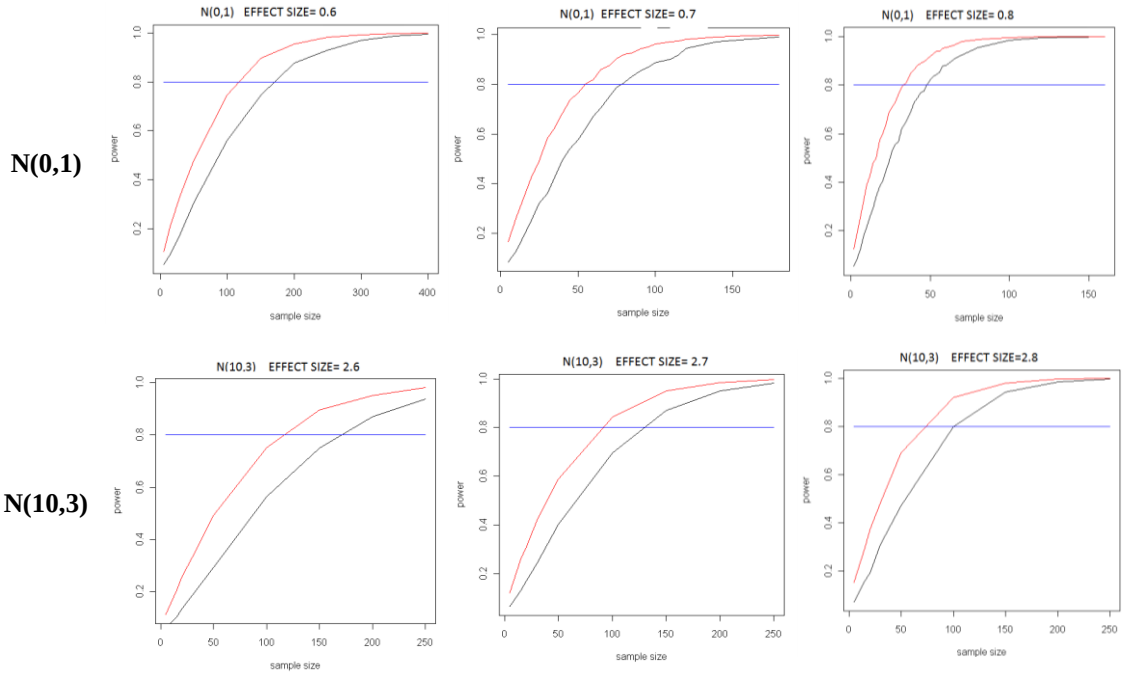


Figure 2. *Power per sample of size n with different effect sizes for both IT and LRT curves*

Figure 2 provides appropriate sample sizes for the data that are simulated from N(0,0.2), N(-2,0.5), N(0,1) and N(10,3) respectively. For example, for the N(0,0.2) when the effect size is given 0.05, we attain 80% power for the LRT with sample size is around 100. On the other hand with same effect size, and sample size we obtain power around 90% for the IT test. However to attain 90% power with LRT test we need sample size around 150 by the likelihood ratio test.

When we increase the effect size to the 0.10, we attain 80% power for the LRT with the sample size is around 35. However, with the same effect size and same sample size we obtain power around 90% for the IT test. Clearly, for small effect sizes we need larger sample sizes to obtain the same power.

Example 2: a study with real data by Rayleigh distribution

The deep water buoy (15.5°N, 69.25°E) recorded daily significant wave height data off Goa by The National Institute of Ocean Technology, Chennai, India have also been considered for this work for a one year period January 1 to December 31, 2000.

The data consist of 2517 data points recorded in different times of a day per a year. For our analysis, since we have large data set, we assume that this data is the real population and can be modeled by using Rayleigh distribution. The distribution of significant wave height data off Goa has mean 1.5958, standard deviation is 0.9769 and the maximum likelihood estimation for the parameter of Rayleigh distribution is 1.3229. So, the following two-sided hypothesis can be constructed:

$$H_0 : \sigma = 1.3229$$

$$H_1 : \sigma \neq 1.3229$$

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1602_19731982

Then, 50 observations were randomly selected to get a random sample for the calculations. The sample arises with mean 1.4126 and standard deviation 0.8129. The maximum likelihood estimation of the Rayleigh parameter for the sample data was calculated as $\hat{\sigma}=1.1496$. LRT and IT test statistics are found as 3.6021 and 0.14048, respectively. Hence, the null hypothesis that claims the data comes from the Rayleigh distribution with parameter $\sigma=1.3229$ accepted.

Our empirical results in Example 2 shows that test accepts the null hypothesis, as it is expected. However, for the same experiment run 100000 times the IT provides higher power than the LRT, as demonstrated by the simulation results in *Figure 3* (Sezer and Asma, 2010).

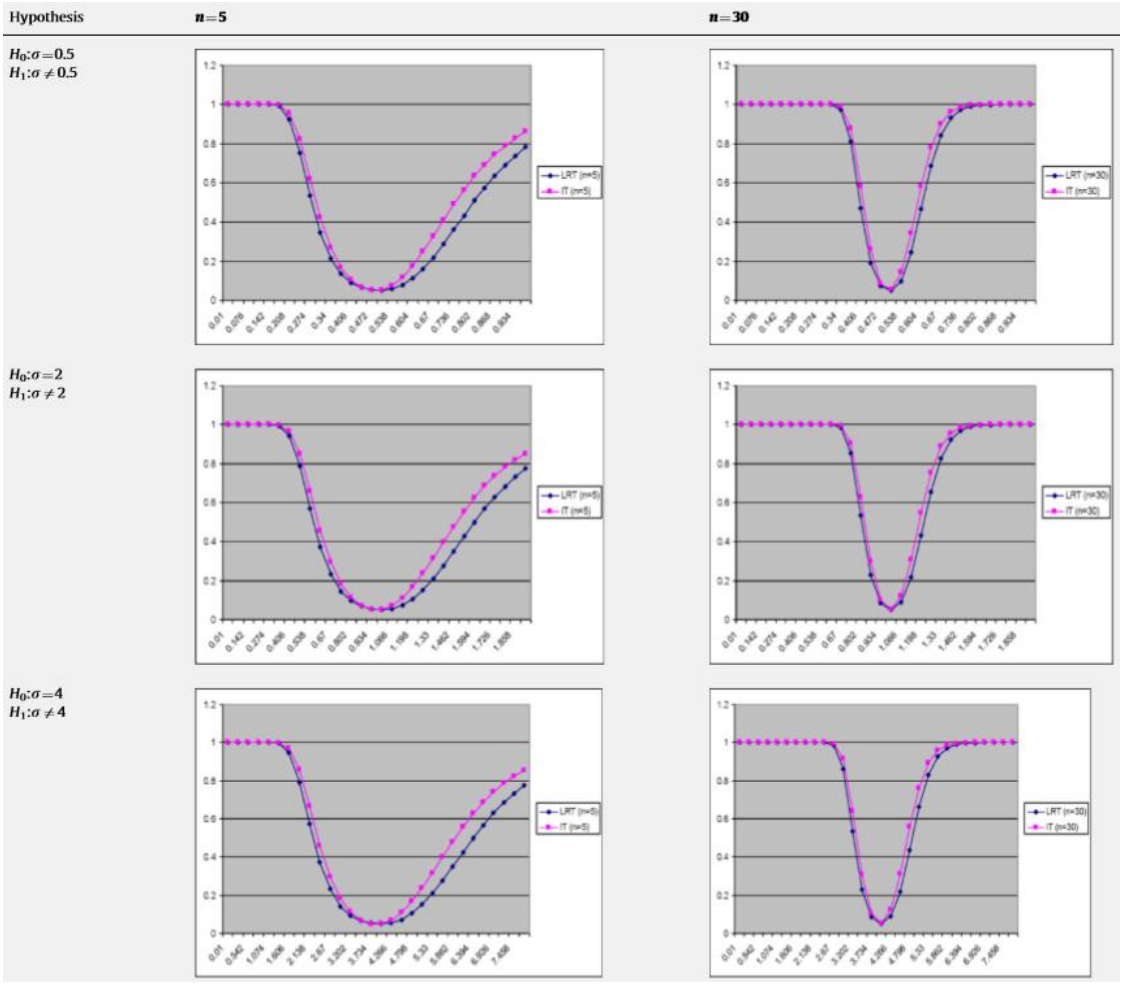


Figure 3. *Power functions for* $H_0:\sigma=\sigma_0;\ H_1:\sigma\neq\sigma_0$

Discussion

In this study, we focused on finding appropriate sample size for testing the variance of the normal distribution and Rayleigh distribution. The power curves are obtained both for the IT and the LRT by the Monte Carlo simulations. When we graphed the power versus sample size with different effect sizes, it can be seen that both IT and LRT curves are monotonically increasing function of the sample size. Based on the power curves, an algorithm is stated to specify an interval for the possible values of the effect

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1602_19731982

size. In order to determine sample size with the LRT and the IT, we should have the following steps:

1. Specify the null hypothesis and the alternative hypothesis.
2. Specify the significance level α.
3. Obtain the power curves for given sample size and significance level.
4. Determine the range of the appropriate effect size based upon the power curves.
5. Specify the target value of the power of the test.
6. Finally, obtain the sample size for the fixed power value.

Conclusion

As conclusion, IT gives more powerful results than LRT with the same sample size. Simulation results indicate that IT outperforms over the LRT in regarding to get smaller sample to attain the same level of the power. Wave height data is used to introduce show the inference procedure works both for the likelihood test and information-based test. In agreement with the theoretical results of Janssen (2014) it is found that wave height obeys the Rayleigh distribution.

Clearly, Rayleigh distribution can be used efficiently to calculate the associated probabilities in marine climatology. We suggest to use IT test to determine appropriate sample size at the beginning of the study. Our simulation studies will be expanded with different skewed distributions belong to other exponential family members such as Log-Normal and three-parameter Weibull distributions.

REFERENCES

[1] Adcock, C. J. (1997): Sample size determination: a review. – The Statistician 46: 261-283.

[2] Burnham, K. P., Anderson, D. R. (2002): Fitting the negative binomial distribution to biological data. – Biometrics 9: 176-200.

[3] Dziak, J. J., Lanza, S. T., Tan, X. (2014): Effect size, statistical power, and sample size requirements for the bootstrap likelihood ratio test in latent class analysis. – Structural Equation Modeling: A Multidisciplinary Journal 21(4): 534-552.

[4] Esteban, M. D., Castellanos, M. E., Morales, D., Vajda, I. (2001): Monte Carlo comparison of four normality tests using different entropy estimates. – Communication in Statistics-Simulation and Computation 30: 761-785.

[5] Gupta, A. K., Harrar, S. W., Pardo, L. (2007): On testing homogeneity of variances for nonnormal models using entropy. – Statistical Methodology and Application 16: 245-261.

[6] Janssen, P. A. E. M. (2014): On a random time series analysis valid for arbitrary spectral shape. – Journal of Fluid Mechanics 759: 236-256.

[7] Kass, R. E., Caffo, B. S., Davidian, M., Meng, X. L., Yu, B., Reid, N. (2016): Ten simple rules for effective statistical practice. – PLOS Computational Biology 12: p.e. 1004961.

[8] Kramer, M., Font, E. (2017): Reducing sample size in experiments with animals: historical controls and related strategies. – Biological Reviews 92(1): 431-445.

[9] Lehmann, E. L. (2006): On Likelihood Ratio Tests. – In: Rojo, J., Pérez-Abreu, V. (eds.) The Second Erich L. Lehmann Symposium Optimality. IMS Lecture Notes-Monograph Series. Vol. 49, pp. 1-8. IMS, Beachwood, OH.

[10] Lenth, R. V. (2001): Some practical guidelines for effective sample size determination. – The American Statistician 55: 187-193.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1602_19731982

[11] Low-Decarie, E., Chivers, C., Granados, M. (2014): Rising complexity and falling explanatory power in ecology. – Frontiers in Ecology and the Environment 12: 412-418.
[12] Lynch, T. B. (2017): Optimal sample size and plot size or point sampling factor based on cost-plus-loss using the Fairfield Smith relationship for plot size. – Forestry 90(5): 697-709.
[13] Martins, P. D., Poleto, C. (2018): Entropy for determination of suspended sediment concentration: parameter related to granulometry. – Journal of Environmental Engineering 144(3): 04017111-04017111-7.
[14] Men'endez, J. A., Rueda, C., Salvador, B. (1992): Dominance of likelihood ratio tests under cone constraints. – American Statistician 20: 2087-2099.
[15] Perlman, M. D., Wu, L. (1999): The Emperor's new tests (with discussion). – Statistical Science 14: 355-381.
[16] Richards, S. A. (2005): Testing ecological theory using the information-theoretic approach: examples and cautionary results. – Ecology 86: 2805-2814.
[17] Roach, T. N. F., Nulton, J., Sibani, P., Rohwer, F., Salamon, P. (2017): Entropy in the tangled nature model of evolution. – Entropy 19(5): 192.
[18] Schepsmeier, U. (2015): Efficient information based goodness-of-fit tests for vine copula models with fixed margins: A comprehensive review. – Journal of Multivariate Analysis 138: 34-52.
[19] Self, S., Mauritsen, R. (1988): Power/sample size calculations for generalized linear models. – Biometrics 44: 79-86.
[20] Self, S., Mauritsen, R., Ohara, J. (1992): Power calculations for likelihood ratio tests in generalized linear models. – Biometrics 48: 31-39.
[21] Sezer, A., Asma, S. (2010): Statistical power of an information-based test and its application to wave height data. – Computers and Geosciences 36: 1316-1324.
[22] Shieh, G. (2000): On power and sample size calculations for likelihood ratio tests in generalized linear models. – Biometrics 56: 1192-1196.
[23] Vranken, I., Baudry, J., Aubinet, M., Visser, M., Bogaert, J. (2015): A review on the use of entropy in landscape ecology: heterogeneity, unpredictability, scale dependence and their links with thermodynamics. – Landscape Ecology 30(1): 51-65.
[24] Wood, S. N., Goude, Y., Shaw, S. (2015): Generalized additive models for large data sets. – Journal of the Royal Statistical Society: Series C (Applied Statistics) 64: 139-155.
[25] Zuur, A. F., Ieno, E. N. (2016): A protocol for conducting and presenting results of regression type analyses. – Methods in Ecology and Evolution 7: 636-645.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1602_19731982