The Decision Case on Gaussian Binary Data

Budi Pratikno, Parwati, M.S.R
Department of Mathematics, Faculty of Mathematics and Natural Sciences,
Universitas Jenderal Soedirman, Indonesia
bpratikto@gmail.com

Abdul Talib Bon
Department of Production and Operations,
University Tun Hussein Onn Malaysia, Malaysia
talibon@gmail.com

Abstract

The research studied decision theory in case of binary data. The Gaussian assumption is then used. Due to this assumption, we then used the maximum likelihood to decide the decisions. We also compute the probability of the decision using the Neyman-Pearson method. A simulation is given to get the eligible result, and they showed the close result.

Keywords:
Decision, Gaussian binary data, and maximum likelihood and Neyman-Person methods

1. Introduction

Inference population can be drawn from sample. There are several statistical techniques to investigate the eligible conclusion, namely power and size techniques, decision theory, forecasting, quality control, statistical modeling and soon. In term of the power and size, many authors studied about improving inference population, such as Pratikno (2012), Khan and Pratikno (2013), Khan (2003), Khan and Saleh (1995), Khan and Hoque (2003), Saleh (2006), Yunus (2010), and Yunus and Khan (2007). Moreover, in term of statistical analysis of the decision theory, a lot people used it in getting the right choice (best conclusion of the population) of the decision, for example in the binary decision. Following Melsa and Cohn (1978), binary decision has only two choices, namely right (success) or wrong (fail). Here, the power on the Binomial distribution is suitable to analysis success and fail event, but in term of the decision theory, the maximum likelihood (ML) and Neyman-Pearson (NP) are more suitable and common.

In the context of the decision theory, some authors have studied about ML and NP, such as Yan and Blum (2001) and Trees (2001). Moreover, Yan and Blum (2001) already used Neyman-Perason method for detecting optimization of signal in binary censor, then Trees (2001) already studied many applications of the NP in medical cases of the heart attack. Melsa and Cohn (1978) then declare that binary single observed follow *Z* distribution (*Gaussian* or normal distribution, *Z*) with different mean and variance in signal-1 and signal-2. Furthermore, we note that, in the binary decisions consist of event m_1 correspond to d_1 (decision 1) and m_2 correspond to d_2 (decision 2).

Like the Binomial distribution, the decision theory also give the two choice of the decisions, accept (success) or reject (fail), but it does not forecast the choice. In term of the forecasting, Montgomery (2008) and Chase and Jacobs (2014) discussed how to predict the estimation of data in the future without choices, right or wrong. Note that that the power of the test, forecasting methods and decision theory are often used to make decision (and estimation or conclusion of the population) in or out of control with fixed (2σ or 95%) confidence interval (CI).

In this paper, Section 1 presented the introduction. The method of the decision theory is given in Section 2. A simulation data is obtained in Section 3, and Section 4 described the conclusion of the research.

2. Research Methodology

Step 1: Binomial Case

We studied about the binary data and Binomial case.

Step 2: Central Limit Theorem
We refer to the central limit theorem (CLT) that for large *n* the Binomial tend to be Gaussian (normal) distribution.

Step 3: Several Method to Draw Inference Population
We studied several method to draw inference population in case binary data, namely (1) the use of the power in discrete distribution (Binomial), and (2) the maximum likelihood (ML) and Neyman-Pearson (NP) methods in case binary data.

Step 4: Inference conclusion
We simulate the choice of the decision in getting the right inference conclusion of the population using assumption of the equation of the Gaussian Distribution.

3. Discussion Results
3.1 The Decision Methods
To illustrate how to use the power in the inference population, we present the power of the Binomial distribution. Let, X_i Bernoulli case with parameter θ, $n=12$, and $Y=\sum_{j=1}^{n=12} X_j$. The Y will then follow Binomial with $n=12$ and $p=\theta$, $Y:\ Bin(n,\theta)$. To test $H_0:\theta=0.7$ versus $H_1:\theta\neq 0.7$ on rejection area $\{(x_1,...,x_{12}):\mathrm{Y}\leq 5\}$, we then get the formula and graph of the power (see Figure 1), respectively,

$$\pi(\theta)=P(\text{Reject } H_0 \,|\, \text{under } H_1)=\sum_{y=0}^{5}\binom{12}{y}\theta^y(1-\theta)^{12-y}=(1-\theta)^7\left(1+7\theta+28\theta^2+84\theta^3+210\theta^4+462\theta^5\right)$$

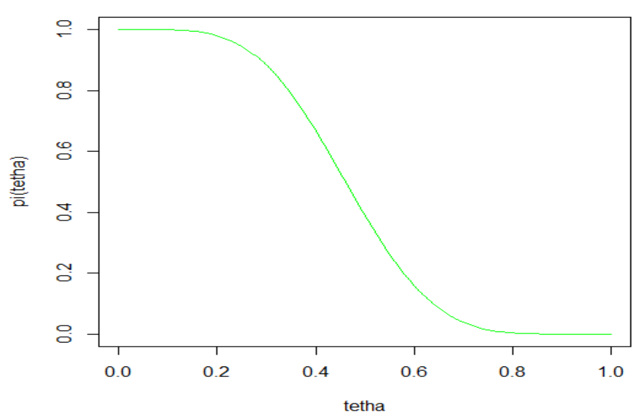

Figure 1. The Power of the Binomial Distribution

From Figure 1., we must choose the maximum power to get the right (eligible) conclusion. Here, we see that maximum power is occurred when θ close to zero ($0\leq\theta\leq 0.25$). Note that for large *n* the Binomial tend to be Gaussian (normal) distribution. So, it can be used to detect the right choice when the power is maximum.

Furthermore, in the context of the decision theory, we use probability to decide the right or wrong choice (test). Following, Bain and Engelhardt (1992), the maximum likelihood estimation (MLE) is defined as maximize join probability density function (pdf) random variables $X_1,...,X_n$,

$$f(x_1,...,x_n;\hat{\theta})=\max_{\theta\in\Omega} f(x_1,...,x_n;\theta) \tag{1}$$

where $\theta \in \Omega$ is parameter and parameter space. Using the equation (1), we use m_1 and m_2 (as binary decision) that follow Z, then it is written as $p(z \mid m_1)$ and $p(z \mid m_2)$, with the criterion of the decision is given as

$$d(z) = \begin{cases} d_1 & \text{jika } p(z \mid m_1) > p(z \mid m_2) \\ \\ d_2 & \text{jika } p(z \mid m_2) > p(z \mid m_1), \end{cases} \quad (2)$$

where $f(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2}$ with mean (μ) and variance (σ^2) .

In other method, following Melsa and Cohn (1978), the Neyman-Pearson (NP) methods has several choices (four decision) on binary case, namely (1) d_1, when m_1 is right, $P(d_1 \mid m_1)$, (2) d_1, when m_2 is right, $P(d_1 \mid m_2)$, (3) d_2, when m_1 is right $P(d_2 \mid m_1)$, and (4) d_2, when m_2 right, $P(d_2 \mid m_2)$. Note that $P(d_j \mid m_i) = \int_{Z_j} p(z \mid m_i) dz,\ i = 1,2 \text{ and } j = 1,2.$ Moreover, Trees [5] then presented that NP method is a simple method to handle decision choice using conditional probability as follow*: (1)* d_1 or d_2, we must then consider the $P(d_1 \mid m_1) + P(d_2 \mid m_1) = 1$ and $P(d_1 \mid m_1) + P(d_2 \mid m_1) = 1$. Here, we note that $P(d_2 \mid m_1)$ is said *level of significance* and $P(d_2 \mid m_2)$ is defined as *power of the test*.

3.2 A Simulation Study

Here, Parwati (*unpublished*) showed a simulation study on the equation $3z^2 + 2z - 6.54 = 0.$ of the Z distribution. We then get the solutions of the equation are $z_1 = 1.2$ and $z_2 = -1.8$. Therefore, the decision area of the d_1 and d_2 are given as $Z_1 = \{z \mid -1.8 < z < 1.2\}$ and $Z_2 = \{z \mid z < -1.8 \text{ atau } z > 1.2\}$. It mean that both d_i *i =1, 2* follow standard Gaussian. Furthermore, using the equation (2) and NP method, we consider *likelihood ratio*, $\Lambda(z) = \frac{p(z \mid m_2)}{p(z \mid m_1)} = \frac{1}{2} e^{\frac{3z^2+2z-1}{8}}$. Due to this, we then get $\frac{1}{2} e^{\frac{3z^2+2z-1}{8}} < \lambda$ to choose d_1 , and $\frac{1}{2} e^{\frac{3z^2+2z-1}{8}} > \lambda$ to choose d_2. Refer to Melsa and Cohn (1978) that $\mathrm{P}(d_2 \mid m_1) = \alpha$ with $\alpha = 0.025$ (95% for two-sides), we then used the α $(\alpha = 0.025)$ to get the decision area, that are $Z_1 = \{z : -1.3 < z < 0.68\}$ and $Z_2 = \{z : z < -1.3 \text{ atau } z > 0.68\}$. From the both Z_1 and Z_2 areas of the both models, we have the lower bound are -1,8 and -1.3, and the upper bound are 1.2 and 0.68. They are so close and not far away.

4 Conclusion

The research studied inference population using power and decision theory in Gaussian case binary data. The maximum power, the maximum likelihood and Neyman-Pearson methods are used as the methods. The result showed that the ML and NP give the close (similar) result of the decision range areas.

References

Bain, L. J. and Engelhardt, M. (1992). Introduction to Probability and Mathematical Statistics, *2nd Edition*. USA: Duxbury Thomson Learning

Chase, R and Jacobs, R. (2014). Operation and Supply Chain Management. *Global Case Edition*, New York: Mc Graw Hill.

Khan, S. and Pratikno, B. Testing Base Load with Non-Sample Prior Information on Process Load. *Statistical Papers*, 54 (3), 605-617, (2013).

Khan, S. Estimation of the Parameters of two Parallel Regression Lines Under Uncertain Prior Information. *Biometrical Journal, **44**, 73-90*, (2003).

Khan, *S.* and Saleh, A. K. Md. E. Preliminary test estimators of the mean based on p-samples from multivariate Student-t populations. *Bulletin of the International Statistical Institute. 50th Session of ISI, Beijing, 599-600*, (1995).

Khan, S. and Hoque, Z. Preliminary test estimators for the multivariate normal mean based on the modified W, LR and LM tests. *Journal of Statistical Research, Vol 37, 43-55, (*2003).

Melsa, J. L. and Cohn, D. L. 1978. Decision and Estimation Theory. Tokyo: McGraw-Hill Kogakusha.

Montgomery, D. (2008). Forecasting and Time Series Analysis. New Jersey: Jhon Wiley and Sons.Inc.

Pratikno, B.. Test of Hypothesis for Linear Models with Non-Sample Prior Information. Unpublished PhD Thesis, University of Southern Queensland, Australia, (2012).

Saleh, A. K. Md. E. Theory of preliminary test and Stein-type estimation with applications. *John Wiley and Sons, Inc., New Jersey,* (2006).

Trees, H. L. V.(2001). Detection, Estimation, and Modulation*,* Part 1. Canada: A Wiley-Interscience.

Yan, Q and Blum, R. S. Distributed Signal Detection Under the Neyman-Pearson Criterion*. Journal of IEEE Transactions on Information Theory* **47**(4) (2001).

Yunus, R. M. Increasing power of M-test through pre-testing. *Unpublished PhD Thesis, University of Southern Queensland, Australia*, (2010).

Yunus ,R. M. and Khan, S. Test for intercept after pre-testing on slope a robust method. *In: 9th Islamic Countries Conference on Statistical Sciences (ICCS-IX): Statistics in the Contemporary World - Theories, Methods and Applications,* (2007).

Acknowledgement
I thankfully to the colleague for providing me the ideas and data.

Biographies

Budi Pratikno, currently works as a lecturer in Department of Mathematics, Universitas Jenderal Soedirman, author of several international journals related to testing intercepts with non-sample prior information (NSPI), which are widely published in international journals such as Statistical Papers, Far East Journal of Mathematics Science, Springer, JSTA, ISSOS, IJET, IJAST, IOP series and others. In the study of NSPI as the main research the author, we try to improve population inference using non-sample. The author completed his undergraduate education at the Mathematics Department (UGM), then continued his Masters in Statistical Science at La Trobe University, Melbourne, Australia, and his S3 in Statistical Science (Ph.D) at the University of Southern Quensland (USQ), Toowoomba, Brisbane, Australia.

Parwati, M.S.R is a student in Department of Mathematics, Universitas Jenderal Soedirman.

Abdul Talib Bon is a professor of Production and Operations Management in the Faculty of Technology Management and Business at the Universiti Tun Hussein Onn Malaysia since 1999. He has a PhD in Computer Science, which he obtained from the Universite de La Rochelle, France in the year 2008. His doctoral thesis was on topic Process Quality Improvement on Beltline Moulding Manufacturing. He studied Business Administration in the Universiti Kebangsaan Malaysia for which he was awarded the MBA in the year 1998. He's bachelor degree and diploma in Mechanical Engineering which his obtained from the Universiti Teknologi Malaysia. He received his postgraduate certificate in Mechatronics and Robotics from Carlisle, United Kingdom in 1997. He had published more 150 International Proceedings and International Journals and 8 books. He is a member of MSORSM, IIF, IEOM, IIE, INFORMS, TAM and MIM.