JOURNAL OF STATISTICAL METHODS AND DATA SCIENCE

Journal of Statistical Methods and Data Science
Volume 01 Nomor 1 June 2023
https://jurnalmipa.unri.ac.id/jsmds

BAYESIAN STATISTICS FOR STUDY POPULATION STATISTICS AND DEMOGRAPHY

Achmad Isya Alfassa[1*]
[1] Universitas Islam Indragiri, Riau, 29212, Indonesia

***e-mail**: *achmadisyaalfassa@gmail.com*

Article Info:
Received: 08-06-2023
Accepted: 19-06-2023
Available Online: 30-06-2023

Keywords:
bayesian statistics
demographic
population
statistics
.

Abstract: Bayesian statistics is a method that belongs to the realm of statistical science which is based on the rules of the science of chance or probability. The Bayesian method is also used in carrying out projection analysis to see a picture of future conditions. This research was conducted to show the relationship between Bayesian Statistics and Demographic and Population Statistics Studies. The results of Bayesian Statistics can be used in the study of Population Statistics and Demography to carry out analysis with previous data and to find out and predict a picture of future conditions to determine the right policies, especially in analyzing population projections, population indicators and other demographics

1. INTRODUCTION

Bayesian is one of the statistical analysis approaches that is based on the rules of chance or probability to make direct judgments about parameters. The Bayesian technique is used in the terminology analysis of population statistics in the creation of scientific statistics. The Bayes method considers parameters to be variables that describe initial knowledge about parameters prior to making observations and are expressed in a distribution known as the prior distribution. In population statistics and demography, the parameters used during the analysis are active observations that begin with initial observations and continue with subsequent observations. Population statistics and demography is a field of study that combines two scientific branches, namely statistics and population science. Alfassa (2022) Population statistics is a science that studies two branches of science, namely statistics and population science. These two branches of science are integrated in population statistics to help solve population problems through statistical analysis and are used as recommendations in making regional population policy decisions.

In population science and demography, there are three basic factors that can be used to execute research studies: fertility, mortality, and migration. Population science has a strong attachment to statistics, which is validated by scientific combinations in conducting analyses and creating policies for the implementation of future development in support of Sustainable Development Goals (SDGs). The Bayesian technique in population statistics and demography is a method that is based on conditional probabilities. This method is also one of the simplest

for dealing with inconsistent or biased data, including both population data and population projection data.

Alfassa (2018) conducted research using statistics and mapping to determine clusters of population data with the same characteristics between countries, this research was conducted using self organizing maps and geospatial mapping. Nurmayanti (2021) conducted research on the application of naïve Bayes in classifying the poor in Lepak village, where testing results were obtained in predicting the results of the classification which indicated that the poor and not poor could be seen from the accuracy value, which was 96.63%, which means that they are included in the category good and shows that the class classification for the Lepak village community is the class with the poor. Alfassa (2020) conducted research using statistics to classify areas prone to forest and land fires using the self organizing maps method, this research was conducted to provide recommendations for preparing for the impact of forest and land fires. Andika (2022) explained the application of Bayes' theorem to an expert system for detecting sheep diseases, where the aim of the study was to create an expert system for diagnosing sheep diseases using Bayes' theorem. The results of Bayes' theorem show that the disease experienced in sheep is helminthiasis with a probability level of 60.71% and the results of the black box test show that the functionality of the system is running well. Imani (2023) conducted research using statistical science on social indicators and population indicators to see the characteristics of social and population conditions in the Province of East Nusa Tenggara, this study used two indicators to show comparisons and be able to explain the characteristics of social conditions and population in the Province of East Nusa Tenggara

In general, statistical scientific terminology is often associated with several sectors such as applied statistics, economic statistics, and actuarial statistics. It is very rare to find writings or references to populations statistics and demography, therefore this study aims to provide an understanding of Bayesian statistics both theoretically and in formulas to be linked to research and studies of population statistics and demography.

2. LITERATURE REVIEW

2.1. Bayesian Theorem

The bayes theorem is a simple mathematical technique or formula for calculating conditional probabilities. The Bayes theorem describes the probability of an event based on other relevant information. Essentially, it involves estimating the probability and then updating that estimate based on more information (Hartshorn, 2016). Furthermore, bayes theorem is understood subjectively or through a Bayesian approach to statistical epistemology and the theory of inductive logic. The subjective in issue is a rule based on probability theory and empirical models on conditional probabilities of proof. Both depend on Bayes' theorem, which simplifies the computation of conditional probabilities and explains subjective perspectives. The deep idea of Bayes' theorem is that a hypothesis may be determined by any individual using data that is known to be true and that it lies at the core of every subjective technique.

The bayes theorem in statistics is a theorem having two different meanings. In the interpretation of the Bayes method, the Bayes theorem is used to state how far the degree of subjective confidence is. This theorem, according to a frequentist interpretation, describes the inverse representation of the probabilities of two events. This theorem also provides the basis for Bayesian statistics and has applications in science, population and demography, economics, medicine, and health. It had previously been used to discuss a human population expansion and

the demographic change (Bongaarts, 2009), understanding future population trends (Vollset et al., 2020), or to simulate a path of stock prices (Das, 2016).

Bayesian inference refers to the use of Bayes' theorem to update beliefs. Take note of the following equation:

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)}, j = 1, 2, \dots, k. \quad (1)$$

Then, based on the total probabilities:

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(A_j)P(B|A_j)}{\sum_{j=1}^{k} P(A_j)P(B|A_j)} \quad (2)$$

Equation (2) is also known as Bayes Theorem. There is an established theory in Bayesian statistics known as the "Monte Carlo Concept," which is a concept that creates sample observations from a specified distribution known as Monte Carlo generating. This technique is a generation technique for replicating difficult procedures and analyzing sample properties.

2.2. Bayesian Statistical Terminology

Bayesian statistics is a statistical theory popularized by Thomas Bayes, who described a special case of Bayes' theorem in his article in 1763. From the late 1700s until the early 1800s, Bayesian statistics were developed. Aside from Thoman Bayes, other figures such as Pierre-Simon Laplace expanded on Bayes' interpretation of probability. Pierre-Simon Laplace solved a variety of statistical problems using what is now known as the Bayesian technique. Other writers later developed many of Bayes' approaches, but the name was rarely used to denote such methods until the 1950s.

For both philosophical and practical reasons, the Bayesian technique fell out of favor with many statisticians throughout the twentieth century. Most of the approaches extensively applied this century are based on frequency interpretation because the Bayesian method requires a large number of computations to complete. The Bayesian technique is increasingly being used in statistics in the twenty-first century. However, as technology advances and becomes more speedy and complex, some academics incorporate the Bayes technique into computer algorithms, giving rise to new algorithms such as the Monte Carlo Markov chain (Temp, 2016).

Bayesian statistics are based on probabilities, which express an event's degree of confidence. The level of confidence might be based on prior knowledge of the event, such as previous experiment findings, or on personal ideas about the event.

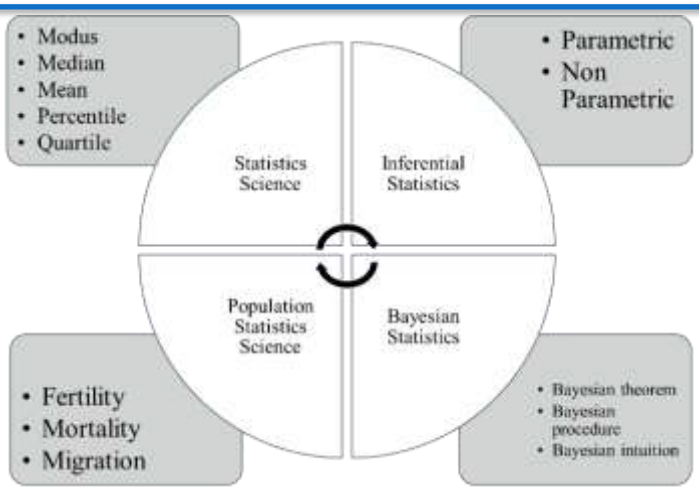

Figure 1. Bayesian Statistical Terminology

3. RESEARCH METHODOLOGY

3.1. Data and Sources

Data is information and real facts; it is often referred to as meaningful numbers. The data in this study is secondary, derived from earlier research material. The data in this study are documents that compare and collect past studies as a reference in the researcher's analysis.

3.2. Method

The study used a descriptive narrative analysis method to explain Bayesian statistics in the study of population statistics and demography. The descriptive narrative analysis method is applied to provide comprehensive, structured, and systematic explanations and can provide insight for readers for future research needs.

3.3. Data Analysis

Figure 2. Step of Analysis

4. RESULTS AND DISCUSSION

4.1 Two Independent Events

Suppose from an experiment that event A has no effect on the value of event B, with P(A) > 0

$$P(B|A) = P(B)$$

If events A and B are mutually independent during the case's learning. As a result, the multiplication rule is as follows:

$$P(A \cap B) = P(A)P(B|A) = P(A)P(B) \quad (3)$$

This will be different if the opposite situation occurs; if P (B) > 0, the rule will be reversed.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A) \quad (4)$$

So, as stated in formula (4), P(A) > 0 and P(B) > 0, implying that the rule P(A ∩ B) = P(A)P(B). will apply. With the rule P(A ∩ B) = P(A)P(B), event A and event B are mutually exclusive. Ronald E. Walpole (2007) "The probability of an event B occurring when it is known that some event A has occurred is called a conditional probability and is denoted by P(P|A). The symbol P(P| A) is usually read "the probability that B occurs given that A occurs" or simply "the probability of B, given A."

4.2 Monte Carlo Concept

The concept of Monte Carlo, also known as Monte Carlo Simulation, is a statistical technique used to anticipate the possible result of an uncertain event. The Monte Carlo idea is frequently used in computer programs to examine past data and predict possible future events based on a choice action. Furthermore, Monte Carlo Simulation is a probabilistic model that can include an element of uncertainty or randomness in producing predictions.

The Monte Carlo Simulation technique has been divided into five simple paths:

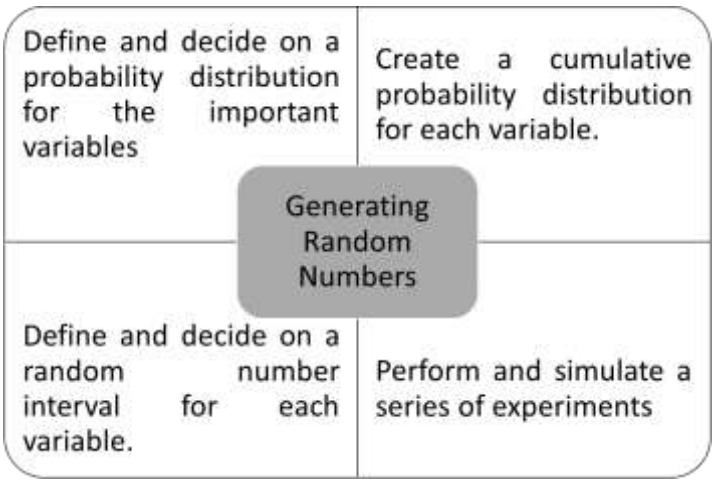

Figure 3. Five paths of Monte Carlo

Monte Carlo Simulation is also used to produce sample observations from a given distribution, which is referred to as Monte Carlo generation. This simulation approach is used to duplicate difficult procedures and investigate sample variables. Assume X is a random variable with a value of 1, and if the value of a dice that emerges is a value of 1 or 2, otherwise X will have a value of 0. X can thus be expressed as follows:

$$X = \begin{cases} 1, if\ 0 < U\ \leq \frac{1}{3} \\ 0,\ \ if\ \frac{1}{3} < U \leq 1 \end{cases} \tag{5}$$

4.3. Bayesian Procedure

As a statistical technique, the Bayesian procedure is predicated on probabilistic assumptions and is sensitive to assumption incidents. Bayesian techniques are typically based on the same assumptions as conventional procedures, plus extra assumptions about the distribution of any parameters (such as μ dan σ^2 di in the t-test) that might be present in the statistical model. This last assumption is commonly referred to as the distribution of the former parameters. Bayesian techniques are frequently utilized with the gamma function characteristic.

$$\begin{gathered} \Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx\ if\ z\ \in \mathbb{C}, Re(z) \geq 0 \\ \Gamma(n) = (n-1)!\ if\ n\ \in\ \mathbb{Z}, n\ \geq 1 \\ \int_0^\infty x^n e^{-ax}dx = \frac{\Gamma(n+1)}{a^{n+1}}\ if\ n > -1, a > 0. \end{gathered} \tag{6}$$

4.4. The Relationship between Bayesian Statistics and Population and Demographic Statistics

Bayesian statistics has several procedures to be used in the analysis of population and demographic statistical studies starting from the scientific basis that can support population and demographic research. One of the methods in Bayesian statistics, namely naïve Bayes, can be used as an analytical method to predict population numbers and project both fertility, mortality and migration indicators.

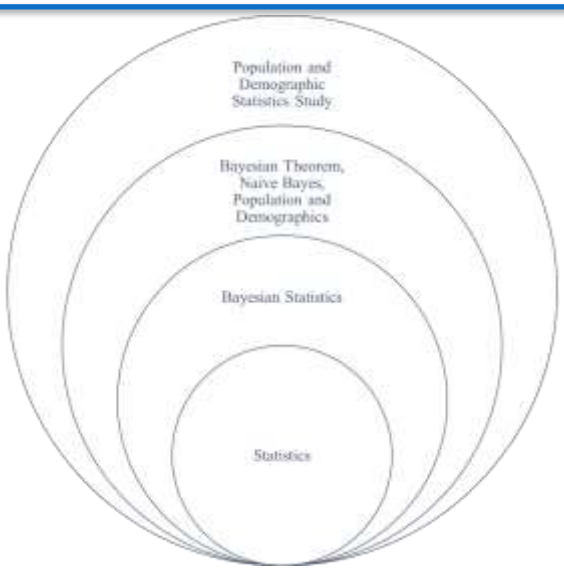

Figure 4. Relations Statistics, Bayesian Statistics, and Population Statistics and Demographics

In conducting research on population statistics and demography, several main indicators are used, such as indicators of fertility, mortality, and migration, as well as several supporting indicators, such as total population, total mobility, number of births, maternal mortality, infant mortality, and other indicators. as in the models below:

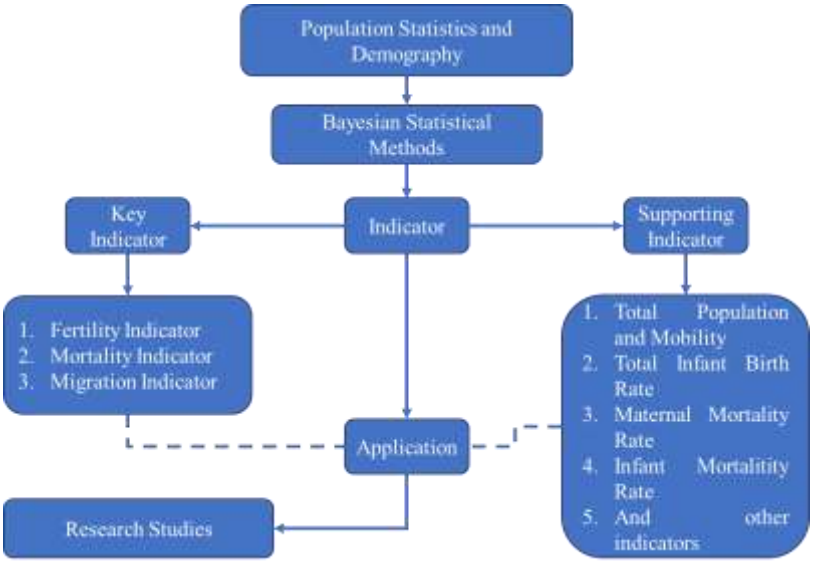

Figure 5. Relationship Model of Bayesian Statistics and Population Statistics and Demographics

5. CONCLUSION

Bayesian Statistics can be used in the study of Population Statistics and Demography to carry out analysis with previous data and to find out and predict a picture of future conditions to determine the right policies, this is due to the relationship between the two theories in scientific terminology and their application to society. Monte Carlo Simulation is a method that can be used in statistical analysis of population and demography, especially in analyzing

population projections and other population and demographic indicator projections. In the study of population statistics and demography using the Bayesian statistics approach, there are two indicators that can support the application of the study, namely the main indicator which consists of indicators of fertility, mortality and migration. Also several supporting indicators which consist of total population, total mobility, total birth rate, maternal mortality rate, infant mortality rate, and other indicators that can support scientific studies of population statistics and demography.

REFERENCES

Alfassa, A. I. (2018). Aplikasi Self Organizing Maps dan Webgis dengan menggunakan R dan QGIS untuk Analisis Kependudukan 100 Negara di Dunia. Skripsi. Statistika, Universitas Islam Indonesia.

Alfassa, A. I., & Kesumawati, A. (2020). Segmentation of Karhutla Hotspot Point of Indragiri Hilir Regency 2015 and 2016 using Self Organizing Maps (Soms). In *Proceedings Ofthe International Conference on Mathematics and Islam (ICMIs 2018). UIN Mataram Indonesia and ADMAPETA (Asosiasi dosen matematika dan pendidikan/Tadris Matematika), Mataram, Indonesia* (pp. 336-341).

Alfassa, A. I. (2022). Statistika Kependudukan Untuk Rencana Kebijakan Kependudukan Daerah. *DEMOS: Journal of Demography, Ethnography and Social Transformation*, *2*(2), 76-85.

Andika, I., Maharani, D., & Mardalius, M. (2022). Penerapan Teorema Bayes pada Sistem Pakar Pendeteksi Penyakit Domba. *Edumatic: Jurnal Pendidikan Informatika*, *6*(2), 252-259.

Bongaarts, J. (2009). Human population growth and the demographic transition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1532), 2985–2990. https://doi.org/10.1098/rstb.2009.0137

Das, S. R. (2016). *Data Science: Theories, Models, Algorithms, and Analytics*. Apache License. http://algo.scu.edu/∼sanjivdas/

Hartshorn, S. (2016). *Bayes Theorem Examples: A Visual Guide For Beginners*. On Kindle Scribe. http://pzs.dstu.dp.ua/DataMining/bayes/bibl/bayes_theorem_examples.pdf

Imani, N., Alfassa, A. I., & Yolanda, A. M. (2023). Analisis Cluster Terhadap Indikator Data Sosial Di Provinsi Nusa Tenggara Timur Menggunakan Metode Self Organizing Map (SOM). *Jurnal Gaussian*, *11*(3), 458-467.

Nurmayanti, W. P., Saky, D. A. L., Malthuf, M., Gazali, M., & Hirzi, R. H. (2021). Penerapan Naive Bayes dalam Mengklasifikasikan Masyarakat Miskin di Desa Lepak. Geodika: Jurnal Kajian Ilmu dan Pendidikan Geografi, 5(1), 123-132.

Temp. (2016). Simulation for Data Science with R. *Packt Publishing*. https://www.oreilly.com/library/view/simulation-for-data/9781785881169/

Vollset, S. E., Goren, E., Yuan, C. W., Cao, J., Smith, A. E., Hsiao, T., Bisignano, C., Azhar, G. S., Castro, E., Chalek, J., Dolgert, A. J., Frank, T., Fukutaki, K., Hay, S. I., Lozano, R., Mokdad, A. H., Nandakumar, V., Pierce, M., Pletcher, M., … Murray, C. J. L. (2020). Fertility, mortality, migration, and population scenarios for 195 countries and territories from 2017 to 2100: a forecasting analysis for the Global Burden of Disease Study. *The Lancet*, *396*(10258), 1285–1306. https://doi.org/10.1016/S0140-6736(20)30677-2