

A New Definition of P-Value Involving the Maximum Likelihood Estimator

H. W. Corley

Center on Stochastic Modeling, Optimization, and Statistics (COSMOS), The University of Texas at Arlington, Arlington, USA
Email: corley@uta.edu

How to cite this paper: Corley, H.W. (2024) A New Definition of P-Value Involving the Maximum Likelihood Estimator. *Open Journal of Statistics*, 14, 546-552.
<https://doi.org/10.4236/ojs.2024.145024>

Received: September 16, 2024

Accepted: October 27, 2024

Published: October 30, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

We present here an alternative definition of the P-value for statistical hypothesis test of a real-valued parameter for a continuous random variable X . Our approach uses neither the notion of Type I error nor the assumption that null hypothesis is true. Instead, the new P-value involves the maximum likelihood estimator, which is usually available for a parameter such as the mean μ or standard deviation σ of a random variable X with a common distribution.

Keywords

Hypothesis Testing, P-Value, Evidence, Statistics

1. Introduction

A principal goal of statistics is to obtain evidence from data for comparing alternative decisions. For example, statistical evidence may allow one to decide that a population mean μ satisfies $\mu \leq \mu_0$ as opposed to $\mu > \mu_0$ for some specified μ_0 . Unfortunately, evidence is an ambiguous concept in statistics, though [1]-[7] among others have attempted to define it. Currently, the P-value [8] [9] is perhaps the most frequently applied measure of evidence used to reject or fail to reject a hypothesis. In Section 2, we review the standard P-value and the notion of a maximum likelihood estimator (MLE). We then propose the new MLE P-value involving the MLE of some parameter of interest. The MLE P-value is not related to significance levels and not defined under the assumption that the null hypothesis is true. Both the standard and new MLE P-values may be considered measures of evidence, but the latter is more intuitive. In Section 3, we present four examples and compare the two approaches.

2. The Standard and MLE P-Values

The notion of P-value is a fundamental tool in statistical inference and has been

widely used for reporting outcomes of hypothesis tests. Yet in practice, the P-value is often misinterpreted, misused, or miscommunicated. Moreover, it does not unequivocally reflect the available evidence for the null hypothesis since H_0 is assumed to hold. In this section, we propose the new MLE P-value that may give different values in some cases than do existing definitions. The MLE P-value provides a simple and intuitive interpretation of P-value. Our definition appears applicable to a wide range of hypothesis testing problems and yields an interpretation of P-value as both a cardinal and ordinal measure of the evidence. We restrict our development to standard one-sided hypothesis testing, but Example 3 illustrates that the approach here can also be used for two-sided hypothesis testing.

We first summarize the two standard ways of defining P-value for the general hypothesis test $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \notin \Theta_0$ with a parameter space Θ_0 . Let $T(\mathbf{X})$ be a test statistic for a random sample $\mathbf{X} = (X_1, \dots, X_n)$ from a random variable X with a single scalar parameter θ . Denote the observed data for \mathbf{X} as $\mathbf{x} = (x_1, \dots, x_n)$.

Definition 1 (Definitions of P-value from [10] and [11]). Under the assumption that $H_0 : \theta \in \Theta_0$ is true, the P-value associated with the observed data \mathbf{x} is defined as either

$$PV1(\mathbf{x} | \Theta_0) = \sup_{\theta \in \Theta_0} P\{T(\mathbf{X}) \geq T(\mathbf{x}) | \theta\} \quad (1)$$

or

$$PV2(\mathbf{x} | \Theta_0) = \inf \{\alpha : T(\mathbf{x}) \in R_\alpha\}, \quad (2)$$

where R_α is the rejection region for a level of significance α .

Equation (1) is usually interpreted as follows. Under the assumption that H_0 is true, P-value is the probability that $T(\mathbf{X})$ is as least as extreme as its observed value $T(\mathbf{x})$. This interpretation can lead to the common misunderstanding that this definition of P-value is the probability that H_0 is true. On the other hand, under the assumption that H_0 is true, Equation (2) is based on significance levels (Type I Error probabilities) and can lead to the misunderstanding that P-value is only a measure of Type I Error and not related to the likelihood that H_0 is true.

Both definitions lead to the question: how can the assumption that H_0 is true produce evidence that H_0 is true? The answer is that H_0 is rejected when there is a probability less than or equal to 0.05 (for example) that H_0 is true when it is assumed true. In other words, a contradiction involving probability is reached. For other issues with these definitions, see [10]-[12], for example. To address such issues, we utilize the well-known maximum likelihood estimator (MLE) [13] [14] and define a P-value utilizing the maximum likelihood estimator for the parameter under consideration.

Definition 2 (Likelihood Function and MLE). Let $\mathbf{x} = (x_1, \dots, x_n)$ be sample data from a random sample $\mathbf{X} = (X_1, \dots, X_n)$ from a random variable X with sample space S and real-valued parameter θ . For the joint pdf $f(\mathbf{x} | \theta)$ of the random sample \mathbf{X} . For any sample data \mathbf{x} , the likelihood function of θ is defined as

$$L(\theta | \mathbf{x}) = f(\mathbf{x} | \theta), \quad (3)$$

where $L(\theta | \mathbf{x})$ in (3) is a function of the variable θ for given data \mathbf{x} . The MLE $\hat{\theta}(\mathbf{x})$ maximizes $L(\theta | \mathbf{x})$ for the data sample \mathbf{x} ; i.e.,

$$\hat{\theta}(\mathbf{x}) = \arg \max_{\theta} L(\theta | \mathbf{x}) \text{ in terms of } \mathbf{x}.$$

Definition 3 (MLE P-value). Let $\mathbf{x} = (x_1, \dots, x_n)$ be observed random sample data for a random sample $\mathbf{X} = (X_1, \dots, X_n)$ from a continuous random variable X with a single real-valued parameter θ and pdf $f_X(x | \theta)$. Let $Y = \hat{\theta}(\mathbf{X})$ denote the MLE for θ with pdf $f_Y(y | \theta)$. For the hypothesis test $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \notin \Theta_0$, the MLE P-value (MPV) at the sample data \mathbf{x} for the null hypothesis $H_0 : \theta \in \Theta_0$ is defined as

$$\text{MPV}(\mathbf{x} | \Theta_0) = \int_{\Theta_0} f_Y(y | \theta = \hat{\theta}(\mathbf{x})) dy, \quad (4)$$

where the integration is over the scalar values $y \in \Theta_0$ of $Y = \hat{\theta}(\mathbf{X})$.

In the integration of (4), θ is estimated by the number MLE $\hat{\theta}(\mathbf{x})$ from the sample data \mathbf{x} . Then the pdf $f_Y(y | \theta = \hat{\theta}(\mathbf{x}))$ of the MLE $Y = \hat{\theta}(\mathbf{X})$ is integrated over the values θ in the null hypothesis $H_0 : \theta \in \Theta_0$. For example, in the one-tailed hypothesis test $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$, the set Θ_0 would be $(-\infty, \theta_0]$. In this case (4), $H_0 : \theta \in \Theta_0$ approximates the frequentist probability that $Y = \hat{\theta}(\mathbf{X}) \in \Theta_0$, and so $\text{MPV}(\mathbf{x} | \Theta_0)$ approximates the likelihood that is true. Any inaccuracies are due (i) to integrating over $Y = \hat{\theta}(\mathbf{X})$ since θ has no prior or posterior distribution and (ii) to setting $\theta = \hat{\theta}(\mathbf{x})$. Both result from using the MLE $\hat{\theta}(\mathbf{x})$ as a surrogate for the unknown parameter θ . Doing so, however, utilizes the fact that distributions of MLEs for a parameter θ are often known for a continuous random variable X with a common distribution such as the normal or exponential. In such cases, (4) may be analytically integrable. If not, a numerical integration could possibly be employed. However, only analytical integration is considered here.

3. Examples

Four examples are now presented to illustrate the MLE P-value approach. The first two involve the normal random variable, and the third involves an exponential random variable. The fourth demonstrates a possible limitation of the method.

Example 1. For a random sample of size n , consider the hypothesis test $H_0 : \mu \leq \mu_0$ vs $H_1 : \mu > \mu_0$ for the parameter μ of a normal random variable $X \sim N(\mu, \sigma^2)$. In this case, we rewrite $\text{MPV}(\mathbf{x} | \Theta_0)$ as $\text{MPV}(\mathbf{x} | \mu_0)$. We also show that (4) gives the standard P-value for $\theta = \mu$.

Result 1. Let X_1, \dots, X_n be a random sample from a random variable $X \sim N(\mu, \sigma^2)$ with unknown μ , and consider the one-sided hypothesis test $H_0 : \mu \leq \mu_0$ vs $H_1 : \mu > \mu_0$. Let \bar{x} be the sample mean for sample values x_1, \dots, x_n . When σ^2 is known,

$$\text{MPV}(\mathbf{x} | \mu_0) = \phi\left(\frac{\mu_0 - \bar{x}}{\sigma/\sqrt{n}}\right); \quad (5)$$

and when σ^2 is unknown,

$$\text{MPV}(\mathbf{x} | \mu_0) = F_{n-1}\left(\frac{\mu_0 - \bar{x}}{s/\sqrt{n}}\right), \quad (6)$$

where $\phi(z)$ is the cdf for the standard normal distribution and $F_{n-1}(t)$ is the cdf for the student t distribution with $n - 1$ degrees of freedom.

Proof. When σ^2 is known, to prove (5) we use the fact [9] that the MLE $\hat{\mu}(X)$ for μ is the sample mean $\bar{X} \sim N(\mu, \sigma^2/n)$. The integral of (4) therefore becomes

$$P[\bar{X} \leq \mu_0] = \phi\left(\frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right). \quad (7)$$

Substituting \bar{x} for μ in (7) gives (5). When σ^2 is unknown, Equation (6) follows from the definition of the student t distribution [9]. ■

Numerically, when $n = 9$, $\mu_0 = 12$, $\sigma = 2$, and $\bar{x} = 13$, the MLE P-value of (5) for $H_0 : \mu \leq \mu_0$ vs $H_1 : \mu > \mu_0$ is $\phi\left(\frac{12-13}{2/3}\right) \approx 0.067$ from the standard normal table [9]. If $n = 25$ instead, the MLE P-value of (5) is

$\phi\left(\frac{12-13}{2/5}\right) \approx 0.006$, and it is much less likely that $\mu \leq \mu_0$. It should be noted that the right sides of (5) and (6) are the standard P-values for the hypothesis test $H_0 : \mu \leq \mu_0$ vs $H_1 : \mu > \mu_0$ using the definition of (2). In particular, the usual test statistic for this case is $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ with a critical region of $z > z_\alpha$ [9],

where α is the level of significance, *i.e.*, the probability of committing a Type I error. The P-value is then the lowest level of α for which the observed z is significant. But this lowest level of α in standard hypothesis testing is simply

$$1 - \phi\left(\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right) = \phi\left(\frac{\mu_0 - \bar{x}}{\sigma/\sqrt{n}}\right), \text{ which is the right side of (5).}$$

Example 2. For a random sample of size n , consider the hypothesis test $H_0 : \sigma^2 \leq \sigma_0^2$ vs $H_1 : \sigma^2 > \sigma_0^2$ for the variance σ^2 of a normal random variable $X \sim N(\mu, \sigma^2)$. From [9], the MLE for the variance of a normal distribution is

$\hat{\sigma}^2(X) = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$, so $\frac{n\hat{\sigma}^2(X)}{\sigma^2} \sim \chi^2(n-1)$, from which we have the following.

Result 2. Let X_1, \dots, X_n be a random sample from a random variable $X \sim N(\mu, \sigma^2)$ with unknown σ^2 , and consider the one-sided hypothesis test $H_0 : \sigma^2 \leq \sigma_0^2$ vs $H_1 : \sigma^2 > \sigma_0^2$. Then the integral of (4) becomes

$$\text{NPV}(\mathbf{x} | \sigma_0^2) = F_{\chi^2(n-1)}\left(\frac{n\sigma_0^2}{\hat{\sigma}^2(x)}\right). \quad (8)$$

Proof. From [9], the MLE $\hat{\sigma}^2(X)$ for σ^2 of a normal distribution is

$$\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} = \frac{(n-1)S^2}{n} \quad \text{and so} \quad \frac{n\hat{\sigma}^2(X)}{\sigma^2} \sim \chi^2(n-1). \quad \text{Hence}$$

$$P[\hat{\sigma}^2(X) \leq \sigma_0^2] = P\left[\frac{n\hat{\sigma}^2(X)}{\sigma^2} \leq \frac{n\sigma_0^2}{\sigma^2}\right] = F_{\chi^2(n-1)}\left(\frac{n\sigma_0^2}{\sigma^2}\right). \quad (9)$$

Substituting $\hat{\sigma}^2(x)$ for σ^2 in $F_{\chi^2(n-1)}\left(\frac{n\sigma_0^2}{\sigma^2}\right)$ in (9) yields (8). ■

Numerically, when $n = 10$, $\sigma_0^2 = 4$, and $s^2 = 9$, the MLE P-value of (8) for $H_0: \sigma^2 \leq \sigma_0^2$ vs $H_1: \sigma^2 > \sigma_0^2$ becomes $F_{\chi^2(9)}(4.94) \approx 0.16$.

Example 3. In this example, we illustrate that the MLE approach is applicable in two-sided hypothesis testing. For a random sample of size n , consider the hypothesis test $H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$ for the parameter μ of

$X \sim N(\mu, \sigma^2)$ with σ known. To be more realistic, we modify this test to

$H_0: \mu \in [\mu_0 - \delta, \mu_0 + \delta]$ vs $H_1: \mu \notin [\mu_0 - \delta, \mu_0 + \delta]$, where δ is an acceptable tolerance level, i.e., an acceptable deviation from μ_0 . In this case, (4) becomes

$$\text{MPV}(x | \mu_0) = \phi\left(\frac{\mu_0 + \delta - \bar{x}}{\sigma/\sqrt{n}}\right) - \phi\left(\frac{\mu_0 - \delta - \bar{x}}{\sigma/\sqrt{n}}\right). \quad (10)$$

Numerically, when $n = 25$, $\mu_0 = 12$, $\sigma^2 = 16$, $\bar{x} = 12.3$, and $\delta = 0.5$, the MLE P-value of (10) becomes $\phi(0.25) - \phi(-1.00) \approx 0.4400$.

Example 4. Let the random variable X have an exponential distribution with pdf $f_X(x | \lambda) = \lambda e^{-\lambda x}$, $x > 0$, with parameter $\lambda > 0$. For a random sample of size n , consider the hypothesis test $H_0: \lambda \leq \lambda_0$ vs $H_1: \lambda > \lambda_0$. From [15], the MLE for λ is $Y = 1/\bar{X}$. It can be shown [16] that \bar{X} for the exponential random variable X follows the gamma distribution $\Gamma(n, n\lambda)$. Hence Y follows an inverse gamma distribution [16], so the right side of (4) becomes the regularized gamma function $Q\left(n, \frac{n}{\lambda_0 \bar{X}}\right)$, which is available in various software packages such as MATLAB.

Example 5. To illustrate the difficulty of finding the pdf $f_Y(y | \theta = \hat{\theta}(x))$ for MLE $Y = \hat{\theta}(X)$ in the integral of (4), consider the random variable X with pdf $f_X(x | \theta) = \theta x^{-(\theta+1)}$ for $x > 1$ and the parameter $\theta > 0$. Then it can be shown

[9] that $\hat{\theta}(x) = \frac{n}{\sum_{i=1}^n \ln(X_i)}$, whose pdf is difficult to determine.

4. Conclusions

The MLE P-value defined here gives a value that estimates the probability that a one-sided null hypothesis on a single parameter is true. Two-sided hypotheses can be similarly treated using tolerance levels as in Example 3. In the approach of this

paper, the MLE $Y = \hat{\theta}(X)$ may be thought of as a new type of test statistic that is integrated over Θ_0 , while its numerical value $\hat{\theta}(x)$ from the data x is used to estimate θ in this integration. Obtaining such an approximation of the probability that the null hypothesis is true is, in fact, the ultimate goal of hypothesis testing. Certainty is not possible. But given the MLE P-value, a decision maker would need to decide if this value is sufficiently large to accept the null hypothesis. Different decision makers might judge a given MLE P-value probability differently, but a metric for the decision has now been provided.

The principal limitation of the analytical approach here is that the pdf of the maximum likelihood estimator needs to be both known and reasonable to integrate. Future work should be directed at the numerical integration of these integrals. Tables could be developed for certain parameters of particular random variables X . Although MLEs have been used here as the surrogate for single parameters, other estimators could also be used. The advantage of MLEs is that they are well-studied and often immediately available. Finally, an approach similar to that of this paper could be applied to hypothesis testing for parameters of discrete random variables.

Acknowledgements

The author thanks his former student Maryham Moghimi for her discussion about statistics and the P-value.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Shafer, G. (1976) A Mathematical Theory of Evidence. Princeton University Press.
- [2] Evans, M. (2015) Measuring Statistical Evidence Using Relative Belief. Chapman and Hall/CRC. <https://doi.org/10.1201/b18587>
- [3] Hacking, I. and Romeijn, J. (2016) Logic of Statistical Inference. Cambridge University Press. <https://doi.org/10.1017/cbo9781316534960>
- [4] Dollinger, M.B., Kulinskaya, E. and Staudte, R.G. (1996) When Is a P-Value a Good Measure of Evidence? In: *Lecture Notes in Statistics*, Springer, 119-134. https://doi.org/10.1007/978-1-4612-2380-1_8
- [5] Hubbard, R. and Lindsay, R.M. (2008) Why P-Values Are Not a Useful Measure of Evidence in Statistical Significance Testing. *Theory & Psychology*, **18**, 69-88. <https://doi.org/10.1177/0959354307086923>
- [6] Singh Chawla, D. (2017) Big Names in Statistics Want to Shake up Much-Maligned P Value. *Nature*, **548**, 16-17. <https://doi.org/10.1038/nature.2017.22375>
- [7] Wang, B., Zhou, Z., Wang, H., Tu, X.M. and Feng, C. (2019) The P-Value and Model Specification in Statistics. *General Psychiatry*, **32**, e100081. <https://doi.org/10.1136/gpsych-2019-100081>
- [8] Chavalarias, D., Wallach, J.D., Li, A.H.T. and Ioannidis, J.P.A. (2016) Evolution of Reporting P-Values in the Biomedical Literature, 1990-2015. *Journal of the American*

- Medical Association*, **315**, 1141-1148. <https://doi.org/10.1001/jama.2016.1952>
- [9] Walpole, R.E., Myers, R.H., Myers, S.L. and Keying, Y. (2017) Probability and Statistics for Engineers and Scientists. 9th Edition, Pearson Education.
 - [10] Abell, M.L., Braselton, J.P. and Rafter, J.A. (1999) Statistics with Mathematica. Academic Press.
 - [11] Lehmann, E.L. and Romano, J.P. (2005) Testing Statistical Hypotheses. Springer.
 - [12] Goodman, S. (2008) A Dirty Dozen: Twelve P-Value Misconceptions. *Seminars in Hematology*, **45**, 135-140. <https://doi.org/10.1053/j.seminhematol.2008.04.003>
 - [13] Pawitan, Y. (2013) In All Likelihood. Statistical Modeling and Inference Using Likelihood. 1st Edition, The Clarendon Press.
 - [14] Casella, G. and Berger, R.L. (2002) Statistical Inference. 2nd Edition, Cengage Learning.
 - [15] Hines, W.W., Montgomery, D.C., Goldman, D.M. and Borror, C.M. (2003) Probability and Statistics in Engineering. 4th Edition, Wiley.
 - [16] <https://math.stackexchange.com/questions/155296/distribution-of-the-sample-mean-of-a-exponential>
 - [17] https://en.wikipedia.org/wiki/Inverse-gamma_distribution