International Journal of Rock Mechanics and Mining Sciences 106 (2018) 374–387

Contents lists available at ScienceDirect

International Journal of Rock Mechanics and Mining Sciences

journal homepage: www.elsevier.com/locate/ijrmms

Comprehensive statistical analysis of intact rock strength for reliability-based design

Nezam Bozorgzadeh[a,*], Michael D. Escobar[b], John P. Harrison[a]

[a] Department of Civil Engineering, University of Toronto, Toronto, Canada
[b] Division of Biostatistics, University of Toronto, Toronto, Canada

ARTICLE INFO

Keywords:
Parameter estimation
Uncertainty quantification
Bayesian regression analysis
Rock strength

ABSTRACT

Rock engineering design is currently evolving from the traditional use of safety factors and towards the more rational reliability-based design (RBD), as witnessed by the introduction of a number of geotechnical limit states design (LSD) standards worldwide. The probabilistic nature of RBD requires statistical characterization of design parameters, including the strength of intact rock. The triaxial compressive strength of intact rock is commonly characterized as a function of confining pressure, and values of the parameters that define this function are generally obtained by regression analysis of laboratory test data. However, the problem of fitting strength criteria to intact rock strength data has been historically tackled as a problem of obtaining best fit curves only, and has inadvertently omitted characterization of the variability inherent in the strength data. As a result, the uncertainty in parameter estimations resulting from this variability are often not quantified rigorously. Not only does this omission render such regression analyses incomplete, it also limits development of reliability-based design protocols for rock engineering which require statistical characterization of design parameters. To overcome both of these deficiencies, this paper presents frequentist (i.e. classical) and Bayesian regression models that rigorously incorporate variability and uncertainty associated with estimated Hoek-Brown strength parameters. In particular, it discusses the limitations of the frequentist model when dealing with limited data or a combination of tensile and compressive strength data. It discusses the potential of Bayesian data analysis techniques to overcome the issue of limited data in rock engineering design by using informative prior distributions. The paper also demonstrates the main issue in fitting strength criteria to tensile and compressive data simultaneously from a statistical perspective, and presents a Bayesian regression model that rigorously fits the strength criterion to a combination of tensile and compressive data. The paper concludes with suggested statistical approaches – Bayesian or frequentist – for conditions encountered in the analysis of intact rock strength data.

1. Introduction

Rock engineering design is currently evolving from the traditional use of safety factors and towards the more rational reliability-based design (RBD) (e.g.[1]), as witnessed by the introduction of geotechnical limit states design (LSD) standards such as EN-1997,[2] Canadian Highway Bridge Design Code (CHBDC),[3] and LRDF Bridge Design Specifications.[4] In the RBD context, the designer is required to first identify all of the limit states (i.e. modes of unsatisfactory/undesirable performance), and then demonstrate that the probability of exceeding each limit state is less than a predefined target value (e.g.[5–7]). To this end, statistical characterization of design parameters is an essential element of RBD. Fig. 1 illustrates a basic load-resistance structural reliability analysis. The limit state line separates the region of satisfactory from unsatisfactory performance, and the iso-density contours of the joint probability distribution of the load and resistance show the probabilistic nature of design, and emphasize the need for statistical characterization of design parameters.

In geotechnical (soil and rock) engineering, many parameters are defined as functions of some other parameter(s). As an example, triaxial compressive strength of intact rock is a fundamental design parameter in rock engineering which is commonly characterized as a function of confining pressure. It is often determined by testing small laboratory scale cylindrical specimens of rock[8] to which a strength criterion is fitted. If this fitting is to be statistically rigorous, then regression analysis is required.

The empirical Hoek-Brown (H-B) strength criterion for intact rock, which defines triaxial strength as a nonlinear function of confining

* Corresponding author.
E-mail address: Nezam.bozorgzadeh@mail.utoronto.ca (N. Bozorgzadeh).

https://doi.org/10.1016/j.ijrmms.2018.03.005
Received 12 May 2017; Received in revised form 13 March 2018; Accepted 14 March 2018
1365-1609/

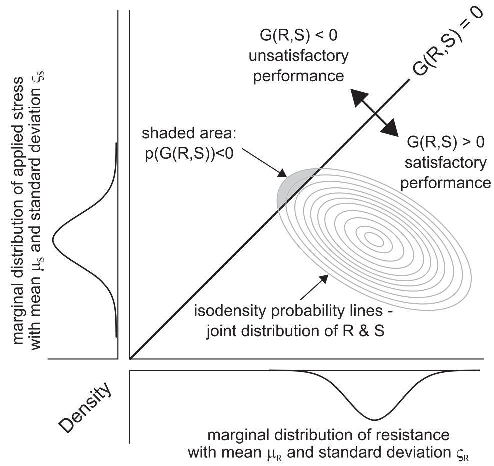

Fig. 1. Simple load-resistance structural reliability analysis.

pressure, has been widely used due to its relative simplicity, small number of parameters (only 2), and its link to the H-B criterion for the rock masses. Unfortunately, the problem of fitting this criterion to triaxial strength data has historically been tackled as a problem of obtaining a best fit (least squared errors) curve without considering the potential variability inherent in strength data. Such analyses result in deterministic parameter values and strength curves, none of which reflect the true (variable) nature of the strength data that they were estimated from. Also, and importantly, the results of such analyses do not provide a measure of variability (e.g. variance or standard deviation) as is required for application of RBD. More recently, and motivated by potential application of RBD, statistical methods have been proposed to quantify the uncertainties associated with estimated strength of intact rock and rock masses.[9] This approach applies expressions for confidence and prediction intervals obtained for simple linear regression to characterize the uncertainties associated with the nonlinear H-B criterion. Moreover, while it recognizes there to be a difference in the nature of errors when fitting the criterion to a combination of tensile and compressive strength, it does not explicitly reflect this difference in the formulation of dispersion parameters (variance or standard deviation) of these two data types. This will be discussed further in Section 4.

This paper presents frequentist and Bayesian statistical models for fitting the H-B criterion to intact rock strength data. Crucially, the models consider the actual variance structure of the data and the assumptions made in a regression analysis. To support introduction of these models, Section 2 reviews classical (frequentist) regression analysis and its application to the nonlinear H-B criterion. Section 3 introduces a Bayesian regression model as an alternative to the frequentist model, and highlights the benefits that the Bayesian framework brings to statistical characterization of rock strength data. Section 4 begins with a discussion of the difficulty of including tensile strength data in the classical regression model, and then goes on to expand the Bayesian model presented in Section 3 to robustly include tensile strength data in the regression analysis. A previously published extensive strength data set of Ankara andesite[10] is used throughout the paper to give numerical illustrations of these statistical models.

2. Review of the classic nonlinear regression

There are many occasions in science and engineering where a response variable y (also known as a dependent or measured variable) is modelled as an empirical or theoretical function $f(\cdot)$ of some predictor variables x (also known as covariates or independent variables) with parameters β:

$$y = f(x; \beta). \tag{1}$$

For the case of measured y values, Eq. (1) represents an ideal deterministic situation where function $f(\cdot)$ perfectly predicts the measured values, i.e. a situation when there is no variability in y. In reality however, variability in the measured values of y means that such observations – i.e. pairs of (x_i, y_i) with $i = 1,2,3,\ldots,n$ –will not exactly conform to Eq.(1). In such conditions, we can assume that the functional relationship holds only on average, i.e.

$$E(y_i|x_i) = f(x_i; \beta), \tag{2}$$

where $E(y_i|x_i)$ is the expected value (i.e. mean) of the observed y_i's, conditional on (i.e. knowing) the value of x_i's. However, we consider that at any value of x the corresponding observations will deviate from the expected value due to the presence of some random error ε. Consequently, Eq. (1) can be re-written as

$$y_i = E(y_i|x_i) + \varepsilon_i = f(x_i; \beta) + \varepsilon_i, \tag{3}$$

which gives the relationship between the dependent and independent variables in the presence of variability denoted by ε. Eq. (3) is the regression model, and is called a nonlinear regression model if the function $f(\cdot)$ is nonlinear in the parameters β; it is important to note that the nonlinearity does not refer to the relation between the dependent and the independent variables. Further assumptions often made about the errors ε_i are that they are independent and identically distributed (i.i.d) normal random variables with zero mean and variance equal to that of the response variables, i.e.

$$\varepsilon_i \overset{i.i.d}{\sim} \text{Normal}\,(0, \varsigma_y^2), \tag{4}$$

where ς_y^2 is the variance of the y values. In this context independent means that observing a particular value (i.e. realization) of ε provides no information about other ε values, and identically distributed means that each ε has the same probability distribution as the others. Also, note that the errors are assumed to be homoscedastic, i.e. their variance is constant across all values of the independent variable x. The regression model described by Eqs. (3) and (4) can be summarized with the alternative notation

$$y_i \Big| x, \beta \overset{i.i.d}{\sim} \text{Normal}\,(f(x_i; \beta), \varsigma_y^2), \tag{5}$$

which shows that the dependent variable y, given the independent variable x and parametersβ, is normally distributed with mean $f(x_i, \beta)$ and variance ς_y^2. This alternative notation will be used again in Section 3 during introduction of the Bayesian regression model and its comparison to the classical model.

2.1. Estimating the model parameters

In the classical (frequentist) statistical framework, the values of the parameters β and the variance of the errors ς_y^2 are assumed to be fixed but unknown values that can be estimated according to different criteria. Minimizing the sum of the squares of the errors (least-squares, or LS) is by far the most commonly used estimation method in engineering, to the extent that regression analysis and least squares analysis are often regarded as synonymous. However, it is critical to realize that a regression model describes the relationship between the dependent and independent variables in presence of randomness, and that LS is only one of the available methods for fitting a regression model, i.e. estimating the parameters.[11–13] Alternative methods of parameter estimation include the method of moments or maximum likelihood estimation (MLE), but a discussion of such alternatives and how they relate to LS is beyond the scope of this paper; these techniques are well presented elsewhere.[11–13]

In an LS analysis, the difference between an observed y value and the fitted function is termed a residual, and the residual sum of squares (RSS) is written as