GSP 282

Selected Papers from the Proceedings of Geo-Risk 2017

Geo-Risk 2017

Keynote Lectures

Edited by

D. V. Griffiths, Ph.D., P.E., D.GE
Gordon A. Fenton, Ph.D., P.Eng
Jinsong Huang, Ph.D.
Limin Zhang, Ph.D.

ASCE

G-I GEO-INSTITUTE

GEOTECHNICAL SPECIAL PUBLICATION NO. 282

GEO-RISK 2017

KEYNOTE LECTURES

SELECTED PAPERS FROM SESSIONS OF
GEO-RISK 2017

June 4–7, 2017
Denver, Colorado

SPONSORED BY

Geo-Institute of the American Society of Civil Engineers

EDITED BY
D. V. Griffiths, Ph.D., P.E., D.GE
Gordon A. Fenton, Ph.D., P.Eng.
Jinsong Huang, Ph.D.
Limin Zhang, Ph.D.

Published by the American Society of Civil Engineers

Published by American Society of Civil Engineers
1801 Alexander Bell Drive
Reston, Virginia, 20191-4382
www.asce.org/publications | ascelibrary.org

Errata: Errata, if any, can be found at https://doi.org/10.1061/9780784480694

ISBN 978-0-7844-8069-4 (PDF)
Manufactured in the United States of America.

Preface

Interest and use of probabilistic methods and risk assessment tools in geotechnical engineering has grown rapidly in recent years. The natural variability of soil and rock properties, combined with a frequent lack of high quality site data, makes a probabilistic approach to geotechnical design a logical and scientific way of managing both technical and economic risk. The burgeoning field of geotechnical risk assessment is evidenced by numerous publications, textbooks, dedicated journals and sessions at general geotechnical conferences. Risk assessments are increasingly becoming a requirement in many large engineering construction projects. Probabilistic methods are also recognized in design codes as a way of delivering reasonable load and resistance factors (LRFD) to target allowable risk levels in geotechnical design.

This Geotechnical Special Publication (GSP), coming out of the *Geo-Risk 2017* specialty conference held in Denver, Colorado from June 4-7, 2017, presents eight outstanding contributions from the keynote speakers. Four of the contributions are from practitioners and the other four are from academics, but they are all motivated by a desire to promote the use of risk assessment and probabilistic methodologies in geotechnical engineering practice. Honor Lectures are presented by Greg Baecher (Suzanne Lacasse Lecturer) on Bayesian thinking in geotechnical engineering and Gordon Fenton (Wilson Tang Lecturer) on future directions in reliability based design. The reliability-based design theme is continued by Dennis Becker who includes discussion of risk management, and Brian Simpson, who focuses on aspects of Eurocode 7 and the rapidly growing importance of robustness in engineering design. The evolution and importance of risk assessment tools in dam safety is covered in lectures by John France and Jennifer Williams, and Steven Vick. The challenges of liquefaction modeling and the associated risks of problems due to instability and deformations are covered in lectures by Hsein Juang and Armin Stuedlein.

These contributions to the use of risk assessment methodologies in geotechnical practice are very timely, and will provide a valuable and lasting reference for practitioners and academics alike.

All the papers in this GSP went through a rigorous review process. The contributions of the reviewers are much appreciated.

The Editors

D.V. Griffiths, Ph.D., P.E., D.GE, F.ASCE, Colorado School of Mines, Golden, CO, USA

Gordon A. Fenton, Ph.D., P.Eng., FEIC, FCAE, M.ASCE, Dalhousie University, Halifax, Canada

Jinsong Huang, Ph.D., M.ASCE, University of Newcastle, NSW, Australia

Limin Zhang, Ph.D., F.ASCE, Hong Kong University of Science and Technology, PR China

Acknowledgments

The following individuals deserve special acknowledgment and recognition for their efforts in making this conference a success

- Conference Chair: D.V. Griffiths, Colorado School of Mines, Golden, Colorado, USA
- Conference Co-Chair: Gordon A. Fenton, Dalhousie University, Halifax, Canada
- Technical Program Chair: Jinsong Huang, University of Newcastle, NSW, Australia
- Short-Courses: Limin Zhang, Hong Kong University of Science and Technology
- Student Program co-Chairs: Zhe Luo, University of Akron; Jack Montgomery, Auburn University
- Sponsorships and Exhibits Chair: Armin Stuedlein, Oregon State University

The Editors greatly appreciate the work of Ms. Helen Cook, Ms. Leanne Shroeder, Ms. Brandi Steeves, and Mr. Drew Caracciolo of the ASCE Geo-Institute for their administration of many important conference organizational issues, including management of the on-line paper submissions, the conference web site and sponsorship.

Contents

Bayesian Thinking in Geotechnics

Gregory B. Baecher, Ph.D., M.ASCE[1]

[1]Glenn L. Martin Institute Professor of Engineering, Univ. of Maryland, College Park, MD 20003. E-mail: gbaecher@umd.edu. ORCID ID 0000-0002-9571-9282.

Abstract

The statistics course most of us took in college introduced a peculiar and narrow species of the subject. Indeed, that species of statistics—usually called, Relative Frequentist theory—is not of much use in grappling with the problems geotechnical engineers routinely face. The sampling theory approach to statistics that arose in the early 20th C. has to do with natural variations within well-defined populations. It has to do with frequencies like the flipping of a coin. Geotechnical engineers, in contrast, deal with uncertainties associated with limited knowledge. They have to do with the probabilities of unique situations. These uncertainties are not amenable to Frequentist thinking; they require Bayesian thinking. Bayesian thinking is that of judgment and belief. It leads to remarkably strong inferences from even sparse data. Most geotechnical engineers are intuitive Bayesians whether they know it or not, and have much to gain from a more formal understanding of the logic behind these straightforward and relatively simple methods.

BAYESIAN THINKING

Most geotechnical engineers are intuitive Bayesians. Practical examples of Bayesian thinking in site characterization, dam safety, data analysis, and reliability are common in practice; and the emblematic *observational approach* of Terzaghi is a pure Bayesian concept although in a qualitative form (Lacasse 2016).

The statistics course one took in college most likely introduced a peculiar and narrow form of statistics, generally known as *Relative Frequentist* theory or *Sampling Theory* statistics. In the way normal statistics courses are taught, one is led to believe that this is all there is to statistics. That is not the case. As one of the reviewers of this paper said, it's not your fault if you haven't thought about Bayesian methods until now; and it's not too late.

This traditional frequentist form of statistical thinking is not particularly useful except in narrowly defined problems of the sort one finds in big science, like medical trials, or in sociological surveys like the US Census. It is tailored to problems for which data have been acquired through a carefully planned and randomized set of trials. It is tailored to aleatory uncertainties, that is, uncertainty dealing with variations in nature. This almost never describes the problems a normal person faces, and especially not geotechnical engineers. Most geotechnical uncertainties are epistemic: they deal with limited knowledge, with uncertainties in the mind not variations in nature.

Two concepts of probability. The reason that college statistics courses deal with this peculiar form of statistics and not something more useful in daily life has to do with intellectual battles in the history of probability, and in how the pedagogy of statistical teaching evolved in the early 20thC. Even though concepts of uncertainty, inference, and induction arose in antiquity, what we think of as modern *probability theory*, at least its mathematical foundations, arose only around 1654. Hacking (1975) and Bernstein (1996) trace this history.

From that time, two concepts of probability evolved in parallel. These deal with different problems, and while they have evolved to use the same mathematical theory, and while they are commonly confused with one another, in fact they are philosophically distinct. One concept, the one taught in undergraduate courses, deals with the relative frequency with which particular events occur in a long series of similar trials. For example, if you roll a pair of dice a thousand times, "snake eyes" (double-1) will occur in about 8.3% of the tosses. This is the sort of probability that is involved with large clinical trials. One exposes 1000 subjects to a test drug, and 1000 subjects to a placebo, and then compares the frequency with which particular outcomes occur in each group.

The other concept of probability deals with degrees of belief that one should rationally hold in the likely outcome of some experiment or in the truth of a proposition. This species of statistics has nothing to do with frequencies in long series of similar trials, but rather with how willing one is to make decisions or to take action when faced with uncertainties. For example, frequency statistics might be used to describe the rates of false positive or false negative results when a medical test is applied to a large number of subjects; but the probability that you as a unique individual are sick if a diagnostic test comes back positive is not a matter of frequencies, it is a matter of one unique individual, namely, you. You are either sick or well. Probability in this case is a matter of the degree of belief about which of those two conditions you think obtains. Vick (2002) interprets this theory of degrees-of-belief as a formalization of "engineering judgment."

Scope of this paper. This paper focusses on inferences which at first glance seem difficult or impossible to make—and indeed they are, using frequentist thinking. But they are easy when viewed through the lens of Bayesian thinking. Bayesian methods have been used across the spectrum of geotechnical applications since the 1970's, as reflected in the early work of Tang (Lacasse *et al.* 2013), Wu (2011), Einstein (Einstein *et al.* 1978), Marr (2011), and many others. These methods have revolutionized many fields of engineering and continue to do so (McGrayne 2012). "Clippy" the annoying Microsoft self-help wizard was a Bayesian app. Spam filtering of your email inbox is, too. The Enigma Code of the German Kriegsmarine was broken using Bayesian methods at Bletchley Park. And the wreckage of Air France flight 447 was found using a Bayesian search algorithm. Recent reviews of the use of Bayesian methods in geotechnical engineering have been provided by Yu Wang (2016), Zhang (2016), and Juang and Zhang (2017). For reasons of space and to avoid complicating the 'message,' advanced topics in Bayesian methods such as belief nets and Markov-chain Monte-Carlo are not discussed here.

LEARNING FROM EXPERIENCE

The application of statistics to practical problems is of two sorts. On the one hand, we use statistics to describe the variability of data using summaries such as measures of central tendency and spread, or frequency distributions such as histograms or probability density functions. On the other hand, we use statistics to infer probabilities over properties of a population that we have not observed and based on a limited sample that we have observed. It is this latter meaning of statistics that we deal with here. It is the inductive use of statistics, which in the 19th C. was called *inverse reasoning*.

Bayes' rule. Bayes' Rule tells us that the *weight of evidence* in observations is wholly contained in the *Likelihood Function*, that is, in the conditional probability of the observations, given the true state of nature (see Hacking 2001 for a comprehensible introduction). Statisticians might call the true state of nature, *the hypothesis*, thus,

$$P(H|data) = N \times P(H) \times P(data|H) \tag{1}$$

in which *H*=the hypothesis is true, $P(H)$=the (prior) probability of the hypothesis being true before seeing the data, $P(data|H)$=the probability of the observed data were the hypothesis true, $P(H|data)$=the (posterior) probability of the hypothesis being true, and *N*=a normalizing constant.

The term, $P(data|H)$ is called, the *Likelihood*,

$$Likelihood\ of\ H\ for\ the\ observed\ data\ =\ P(\ data\ |\ H\ ) \tag{2}$$

The Likelihood might be thought of as the degree of plausibility of the data in light of the hypothesis (Schweckendiek 2016). In the Bayesian literature, the Likelihood is sometimes written as, $L(H|data)$ (O'Hagan and Forster 2004).

The normalizing constant in Eq. (1) is just that which makes the sum of the probabilities for and against the hypothesis (*H*) equal 1.0. In practical applications, *N* is often and most easily obtained numerically, but in the simple case above it can be calculated from the Total Probability Theorem as, $N = \{P(H) \times P(data|H) + P(\bar{H}) \times P(data|\bar{H})\}$, in which $\bar{H}$=the hypothesis is not true.

Dividing Eq. (1) by its complement for *not-H*,

$$\frac{P(H|data)}{P(\bar{H}|data)} = \frac{P(H)}{P(\bar{H})} \times \frac{P(data|H)}{P(data|\bar{H})} \tag{3}$$

The normalizing constant, *N*, which is same in numerator and denominator, cancels out. In every-day English, this reads, "the posterior odds for the hypothesis equals the prior odds times the likelihood (LR) ratio." What one thought before seeing the data is entirely contained in the prior odds, while the weight of information in the data is entirely contained in the *Likelihood Ratio*. The Likelihood Ratio is the relation of the Likelihood for a true hypothesis to that for a false hypothesis, $LR = P(data|H)/(P(data|\bar{H})$.

The weight of evidence. The crucial thing about Eq. (3) is the unique role of the Likelihood Ratio. The Likelihood Ratio contains the entire weight of evidence contained in the observations. This is true whether there is one observation or a large number, which means that inferences can be made even if the data are relatively weak (Jaynes 2003), and sometimes these inferences from weak data can actually be relatively strong (Good 1996).

Sir Harold Jeffreys (1891-1989), late Professor of Geophysics at Cambridge and ardent defender of Bayesianism (although an opponent of continental drift), proposed that this weight of evidence—for purposes of testing scientific hypotheses—be characterized as in Table 1. Whether one agrees with the verbal descriptions and corresponding LR's is left to the reader's judgment. In the modern literature, this weight of evidence in the *LR* is called, the *Bayes factor* (Kass and Raftery 1995). For Millennial readers, the logarithm of the Bayes Factor will be recognized as the decimal version of binary *bits* of information in the Shannon and Weaver (1949) sense. For readers of the author's age the odds might be compared to bets at the horse track.

Table 1. Qualitative scale for the degree of support provided by evidence (Jeffreys 1998)

LR = Likelihood Ratio (from-to)		Weight of evidence to support the hypothesis
1	10	Limited evidence
10	100	Moderate evidence
100	1000	Moderately strong evidence
1000	10,000	Strong evidence
10,000	∞	Very strong evidence

SIMPLE BAYESIAN INFERENCE (A/K/A "BAYESIAN UPDATING")

A straightforward but powerful example of simple Bayesian inference is given by the work of Chen and Gilbert (2014) on Gulf of Mexico offshore structures. These inferences might be based on laboratory tests, *in situ* measurements, performance data, or even quantified expert opinion. This contrasts to an earlier time when such inferences were almost always made using Frequentist methods (Lumb 1974).

Chen and Gilbert use Bayes' Rule to update bias factors in engineering models for pile system capacity based on observed performance in Gulf of Mexico hurricanes. The initial work included events up to Hurricane Andrew in 1992, and in a subsequent paper up to more recent hurricanes between 2004 and 2008 (Chen and Gilbert in press). The analysis addresses model bias in four predictions: wave load, base shear capacity, overturning in clay, and overturning in sand.

The question addressed is, what is the systematic bias in the predictions of the engineering models being used to forecast pile system capacity, given observations of how the platforms performed in various storm events, and given the prior forecasts of how they would perform. Rewritten in the notation of the present paper (O'Hagan and Forster 2004),

$$f(B|data) = N \times f(B) \times P(data|B) \tag{4}$$

in which, *f(.)* is a probability density function (pdf), B = model bias factor, $data$ = the observed performance of the pile system, and the normalizing constant, *N*, is that which makes the integral over all B equal to 1.0, $N = \left\{\int_{-\infty}^{\infty} f(B) \times P(data|B)\, dB\right\}^{-1}$ (*i.e.,* the area under $f(B|data)$ has to be unity for the pdf to be proper). This is simply a restatement of Eq. (1).

The updated pdf's of Figure 1 show how the storm loading data led to a re-evaluation of the uncertainty in the model bias factors for wave loading and to overturning. The dotted curves show the prior pdf's of model bias. The solid curves show the posterior or "updated" pdf's. In both cases the performance data led to a lowering of the best estimate of model bias and to a slight reduction in the uncertainty in the bias. As noted by the authors, however, there is no assurance that observed performance will always reduce uncertainty. If the observations are inconsistent with what was thought *ex ante*, the variance of the pdf might, in fact, increase rather than decrease. The weight of evidence in the Likelihood Function works both ways, either in favor of an hypothesis or opposed to it. Similar applications are provided by Zhang (2004) inferring pile capacities based on incomplete load tests, and by Huang, *et al.* (2016) inferring the reliability of