Selection of Characteristic Values for Rock and Soil Properties using Bayesian Statistics and Prior Knowledge: Theory, Software and Application Examples

Yu Wang
City University of Hong Kong

Innotech Forum on Geotechnology, Hong Kong, 17 April 2018
Acknowledgement

• Joint TC205/TC304 Working Group on “Discussion of statistical/reliability methods for Eurocodes”

• Discussion group on “Selection of characteristic values for rock and soil properties using Bayesian statistics and prior knowledge” led by Yu Wang

• Discussers (alphabetical order):
  – Marcos Arroyo (Spain), Zijun Cao (China), Jianye Ching (Taiwan), Tim Länsivaara (Finland), Trevor Orr (Ireland), Kok-Kwang Phoon (Singapore), Hansruedi Schneider (Switzerland), Brian Simpson (UK)
Outline

• Background and Motivation
• Bayesian Methods
• Software and Android App
• Illustrative Examples
• Concluding Remarks
Background

• Soil or rock properties are necessary input to geotechnical designs and analyses.
• During site investigation, soil or rock properties are often measured sparsely due to time, resource, or technical constraints.
• How to select characteristic values for soil or rock properties from limited data?
Example: what $C_u$ value to use?

Figure 1. Characteristic $c_u$ values for determination of the shaft and base resistance of a pile (Frank et al. 2004).
Example: what SPT profile to use?

- Bond and Harris (2008) asked about one hundred engineers to select the SPT profile.
- The maximum profile was about 3 to 5 times greater than the minimum one.

*Figure 2.* Results of SPTs on London and Lambeth clays, with engineers’ interpretations of the characteristic value (Bond and Harris 2008).
Motivation

• To select characteristic values for soil or rock properties in a rational and consistent manner using Bayesian statistics and prior knowledge

• Eurocode 7, Clause 2.4.5.2(10):
  “If statistical methods are employed in the selection of characteristic values for ground properties, such methods should differentiate between local and regional sampling and should allow the use of a priori knowledge of comparable ground properties.”

• Bayesian method combines information from different sources in a rational manner
Geotechnical Site Investigation

**PROCEDURE**
- I: Desk-study
- II: Site reconnaissance
- III: In-situ investigation
- IV: Laboratory testing
- V: Interpretation of site observation data
- VI: Inferring geotechnical properties and underground strata

**INFORMATION**
- Prior knowledge (e.g., geological maps, geotechnical reports, engineering experience and judgment, etc.)
- Site observation data (e.g., data from test boring, in-situ testing and/or laboratory testing)
- Transformation model (e.g., empirical regression)
- Updated knowledge

**UNCERTAINTY**
- Uncertainties in existing information
- Inherent variability
- Measurement error
- Statistical uncertainty
- Transformation uncertainty

**CHALLENGES**
- Multi-sources information
- Various uncertainties
- Limited site-specific data

How to combine them Systematically?
Bayesian Framework for Geotechnical Site Investigation

- Bayesian approach combines systematically information from different sources for uncertainty quantification

Diagram:

- Prior knowledge
  - Prior distribution
    - Likelihood model $M_L$
      - Likelihood function
        - Site observation data $X_M$
          - Statistical uncertainty and measurement errors
    - Transformation model $M_T$:
      - Inherent variability
    - Transformation uncertainty
  - Posterior distribution
    - Posterior knowledge

Bayesian approach combines systematically information from different sources for uncertainty quantification.
Bayesian Equivalent Sample Method

• **Likelihood Function**
  - Probability model $M_p$ of $X_D$ with model parameters $\Theta_p$
    (e.g., random variable, random field)
  - Transformation model, $X_D = f_T(X_M; \varepsilon_T)$

• **Prior Distribution**
  - **Non-informative** — e.g., joint uniform distribution
  - **Informative** — subjective probability assessment framework (Cao et al., 2016)

• **Posterior Distribution**

\[
P(\Theta_p \mid \text{Data, Prior}) = K \times P(\text{Data} \mid \Theta_p, \text{Prior}) \times P(\Theta_p \mid \text{Prior})
\]

\[
P(X_D \mid \text{Data, Prior}) = \int P(X_D \mid \Theta_p) P(\Theta_p \mid \text{Data, Prior}) d\Theta_p
\]

Markov Chain Monte Carlo simulation

Equivalent samples of $X_D$
Example: Estimating $E_u$ of clay from SPT-N

- Limited site-specific SPT-N data
- Formulation of likelihood function
- Formulation of prior distribution
- Posterior distribution of $E_u$
- Bayesian equivalent samples of $E_u$ from Markov Chain Monte Carlo simulation
A statistically homogenous clay layer

**Undrained Young’s modulus**

- \( E_u \sim \text{LogNormal} (\mu, \sigma) \)
- \( \mu = \text{Mean value of } E_u \)
- \( \sigma = \text{Standard deviation of } E_u \)

- \( \ln E_u \sim \text{Normal} (\mu_N, \sigma_N) \)
- \( \mu_N = \text{Mean value of } \ln E_u \)
- \( \sigma_N = \text{Standard deviation of } \ln E_u \)

**Depth (D)**

\[
\ln E_u = f_N (\mu_N, \sigma_N) = f(\mu, \sigma)
\]
Likelihood Function

\[ \ln\left(\frac{E_u}{p_a}\right) = 0.63 \ln N_{SPT} + 2.96 + \varepsilon_T \]

\( E_u \) values are obtained from pressuremeter tests

\( p_a \): atmospheric pressure

\( \varepsilon_T \): transformation uncertainty

Normal random variable

(Phoon and Kulhawy 1990)

(Ohya et al. 1982, Kulhawy and Mayne 1990)
Likelihood Function

\[ \ln N_{SPT} = f_T(\ln E_u) + \varepsilon \]

\[ \ln E_u = f(\mu, \sigma) \]
Normal variable

\[ \ln N_{SPT} \sim \text{Normal} (\mu_{SPT}, \sigma_{SPT}) \]
\[ \mu_{SPT} \sim f_m(\mu, \sigma); \sigma_{SPT} \sim f_s(\mu, \sigma) \]

\[ P(\ln N_{SPT} | \mu, \sigma) = \text{Normal Probability Density Function (PDF)} \]

\[ DATA = \{ \ln N_{SPT,i}, i=1,...,n_s \} \]

Likelihood function
\[ P(DATA | \mu, \sigma) = \prod_{i=1}^{n_s} P(\ln N_{SPT,i} | \mu, \sigma) \]

Measurement errors
Statistical uncertainty
Inherent variability
Prior Distribution

• When only respective typical ranges $[\mu_{\text{min}}, \mu_{\text{max}}]$ and $[\sigma_{\text{min}}, \sigma_{\text{max}}]$ of $\mu$ and $\sigma$ are available

$$P(\mu, \sigma) = \begin{cases} \frac{1}{\mu_{\text{max}} - \mu_{\text{min}}} \times \frac{1}{\sigma_{\text{max}} - \sigma_{\text{min}}} & \text{For } \mu \in [\mu_{\text{min}}, \mu_{\text{max}}] \\
0 & \text{and } \sigma \in [\sigma_{\text{min}}, \sigma_{\text{max}}] \\
& \text{Others} \end{cases}$$
Posterior Distribution

- **Bayes’ Theorem**
  \[ P(\mu,\sigma|DATA) = KP(DATA|\mu,\sigma)P(\mu,\sigma) \]
  Posterior distribution Likelihood function Prior distribution

  where K is a normalizing constant

- **Total Probability Theorem**
  \[ P(E_u|DATA, PRIOR) = \int P(E_u|\mu,\sigma) P(\mu,\sigma|DATA,PRIOR)d\mu d\sigma \]
  LogNormal PDF \[ P(\mu,\sigma|DATA) \]
• Markov Chain Monte Carlo Simulation (MCMCS) is a numerical process that simulates a sequence of samples of a random variable as a Markov Chain

Limiting stationary PDF of the Markov Chain = PDF of the target random variable (e.g., \( P(E_u | DATA, PRIOR) \))

• A feasible way to generate samples from an arbitrary PDF (e.g., \( P(E_u | DATA, PRIOR) \))
### BEST EXCEL Add-In
(Bayesian Equivalent Sample Toolkit)

#### Key Features:
- **12 Built-in model**
- **User-defined model**

#### Table Example:

<table>
<thead>
<tr>
<th>Type</th>
<th>S/N</th>
<th>Design Property</th>
<th>Distribution Type</th>
<th>Measured Property</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clay</td>
<td>1</td>
<td>$E_u$</td>
<td>Log-normal</td>
<td>$N$</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$r$</td>
<td>Log-normal</td>
<td>$L_i$</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>$r$</td>
<td>Log-normal</td>
<td>$k_D$</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>$S_{u0}$</td>
<td>Log-normal</td>
<td>$N$</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>$\sin \phi'$</td>
<td>Normal</td>
<td>$\Delta u_t$</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>$k_0$</td>
<td>Normal</td>
<td>$k_D$</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>$k_0$</td>
<td>Normal</td>
<td>$N$</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>$k_0$</td>
<td>Normal</td>
<td>$\frac{q_c}{P_a}$</td>
</tr>
<tr>
<td>Sand</td>
<td>1</td>
<td>$\phi'$</td>
<td>Normal</td>
<td>$\frac{q_c}{P_a}$</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$\phi'$</td>
<td>Normal</td>
<td>$N_{100}^{0.66}$</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>$E$</td>
<td>Normal</td>
<td>$N_{0.16}$</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>$E_D$</td>
<td>Log-normal</td>
<td>$N$</td>
</tr>
</tbody>
</table>

#### Transformation Model

$$X_M = a \ln (X_D) + b + \epsilon$$

$X_M$ is the measured property or a function of the measured property. $X_D$ is the design property which follows log-normal distribution.

$a$ and $b$ are the transformation coefficients. $\epsilon$ is a Gaussian random variable.

#### Click to save.
A clay site of US National Geotechnical Experimentation Sites at Texas A&M University

DATA = 5 SPT data

Uniform PRIOR with
\[ \mu \in [5 \text{MPa, 15MPa}] \]
\[ \sigma \in [0.5 \text{MPa, 13.5MPa}] \]
(Phoon and Kulhawy 1999a and 1999b)

42 Pressuremeter test data
| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U |
| SPT N | 30 | 14 | 14 | 19 | 14 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

**Note:** The text in the table appears to be in Chinese.
Application Example

![Graph showing cumulative distribution function (CDF) of Young's modulus E_u with 5 SPT data & prior knowledge, 42 pressuremeter test data, mean - stdev, and 50% probability levels.](image)
BEST App in Android Phone
Concluding Remarks

• Development of practical Bayesian statistical methods and user-friendly tools for selection of characteristic values for soil or rock properties
  ➢ Bayesian equivalent sample method
  ➢ Quantification of prior knowledge
  ➢ User-friendly software (BEST) in EXCEL and Android App

• BEST is applicable to direct or indirect measurements

• Random field modeling of spatial variability is not covered
References


“Probability theory is nothing but common sense reduced to calculation.”

Pierre-Simon Laplace
(1749-1827)
Thank you!