

Real World Applications of Probability Theory

Choi Tim Antony Yung

Department of Electrical and Computer Engineering
California State Polytechnic University, Pomona

Pomona, CA, USA

orcid.org/0000-0002-4361-5043

Abstract—This document provides three use case of probability theory across different fields.

Index Terms—Probability Theory, Bayesian Theory, User Interface Design, Analytic Thinking, Probabilistic Classification

I. INTRODUCTION

This document is a brief summary of some real world applications of probability theory. In particular, two cases of determining correlation between different characteristics in different groups of people via sampling and one case of implementation of classification method base on Bayes' Theorem.

II. USER INTERFACE DESIGN

It was often a frustration for designer that a product does not work the way it was intended to due to user errors from negligence of the user's manual, manifested in one of the common response to questions: "RTFM." In an article aptly titled "Life is too short to RTFM," Blackler et al. [1] concluded that users perceive both manuals and excess features to be the source of their negative experiences and suggested reducing the need for user to access manual by removing excess features and improving other aspect of usability. The article was, in part, based on a study that collected 170 responses from different people. The researchers then applied probability theory on these samples to determine that there are differences in the probability between different demographics on whether or not they read (or claim to have read) manuals.

III. ANALYTIC THINKING

Deepak Chopra, an author and a leader in the New Age movement, have a Twitter feed one might found to contain an abundance of "seemingly impressive assertions that are presented as true and meaningful but are actually vacuous," which was referred to as "pseudo-profound bullshit" by Dr. Gordon Pennycook of University of Waterloo. In a study, Pennycook et al. [2] generated a list of ten meaningless statements constructed with profound-sounding words and words used in Deepak Chopra's tweets that retained a syntactic structure. Pennycook then, from those meaningless statements, created a 5-point Bullshit Receptivity (BSR) scale. Respondents of the study were asked to rate profoundness of the meaningless statements as a random variable ranging from one to five, with one meaning not at all profound and five meaning very profound. The responses of the respondents regarding the

profoundness of the ten meaningless statement is then recorded and the mean score of the ten statements is the BSR score. Other questions were also used to determine the characteristics of the respondents. From this study, Pennycook found that the higher a respondent's BSR score is, the more likely they are to "hold religious and paranormal beliefs" and "to endorse complementary and alternative medicine."

IV. CLASSIFICATION

Classification is an approach used in machine learning and data mining to group data under different labels. One of the way data can be classified is to use a naive Bayes classifier. Yang [3] described how Bayes' theorem can be used to determine the class of data from the probability of data to be classified in a certain class given that the data have certain attributes i.e. $P(C_i|A_1 = v_1 \cap A_2 = v_2 \cap \dots \cap A_n = v_n)$ with C_i being a certain class, A_n being a certain attribute and v_n being a certain value associated with the attribute A_n . Many data do not share the exact value to attributes with any existing data. For example, to determine the classification of a man aged 30, we need to determine the probability of a person belonging to a certain classification given that person is a man and is aged 30, who may not be in the existing data set. Bayes' theorem allows finding that probability $P(C_i|A_{\text{sex}} = \text{male} \cap A_{\text{age}} = 30)$ by finding the probability $\frac{P(A_{\text{sex}} = \text{male} \cap A_{\text{age}} = 30|C_i) \times P(C_i)}{P(A_{\text{sex}} = \text{male} \cap A_{\text{age}} = 30)}$. Since it is compared among different class i.e. different indice of i , the denominator can be dropped i.e. $P(A_{\text{sex}} = \text{male} \cap A_{\text{age}} = 30|C_i) \times P(C_i)$. Assuming naively mutual independent attributes, the probability can be further broken down to $P(A_{\text{sex}} = \text{male}|C_i) \times P(A_{\text{age}} = 30|C_i) \times P(C_i)$, all of which are known.

REFERENCES

- [1] A. L. Blackler, R. Gomez, V. Popovic and M. H. Thompson, "Life Is Too Short to RTFM: How Users Relate to Documentation and Excess Features in Consumer Products," in *Interacting with Computers*, vol. 28, no. 1, pp. 27-46, Jan. 2016, doi: 10.1093/iwc/iwu023.
- [2] G. Pennycook, J. A. Cheyne, N. Barr, D. J. Koehler, and J. A. Fugelsang, "On the reception and detection of pseudo-profound bullshit," *Judgment and Decision Making*, vol. 10, no. 6, pp. 549-563, Nov. 2015.
- [3] F. Yang, "An Implementation of Naive Bayes Classifier," 2018 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 2018, pp. 301-306, doi: 10.1109/CSCI46756.2018.00065.