

A Multidimensional Technique for Measuring Consensus Within Groups via Conditional Probability

Mushtaq K. Abd Al-Rahem

Dissertation submitted to the
Eberly College of Arts and Sciences
at West Virginia University
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in
Mathematics

Marjorie Darrah, Ph.D, Chair
Harvey Diamond, Ph.D
Hong-Jian Lai, Ph.D
Michael Mays, Ph.D
James Nolan, Ph.D

Department of Mathematics

Morgantown, WV
2017

Keywords: Index, Disagreement, Consensus, Measure, Likert
scale, Conditional Probability, n -Dimension space.

Copyright © 2017 Mushtaq K. Abd Al-Rahem

ProQuest Number: 10269283

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10269283

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

ABSTRACT

A Multidimensional Technique for Measuring Consensus Within Groups via Conditional Probability

Mushtaq K. Abd Al-Rahem

A recent increase in the use of the term “consensus” in various fields has led researchers to develop various ways to measure the consensus within and across groups depending on the areas. Numerous studies use the mean or the variance alone as a measure of consensus, or lack of consensus. Most of the time, high variance is viewed as more disagreement in a group. Using the variance as a measure of disagreement is meaningful in an exact comparison cases (same group, same mean). However, it could be meaningless when it is used to compare groups that have different sizes, or if the mean is different. In this thesis, we establish the fact that the range of the variance is a function of the mean, we present a new index of disagreement, ϕ , and measure of consensus, $\psi = 1 - \phi$, that depend on both, the mean and the variance, by utilizing the conditional distribution of the variance for a given mean. Initially, this new index is developed for comparison of data collected using a Likert scale of size 5. This new measure is compared with the results of two other known measures, to show that in some cases they agree, but in other cases the new measure provides additional information. Next, to facilitate generalization, a new algorithmic method to determine the index using a geometric approach is presented. The geometric approach makes it easier to compute the measure of consensus and provides the foundational ideas for generalizing the measure to Likert scales for any n . Finally, a multidimensional computational technique was developed to provided the final step of generalization to Likert scales of any n .

DEDICATION

To My Mother and Father,

I vividly remember how you dressed me up in new clothes and shoes in anticipation for my first-grade primary school.

My Children, Ridha, Tuqa, and Aya,

your bedtime was my work time

Acknowledgments

First and above all, praise and gratitude to Almighty God for providing this opportunity and granting me the ability to conclude my research successfully. Without God blessings, this accomplishment would not have been realized.

I would especially like to show my utmost appreciation to my supervisor, Professor Marjorie Darrah, for her generous encouragement, guidance, and invaluable assistance during my study. Professor Darrah, I really appreciate your mentorship and inspiration toward the field of “Consensus Measurement” and your suggestions throughout the entire period of the research. You made my defense one of the most exciting, enjoyable and challenging moments of my life.

I wish to acknowledge the help, support, and encouragement of all staff and members of the Department of Mathematics, most importantly the committee members. Words are not enough in describing the stimulating discussions and insightful suggestions that led to the inception of this dissertation. In particular, your comments and encouragements broadened my perspectives by indicating ways that led me to clarify my ideas, outcomes, and presentations. Thank you.

Also, I gratefully acknowledge to Dr. Yoshio Akiyama and Dr. James Nolan whose research provided a strong foundation on which to build. Dr. Nolan, your help and support during the first paper means a lot to me. Thank you and I hope I will work again with you in future.

Noor, you know you are not only my wife but my life. Thank you for being so patient during the research period. You consistently challenging me to be a better person, and for reminding me what life is all about when I'm overwhelmed and stressed. You are my perfect complement. I love you.

I would be remiss if I did not also express my sincere gratitude to all my professors, colleagues and friends who extended their support throughout the period of my research study. Last but not least, I gratefully acknowledge all kinds of support from the Higher Committee for Education Development in Iraq (HCED).

Thank You All

List of Figures

2.1	The Range of the Variance	22
4.1	Example for The Area $A_1(v)$ and $A_2(v)$ in R^2	65
4.2	Example for of Planes used to Compute Areas $A_1(v)$ and $A_2(v)$ in R^3	65

List of Tables

2.1	The Range of Variance for a Given Mean	23
2.2	Sample Cumulative Conditional Distributions for Index of Dis- agreement $\Phi(v m)$ and Measure of Consensus Ψ	33
2.3	Comparison of $Cns(X)$ and r_{WG} with Ψ for Q38 and Q39 . . .	34
2.4	Comparison of $Cns(X)$ and r_{WG} with Ψ for Q1 and Q2	34
2.5	Comparison of $Cns(X)$ and r_{WG} with Ψ for Q23 and Q34 . . .	35
2.6	Climate Survey Results.	37
3.1	Examples of Comparison of Akiyama et al. indices Φ and Ψ and our result Φ_{CG} and Ψ_{CG} for selected values of m and v . . .	53

Contents

1	Introduction	1
1.1	Motivation	3
1.2	Literature Survey	4
2	Paper 1: A method for measuring consensus within groups: An index of disagreement via conditional probability	9
2.1	Introduction	11
2.2	Theoretical Foundations	15
2.2.1	Bounds for the Variance	17
2.2.2	Defining the Index of Disagreement	24
2.2.3	Computing the Disagreement Index Φ and Consensus Measure Ψ	31
2.3	Comparing Ψ with r_{WG} and $Cns(X)$	32
2.4	Conclusions and Future Work	35
3	Paper 2: A Geometric Approach for Computing a Measure of Consensus for Groups	40
3.1	Introduction	42
3.2	Preliminaries	44
3.2.1	Computational Geometry	44
3.2.2	Conditional Probability	46
3.3	Theoretical Foundations	47

3.3.1	Defining the Index of Disagreement	48
3.4	New Algorithms for Computing the Index of Disagreement	50
3.5	Conclusions and Future Work	53
4	Paper 3: Using a-Multidimensional Approach for Generalizing a Consensus Measure to Likert Scales of Any Size n	55
4.1	Introduction	57
4.2	Theoretical Foundations	60
4.2.1	Generalization of Calculations to $n > 5$ for Index of Disagreement (conversely Consensus Measure)	61
4.2.2	Determine the Index of Disagreement	65
4.3	Index of Disagreement in 3D	67
4.4	Numerical Example	68
4.5	Conclusions and Future Work	70
5	Discussions and Open Problems	72

Chapter 1

Introduction

The challenge with measuring consensus in groups, as a field of study, and other sciences that depend on opinions or beliefs is that there is no “exact” right answer, but rather methods that try to describe human behavior through numbers. For many years, researchers spent much time on different approaches to create rules or axioms to be a basis for anyone working in this field. Psychology, for example, has a long history of trying to set or define measurements. Psychological testing began with Darwin’s first book *The Origin of Species* in 1859. It acquired its current widespread definition by Stanley Stevens in 1946 [37], [43].

One of the most common approaches for measuring thoughts is by turning them into numbers by the use of scales. Scales developed for this purpose include; the Borg scale [44], the Guttman scale [27], and the Likert scale [39]. In early years, Likert data was used in ordinal ways and interval ways. Therefore, the mean and the variance of this type of data were determined and handled in meaningful ways [24]. Moreover, Norman [44] gave a convincing argument for the use of parametric methods with Likert data, and Steven in 1946 presented an article on different kinds of scales and allowable statistics, where he included an argument about whether Likert data was ordinal or interval [57].

In this thesis, as in many other consensus measurement studies, our data were collected by using the Likert scale. The Likert scale is used when respondents express their strength of agreement regarding several statements. Typically, response options vary from “Strongly Disagree” to “Strongly Agree.” Moreover, the central point refers to neutrality, and the ratings were made to represent unidirectional positions [15], [44].

Even though those who work in the field of consensus share the same word, consensus, there are variations in the definitions of this term. The main reason for this diversity of the definition of consensus is the extensive use of this term in many different subjects, such as Politics, Economics, and Sciences. An illustration of how several scientific fields use consensus is presented by Lehrer and Wagner [42].

Although some of the theories presented in literature have proven to be insufficient, we find various atomistic, and individualistic methods that have gripped philosophical thought from time to time. Therefore, some factors of consensus are needed to give an explanation of central concepts or direction of thinking. In Social Choice Theory, for example, the discussions and analysis of consensus measurement are introduced by Bosch[13]. Absolute and intrinsic measures of consensus were proposed, analyzed and axiomatically characterized.

In this work, the term of “consensus” references to common opinions, beliefs, or perceptions achieved by a group of individuals, while “disagreement” means the variations of belief, opinion, or attitude. Analytically, consensus and disagreement are opposites. In other words, each one is the inverse of the other, or, mathematically, you can say each one is the complement of the other with respect to one or if you add consensus and disagreement you get one. For instance, if you have a 90% consensus in your group work, it means that there is 10% disagreement within the same group.

For purposes of organization and to keep clarity in check while going a little deeper in the introduction, we divide the rest of the preface into two

sections: the motivation and the literature survey.

1.1 Motivation

From the beginning of humankind, people have not lived in seclusion, but rather in groups. The rising population has led to the necessity of making various decisions, most of which are collective. Most of these examples are evident every day, from important cases such as the relations between countries or where to build new nuclear plants, to relatively minor decisions like which restaurant you and your family members or friends will choose for lunch on a particular day.

Moreover, the cases necessitating the need for group decisions has grown even more after the technology has made the world a global village. Even you agree or not, many times you will find yourself in a new group at one of the social websites or cell phone applications, and you might be a member of a group decision.

To obtain a better collective decision, which realizes a greater contentment among members, it is important that decisions are made with consensus. The reason for this, in many examples, is that it is not advantageous for the group as a whole if the decision is enjoined by a subset of the individuals [4], [21].

Even ignoring or missing a person in a group decision might make a big difference, one percent error in the consensus measurement may lead to the wrong direction in formulating valuable decisions. Consequently, the requirement of right tools to measure the consensus in one group or among individual groups is a major step for getting the agreement or minimizing the disagreement [4].

As we mentioned above, the population is increasing, and the consequence is a rising level of diversity of opinions and ideas. In many cases, therefore, the Yes/No responses seems inadequate more often. In fact, even the five

items of Likert scale is not sufficient to work within particular situations. Accordingly, we deliberately spend a lot of time in this endeavor working to build a robust measure initially for the five point Likert scale and then we set out to generalize the method to any number of Likert scale choices.

1.2 Literature Survey

Seeing that groups usually do not attain total agreement, some researchers seek to measure the “distance” from the consensus [30], [41], [13]. Other studies in this field aim at finding out how much consensus a group possesses or, the determination of the exact ratio of disagreement the board of directors maintains. Furthermore, “Do We Have Consensus?” is the central question of Wisecup [60] for his research in 2011.

In the early years of studying in the area of consensus, researchers were concerned with exploring consensus in small decision-making groups. Hare presented one preliminary study in 1952 [30]. The main idea of his work involved assigning to the group a rank ordering task, and then the consensus is the difference between the average correlation of the level before and after the discussion.

Knutson [41] gave groups five choices extending from the highest conservative to the most liberal policy. The procedure of Knutson works determines the distance from the agreement at the conclusion of a discussion. Each group member was supposed to select only one of the five alternatives, and then compute the total number of positions the members were far from the most agreed-upon policy. In fact, this procedure was first presented by Kline [40] but for six choices of alternative instead of five. Hill [34] measured the consensus in his paper by asking individual group members to give a rating for their level of agreement with the final group decision.

Indeed, Hill is not the first or the only one to utilize the use of the percentage as the main idea for measuring agreement. Since percentage agreement

is not hard to compute or to understand, this measure of consensus is one of the most common tests that is often used. However, the percentage can produce the best results if used on binary response and has more application to small group consensus [23],[49].

Spillman, Bezdek, and Spillman [56] believe the measures discussed above seem incompatible with small groups as an ongoing dynamic process. The dynamic structure of small group variables must be studied if these groups are to be considered from a general systems perspective. Therefore, a couple of years after Hill's publications, a new way of measuring the consensus was introduced. The method depends on the application of fuzzy mathematics in the structure and analysis of singular and group preference matrices. The characteristics of their work in comparison to all other studies done before applies at different times during group discussion rather than a single alternative at the conclusion of the discussion [41], [56], [40].

Many researchers use statistical concepts in measuring consensus. The variance that is determined from the Likert data and is also one of the most common tools for measuring consensus or, to be more precise, the deficiency of consensus. There are advantages and disadvantages for using the variance as a measure of consensus. On the one hand, it is significant in an exact comparison status. But, on the contrary, it could be meaningless when it is used to compare groups that have different sizes or have differences in mean value [17].

One of these methods that depend on the variance only for measuring the consensus is called r_{WG} proposed in [35] by James et al. The r_{WG} index is computed by comparing the observed within-group variance with an expected variation from random responses. In other words, you can calculate r_{WG} index by dividing the variance by an estimate of the amount of variance that would be expected by chance alone, and then subtract this value from one. Comparatively, this measure is not difficult to determine and can almost adapt to a scale from 0 to 1. r_{WG} can be used for across time com-

parisons with the consideration that this measure is affected by sample size. Moreover, the major difficulty with the measure is that it cannot be used as a comparison across different studies [17].

If we have enough knowledge in statistics, we can apply the entropy equations. In general, entropy in statistical thermodynamics as a measure of the number of microscopic configurations that a thermodynamic system can have when in a state specified by certain macroscopic variables[25]. You don't get it, right? Me either! Let us focus on the statistical definition then. Entropy, in statistics, means a logarithmic measure of the number of cases with a remarkable probability of being taken. Also, defined as the measure of disorder of a system, the higher the disorder, the higher the entropy. In general, Entropy is provided by:

$$S = k_B \sum_i p_i \log p_i$$

where k_B is the Boltzmann constant and p_i is the probability that the system is in the i^{th} state[22].

The reason for the above description of entropy is that Shannon [54] introduced a widely used formula for this $-\sum p_i \log p_i$ where p_i refers the probability that the i^{th} event takes places. This formula is then used by Tastle and Wierman [59] to present the measure of consensus $Cns(X)$. This measure of consensus, $Cns(X)$, was developed after they established first an essential set of rules that they believed should be satisfied before any measure can be considered a viable solution to the Likert scale consensus problem [58]. Likert scales in Tastle and Wierman measure are considered to be ordinal and then it uses the probability distribution and the distance between categories to produce a value between 0 and 1. $Cns(X)$ defined as:

$$Cns(X) = 1 - \sum_{i=1}^n p_i \log_2 \left(1 - \frac{X_i - \mu_X}{d_X} \right)$$

where p_i is the probability of outcome X_i , μ_X is the mean of X and $d_X =$

$$X_{max} - X_{min}.$$

Not only Tastle and Wierman utilize Shannon's formula to measure consensus, but many other scholars employed entropy to represent the degree of consensus. Researchers use this formula as a method for selecting the best classifier in a decision tree [50], to describe status in a social science system [9], or as an index to measure consensus for economic theory and policy among economists [8].

More recently, work presented by Beliakov et al. [10] focuses on the problem of constructing functions that can measure the degree of consensus for a set of inputs provided over the unit interval. They introduced two consensus operators built from aggregation functions and inclusion factors. Zhang et al. [61] and Alcalde-Unzu and Vorsatz [4] recently proposed rules that, as they believe, will improve the consensus making process.

There are many other studies for measuring the consensus in numerous ways. From an approach based on preference relation modeling that examines consensus in the context of the group decision making process [31],[61], [47], to geometric cardinal consensus index to measure consensus degree among decision makers [20]. Moreover, a direct consensus framework that provides individual preference vectors of alternatives is another approach presented by Dong and Zhang [21].

This work produces a new index to compare the measure of consensus of a group using data composed of Likert items. Likert items are usually employed in many fields to measure situation, priority, and personal reactions. This thesis is structured as follows: Firstly, the introduction of a new index that exploits the conditional distribution of the variance for a given mean that depends on the comparison of data collected using Likert items. A common 5-point Likert scale is used in the second chapter, which is the first paper. Chapter 3, the second paper, shows that the same values of disagreement can be computed, and then the consensus, by using simple geometric and numerical steps rather than intense calculus with many cases. Chapter 3, which is

divided into several subsections, provides two algorithms along with all the necessary facts from the areas of computational geometry, conditional probability, and consensus. A new multidimensional process for generalizing the consensus measure is presented in the Chapter 4, the third paper. Chapter 4 focuses on the generalization, with consideration of the conditional distribution of the variance for a given mean, using multidimensional computational methods to produce comparable results to methods used in Chapters 2 and 3. Chapter 5, the last chapter, provides a brief discussion and various open problems.

Chapter 2

Paper 1: A method for measuring consensus within groups: An index of disagreement via conditional probability

This paper published in February 2016 in the Information Sciences journal in ELSEVIER Inc. See references for the full citation.

Abstract

This paper presents a new index of disagreement (or measure of consensus) for comparison of data collected using Likert items. This new index, which assesses the level of disagreement among group members, exploits the conditional distribution of the variance for a given mean. The variance is often used as a measure of disagreement, with high variance seen as a high disagreement in a group. However, since the range of the variance is a function of the mean, this implies that for a mean close to the end points of the scale, the range of the variance is relatively small and for a mean at the center of the scale the range of the variance is larger. The index of disagreement introduced in this paper takes into account both the mean and the variance and provides a way to compare two groups that is more meaningful than just considering the variance or other measures of disagreement or consensus that only depend on the variance.

2.1 Introduction

In this article, consensus refers to a belief, opinion, or perception reached by a group of persons, specifically the extent to which the group members can agree on Likert scale items. Disagreement refers to a difference of belief, opinion, or perception; it is the inverse or opposite of consensus. The new measure introduced focuses specifically on consensus measured using Likert scale item responses. The idea of measuring consensus has been appearing in many forms in recent literature.

One of the easiest and most common measures of consensus is the percentage agreement measure. This measure, which is used to describe the percentage of group members or estimate the percentage of a population who endorse a particular belief, is easy to compute and to understand. The measure works best for binary responses, has been used in various situations and has been applied to small group consensus [23],[49]. Another very common measure of consensus, or lack of consensus, is variance. There have been many discussions since Stevens [57] 1946 treatise on types of scales and permissible statistics about whether Likert data are ordinal or interval. In the past, Likert data have been handled in both ways [24] and thus the mean and variance of this type of data are computed and have been used in meaningful ways. Norman [44] gives a compelling argument for the use of parametric methods with Likert data. Thus, variance, computed from Likert data, is often used to discuss the disagreement or lack of consensus in a group. There are some disadvantages to using this statistic as a measure of disagreement; variance is only meaningful in an exact comparison context and cannot be meaningfully used to compare groups when the mean or the group size are different [17].

A more refined measure of consensus that can be applied to Likert data is the within-group agreement index (r_{WG}). This index is calculated by dividing the variance by an estimate of the amount of variance that would be expected by chance alone, and then subtracting this value from one [35].

This measure is relatively easy to compute and roughly conforms to a scale from 0 to 1, although values sometimes fall outside this range. While this measure is affected by sample size, it can be used for across time comparisons and r_{WG} does control for chance. However, this measure cannot be used to compare across studies [17].

Another measure of consensus from information theory is Entropy or the measure of the degree of disorder of a system. Shannon [54] proposed the most widely used formula for this $-\sum p_i \log p_i$ where p_i represents the probability that the i^{th} event occurs. Others employed entropy to indicate the degree of consensus as a method to select the best classifier in a decision tree [50], to describe status in a social science system [9], or as an index to measure consensus for economic theory and policy among economists [8].

More recently, Tastle and Weirman [[58],[59]] introduce a measure of consensus that build on Shannon's entropy and also applies to Likert scale responses; however, unlike some of the previously mentioned measures, these authors consider Likert data to be ordinal. Their measure uses the probability distribution and the distance between categories to produce a value between 0 and 1. They define consensus, $Cns(X)$, as follows:

$$Cns(X) = 1 - \sum_{i=1}^n p_i \log_2 \left(1 - \frac{X_i - \mu_X}{d_X}\right)$$

where p_i is the probability of outcome X_i , μ_X is the mean of X and $d_X = X_{max} - X_{min}$. The authors incorporate aspects of the entropy measure in the calculation and although they consider the data as ordinal, they do employ the use of the mean.

In an earlier paper, Tastle and Weirman [58] also establish an important set of rules that they believe must be satisfied before any measure can be considered a viable solution to the Likert scale consensus problem. The four rules follow:

1. For a given (even) number of individuals participating in

a discussion on some question of interest, if an equal number of individuals, $n/2$, separate themselves into two disjoint groups, each centered on the strongly disagree and strongly agree categories, the group is considered to have no consensus.

2. If all participants classify themselves in the same category of the Likert scale, be it to agree or disagree on the question or matter at hand, or if all are neutral on the matter, then the consensus of the group is considered to be complete at 100%.
3. If the mix of participants is such that $n/2 + 1$ participants assign themselves to any one category, the degree of consensus must be greater than 0, for the balance in the group is no longer equal.
4. As the number of categories to which participants classifies himself/herself diminishes, the consensus must increase, eventually approaching 1 on the unit interval. Thus, when all participants place themselves in a single category, consensus has been maximized and it considered to be perfect, and that is given a value of 1. (p. 387)

There is also another approach to consensus based on preference relation modeling that examines consensus in the context of the group decision making process [31]-[61]. This approach is applied to multi-criteria decision-making problems where a finite set of alternatives must be compared and/or ordered and each alternative has multiple criteria to consider [47]. The approach stresses the importance of coming to an acceptable level of agreement among a group of experts, for example doctors deciding on a course of patient treatment, and considers reaching consensus as a systematic course of action that is implemented in a group under the supervision of a moderator with

the intention of reducing discordance. The guiding idea in this approach is to gather relevant information from experts in each round of discussion [48]. In some models, this method includes the use of an index of comparability and an index of concordance to regulate information flow and to invite discordant and confident experts to explain opinions to the group in order to facilitate consensus building [47].

Following this same approach of multiperson decision making with preferences, Dong et al. [20] define the geometric cardinal consensus index and the geometric ordinal consensus index to measure consensus degree among decision makers. More recently, Dong and Zhang [21] have developed a direct consensus framework that provides individual preference vectors of alternatives. Standardized individual preference vectors are then aggregated into the collective preference vector. These help adjust the preference representation structures that are presented to help decision makers reach consensus. Zhang, Dong and Xu [61] also recently introduced new rules that they believe will improve the consensus making process.

There are various ways to define consensus and numerous approaches to measure consensus within and among groups. The research presented in this article creates a new index to compare the disagreement (or measure the consensus) of a group using data collected by Likert items. Likert items are commonly used in many disciplines to measure attitudes, preferences, and subjective reactions. For this discussion a common 5-point Likert scale is used, with the integers 1 through 5 corresponding to the words strongly disagree through strongly agree.

This new index is based on the fact that when using a Likert scale, treating the data as interval, and calculating the mean and variance, the range of the variance is always a function of the mean. For example on a five-point scale, if the mean is closer to either of the end points (one or five) the variance must lie within a smaller range of values than if the mean is closer to the midpoint of three. It is the strong interdependence between the mean and

variance for a particular Likert scale that forms the basis for this work. Thus, any measurement of disagreement or consensus that involves using the variance is subject to the mean. To overcome this problem, the authors have developed the new index of disagreement (and its inverse the measure of consensus) to consider the mean in the calculation and also the range of values for the variance at that mean. Observations about this relationship between the mean and the variance of the data expressed for a particular Likert scale and the implications of this relationship will be developed in this article. The newly defined index of disagreement and measure of consensus are comparable across time, groups and studies.

The paper first outlines the mathematical foundations and then give examples and comparisons with other measures of consensus. Section 2, which is divided into several subsections, gives the Theoretical Foundations of the new index of disagreement (Φ) and measure of consensus (Ψ). The subsections of Section 2 discuss the bounds for the variance, the development of Φ , and then a discussion of how to compute both Φ and Ψ for a five-point scale. Section 3 compares Ψ to two other measures, r_{wg} [35] and $Cns(X)$ [59], showing when they agree and when the new measure will provide additional or different information. Section 4 discusses conclusions and future work.

2.2 Theoretical Foundations

Consider a survey where in the respondents are asked to choose exactly one answer from the scales: $1, 2, \dots, n$. It is assumed that the scales are defined to allow meaningful computations of the mean and variance. This means, for example, that the scale does not represent a categorical variable. Although there is no theoretical restriction on the number n ($n > 1$) of admissible responses, a number n that is too large would make the survey meaningless. On the other hand, a too small number n would make the situation trivial. For example, $n = 2$ yields binomial responses wherein the mean value

uniquely determines the variance. Since the object of this discussion is to address the range of variance (as an indicator of consensus) at a given mean, $n = 2$ leaves nothing to discuss. Therefore, we will place a restriction that $n > 2$. In this paper we will focus on $n = 5$, because (a) it is one of the most prevalent modes of questionnaire design and (b) this choice is within mathematical maneuverability. It is noted, however, that while a similar result can be developed for any number, algebra involved will lose its simplicity for a larger number $n > 5$.

Let m and v denote the mean and the variance computed from the survey defined above. A large value of v suggests a high level of disagreement (equivalently a low level of consensus). For example, a vanishing variance ($v = 0$) represents a complete consensus. Therefore, a survey question that yielded $v = 0$ indicates a higher consensus than a survey question that yielded, say $v = 2$. The latter shows the existence of different responses. However, variances arising from different survey questions or from the same survey question for different groups do not always provide a simplistic basis of comparisons of consensus. This is because the variance ranges in an interval determined by the mean. It will be seen below that, for a given mean m , variance v varies between the minimum and the maximum that are determined by m . This observation leads us to the following question.

Question: Suppose the result of two questionnaires showed means m_i and variances v_i , where $i = 1$ or 2 . In notation, we write $(v_1; m_1)$ and $(v_2; m_2)$. We assume that $v_1 > v_2$. Does the second survey question (that had a smaller variance v_2) indicate a lower disagreement? Or conversely, does the first survey question (that had a larger variance v_1) represent a higher disagreement?

It is the thesis of this paper that the above conclusion does not automatically follow and the variance alone is not a total indicator of disagreement (or consensus). For example, at the mean m_2 the variance v_2 (the smaller of the

two) may represent the maximum value the variance may take on, while at m_1 the variance v_1 (the larger of the two) may be close to the minimum value that the variance may take on. It will be seen below that this is considered as a problem of conditional distribution of variance at a given mean value.

Since variance is used as a measure of disagreement, a low variance suggests a low level of disagreement and a high variance suggests a high disagreement, we will develop an index of disagreement, denoted by $\Phi(v; m)$. Consensus will be viewed as the complement of disagreement, $\Psi(v; m) = 1 - \Phi(v; m)$, which might be termed as the index of consensus. Since consensus and disagreement are complementary to each other, this reversing does not add any information.

The new index of disagreement is a conditional distribution of the variance v for a given mean m . For any given mean m , it has the following properties:

- $0 \leq \Phi(v; m) \leq 1$,
- $\Phi(v; m) = 1$ indicates a maximum disagreement, and
- $\Phi(v; m) = 0$ indicates a minimum disagreement.

The new index of disagreement is based not only on variance, but considers the range of the variance as a function of the mean. Depending on the values of m_1 and m_2 , it is possible that a smaller variance represents a higher disagreement, i.e., $\Phi(v_1; m_1) > \Phi(v_2; m_2)$ is a possibility even when $v_1 < v_2$. It will be seen that $\Phi(v; m)$ and v jointly provide a more reasonable measure of disagreement.

2.2.1 Bounds for the Variance

As mentioned before we consider a survey where a response is made by selecting exactly one value out of the n numerical scales: $1, 2, \dots, n$. Let $\{p_i\}$ denote the distribution of responses so that p_i represents the proportion of responses for the scale value i .

We have $\sum_{i=1}^n p_i = 1$. Thus, we have a variable X whose distribution is $Prob(X = i) = p_i$, where $i = 1, 2, \dots, n$. For simplicity, the number of survey respondents is assumed sufficiently large to treat p_i as a continuous variable.

The mean and the variance of X are given by:

$$m = m(X) = \sum_{i=1}^n ip_i, \text{ and}$$

$$v = v(X) = \sum_{i=1}^n (i - m)^2 p_i.$$

Therefore, we have the system of three equations

$$\begin{aligned} \sum_{i=1}^n p_i &= 1, \\ m &= \sum_{i=1}^n ip_i, \text{ and} \\ v &= \sum_{i=1}^n (i - m)^2 p_i, \end{aligned} \tag{2.1}$$

with the following restrictions:

$$0 \leq p_i \leq 1, \quad i = 1, 2, \dots, n. \tag{2.2}$$

Statement 1: For a given mean value m , $1 \leq m \leq n$, v is bounded by the following inequalities:

$$v \leq (m-1)(n-m) \leq \left(\frac{n-1}{2}\right)^2 \quad (2.3)$$

and

$$v \geq (m - [m])([m] + 1 - m), \quad (2.4)$$

where $[m]$ is the greatest integer $\leq m$.

Note: In (3), we have $(m-1)(n-m) = \left(\frac{n-1}{2}\right)^2$ if and only if $m = \frac{n+1}{2}$.

proof:

The proof for (3) is provided first. Let $\{p'_i\}$ be given by

$$\begin{aligned} p'_1 &= \frac{n-m}{n-1}, \\ p'_n &= \frac{m-1}{n-1}, \text{ and} \\ p'_i &= 0 \text{ for } 1 < i < n. \end{aligned}$$

Then, we have

$$\begin{aligned} \sum_{i=1}^n p'_i &= 1, \quad m' = \sum_{i=1}^n ip'_i = m, \quad \text{and} \\ v' &= \sum_{i=1}^n (i-m)^2 p'_i = (m-1)(n-m). \end{aligned}$$

Therefore, the upper bound stated in (3) is realized by $\{p'_i\}$. It has to be shown that $v \geq v'$ for any variance v arising from any $\{p_i\}$. Since it can be

verified that $(i - m)^2(n - 1) \leq (m - 1)^2(n - i) + (n - m)^2(i - 1)$, we have

$$\begin{aligned}
v &= \sum_{i=1}^n (i - m)^2 p_i \\
&\leq \frac{(m - 1)^2}{n - 1} \sum_{i=1}^n (n - i) p_i + \frac{(n - m)^2}{n - 1} \sum_{i=1}^n (i - 1) p_i \\
&= \frac{(m - 1)^2}{n - 1} (n - m) + \frac{(n - m)^2}{n - 1} (m - 1) \\
&= (m - 1)(n - m) = v'.
\end{aligned}$$

The proof for (4) is provided next. Consider $\{\hat{p}_i\}$ where

$$\begin{aligned}
\hat{p}_{[m]} &= [m] + 1 - m, \\
\hat{p}_{[m]+1} &= m - [m], \text{ and} \\
\hat{p}_i &= 0, \text{ for all other } i.
\end{aligned}$$

Then

$$\begin{aligned}
\sum_{i=1}^n \hat{p}_i &= 1, \quad \sum_{i=1}^n i \hat{p}_i = m, \quad \text{and} \\
\hat{v} &= \sum_{i=1}^n (i - m)^2 \hat{p}_i \\
&= ([m] - m)^2([m] + 1 - m) + ([m] + 1 - m)^2(m - [m]) \\
&= (m - [m])([m] + 1 - m).
\end{aligned}$$

Therefore, the lower bound stated in (4) is realized by the distribution $\{\hat{p}_i\}$. It remains to show that $v \geq \hat{v}$ for any variance v arising from any distribution $\{p_i\}$.

Using the inequality, $(i - m)^2 \geq (m - [m])^2([m] + 1 - i) + ([m] + 1 -$

$m)^2(i - [m])$ for any i , we see that

$$\begin{aligned}
v &= \sum_{i=1}^n (i - m)^2 p_i \\
&\geq \sum_{i=1}^n \{(m - [m])^2([m] + 1 - i) + ([m] + 1 - m)^2(i - [m])\} p_i \\
&= (m - [m])^2 \sum_{i=1}^n ([m] + 1 - i) p_i + ([m] + 1 - m)^2 \sum_{i=1}^n (i - [m]) p_i \\
&= (m - [m])^2([m] + 1 - m) + ([m] + 1 - m)^2(m - [m]) \\
&= (m - [m])([m] + 1 - m) \\
&= \hat{v},
\end{aligned}$$

which completes the proof of Statement 1.

Define $Max(v|m) = (m - 1)(n - m)$ and $Min(v|m) = (m - [m])([m] + 1 - m)$. The range of the variance, $R(m) = [Min(v|m), Max(v|m)]$, therefore is a function of a given mean. When $n = 2$, it can easily be shown that $Max(v|m) = Min(v|m)$. Inequalities (3) and (4) suggest that, without reference to m , a given variance cannot be consider either "large" or "small." Depending on the value of m) a given value of variance v may be in any part of the range $R(m)$.

The range of the variance is sketched below for $n = 5$. The horizontal axis represents the mean and the vertical axis denotes the variance. The upper parabola represents the $Max(v|m) = (m - 1)(n - m)$, while the undulating graph below represents the $Min(v|m) = (m - [m])([m] + 1 - m)$. It is noted that the graph is symmetrical with respect to the vertical line $m = 3$. The graph below in Figure 1 depicts how a given mean dictates a range for the variance. The range of the variance at any mean will lie between the parabola and the undulating graph.

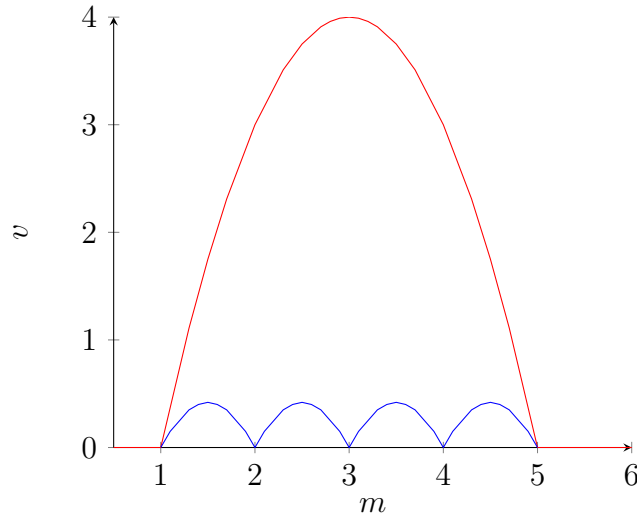


Figure 2.1: The Range of the Variance

For selected values of the mean between 1.0 and 3.0, Table 2.1 numerically illustrates the minimum, maximum, and the width of a range. For example, the table shows that a wide range (0.0 to 4.0) is allowed for the variance if the mean is $m = 3$, while variances corresponding to the mean $m = 1$ are afforded with a zero range of variation. The table only presents mean values from 1.0 to 3.0, since the mean values from 3.0 to 5.0 will be symmetric with these values in the table. If $3 \leq m \leq 5$ then, take $6 - m$ to find the corresponding value and range for the variance in Table 1. For example, if $m = 4.6$ then look at $m = 6 - 4.6 = 1.4$. Using the table value of 1.4, we see that the range of variance for 4.6 will also be from 0.24 to 1.44. Both values 4.6 and 1.4 are 1.6 away from the middle of the scale $i = 3$. Therefore, again, it is worth reiterating the idea that two variances associated to different means are not directly comparable because the range for the variance is dependent on the mean.

Table 2.1: The Range of Variance for a Given Mean

Mean	Variance		
	Minimum	Maximum	Width of Range
1.0	0.00	0.00	0.0
1.1	0.09	0.39	0.3
1.2	0.16	0.76	0.6
1.3	0.21	1.11	0.9
1.4	0.24	1.44	1.2
1.5	0.25	1.75	1.5
1.6	0.24	2.04	1.8
1.7	0.21	2.31	2.1
1.8	0.16	2.56	2.4
1.9	0.09	2.79	2.7
2.0	0.00	3.00	3.0
2.1	0.09	3.19	3.1
2.2	0.16	3.36	3.2
2.3	0.21	3.51	3.3
2.4	0.24	3.64	3.4
2.5	0.25	3.75	3.5
2.6	0.24	3.84	3.6
2.7	0.21	3.91	3.7
2.8	0.16	3.96	3.8
2.9	0.16	3.99	3.9
3.0	0.00	4.00	4.0

In summary of the above observations, we state the following:

- If two survey questions show the same mean, respective variances indicate the hierarchy of disagreement. A larger (or smaller) variance shows a higher (or lower) disagreement in such a case.
- A more complex scenario arises when the two survey questions receive different means. Then, as mentioned above, variance alone is not a valid indicator in comparing the levels of disagreement. Therefore, the purpose of this paper is to present a measurement that will be comparable across means.

2.2.2 Defining the Index of Disagreement

This section goes through the steps necessary to show that $\Phi(v|m)$ is a conditional probability density function and give examples of how this measurement along with variance can be used to measure group disagreement for a Likert item.

It can be verified that the equations in (1) are equivalent to the following:

$$\begin{aligned}
 p_1 &= \frac{1}{2}(v + m^2 - 5m + 6) - \frac{1}{2} \sum_{i=4}^n (i^2 - 5i + 6)p_i \\
 p_2 &= -(v + m^2 - 4m + 3) + \sum_{i=4}^n (i^2 - 4i + 3)p_i \\
 p_3 &= \frac{1}{2}(v + m^2 - 3m + 2) - \frac{1}{2} \sum_{i=4}^n (i^2 - 3i + 2)p_i.
 \end{aligned} \tag{2.5}$$

Statement 2: Let $t = \frac{1}{2}(v + m^2 - 3m + 2)$, Then

$$Max(t) = \frac{n-2}{2}(m-1), \text{ and} \quad (2.6)$$

$$Min(t) = \frac{1}{2}([m]-1)(2m-[m]-2). \quad (2.7)$$

From now on, it is assumed that $n = 5$. We also assume that $1 \leq m \leq 3$. It will be seen that this restriction does not affect generality due to the aforementioned symmetry of the data set with the range of the variance being the same for m and $6 - m$.

To simplify notation and avoid the need for subscripting, we let $(a, b, c, d, e) = (p_1, p_2, p_3, p_4, p_5)$, so we can rewrite (1) and (2) as below:

$$\begin{aligned} a + b + c + d + e &= 1 \\ a + 2b + 3c + 4d + 5e &= m \\ a + 4b + 9c + 16d + 25e &= v + m^2 \end{aligned} \quad (2.8)$$

with

$$0 \leq a, b, c, d, e \leq 1,$$

where m is the mean and v is the variance. The following shows that the maximum and minimum for the variance holds for a given mean m :

$$\begin{aligned} Max(v|m) &= (m-1)(5-m) \\ Min(v|m) &= (m-[m])([m]+1-m) \end{aligned} \quad (2.9)$$

Let $t = \frac{1}{2}(v + m^2 - 3m + 2)$. $u = \frac{m-1}{2}$, and $w = Max\{m-2, 0\}$. Then, for

$n = 5$, (Statement 2) gives:

$$Max(t) = \frac{3}{2}(m - 1) = 3u. \quad (2.10)$$

$$\begin{aligned} Min(t) &= 0 && \text{for } 1 \leq m \leq 2 \\ &= m - 2 && \text{for } 2 \leq m \leq 3; \end{aligned} \quad (2.11)$$

i.e. $Min(t) = w$.

Statement 3: The equations in (8) are equivalent to the following

$$\begin{aligned} a &= \frac{1}{2}(v + m^2 - 5m + 6) - (d + 3e) \\ b &= -(v + m^2 - 4m + 3) + (3d + 8e) \\ c &= \frac{1}{2}(v + m^2 - 3m + 2) - (3d + 6e). \end{aligned} \quad (2.12)$$

where

$$\begin{aligned} d + 3e &\leq t - (m - 2) \\ 3d + 8e &\geq 2t - (m - 1) \\ 3d + 6e &\leq t \\ 0 &\leq d, e \leq 1. \end{aligned}$$

Let

$$\Omega_1 = \{(c, d) \mid 3d + 8e \geq 2t - (m - 1) \quad \text{and } 0 \leq d, e \leq 1\}$$

$$\Omega_2 = \{(d, e) \mid d + 3e \geq t - (m - 2) \quad \text{and } 0 \leq d, e \leq 1\}$$

$$\Omega = (V) = \{(d, e) \mid 3d + 6e \leq t \quad \text{and } 0 \leq d, e \leq 1\}$$

and

$$A_1(v) = \text{Area of } [\Omega_1 \cap \Omega]$$

$$A_2(v) = \text{Area of } [\Omega_2 \cap \Omega]$$

$$A(v) = A_1(v) - A_2(v).$$

Thus, for a given m , we have

$$A(v) = \text{Area of}\{(d, e) \mid (d, e) \text{ satisfies (12)}\}.$$

The ratio $\frac{A(v)}{\int A(v)dv}$ represents the conditional probability density function of $v|m$.

Statement 4: For $1 \leq m \leq 3$, we have

$$\begin{aligned} g_1(t) &= \frac{A_1(v)}{36} = \frac{\tilde{A}_1(t)}{36} = t^2, & w \leq t \leq u & \quad (2.13) \\ &= t^2 - 3(t - u)^2, & u \leq t \leq 2u, & \\ &= (t - 3u)^2, & 2u \leq t \leq 3u. & \end{aligned}$$

$$\text{If } 1 \leq m \leq 2, \text{ then } g_2(t) = \frac{A_2(v)}{36} = \frac{\tilde{A}_2(t)}{36} = 0. \quad (2.14)$$

$$\begin{aligned} \text{If } 2 \leq m \leq 3, \text{ then} & \quad (2.15) \\ g_2(t) &= \frac{A_2(v)}{36} = \frac{\tilde{A}_2(t)}{36} = t^2 - 6(t - w)^2, & w \leq t \leq \frac{3}{2}w \\ &= 3(t - 2w)^2, & \frac{3}{2}w \leq t \leq 2w, \\ &= 0, & 2w \leq t \leq 3w. \end{aligned}$$

Statement 5: Let $g(t) = g_1(t) - g_2(t)$. Then,

$$\int_w^{3u} g(t)dt = 2u^3 - w^3. \quad (2.16)$$

Therefore, $\frac{g(t)}{2u^3 - w^3}$ is the conditional probability density function of $t = t(v)$ for a given m .

Definition: The index of disagreement $\Phi(v|m)$ is defined as below:

(i) If $m = 1$, then we define $\Phi(v|m) = 0$.

(ii) If $1 < m \leq 3$, then we define

$$\Phi(v|m) = \frac{\int_w^\tau g(t)dt}{2u^3 - w^3} \text{ where } \tau = \frac{1}{2}(V + m^2 - 3m + 2).$$

(iii) If $3 \leq m \leq 5$, then we define

$$\Phi(v|m) = \Phi(v|6 - m).$$

In addition, the following two functions are defined to be used later.

$$\begin{aligned} G_1(\tau) &= \int_w^\tau g_1(t)dt + \frac{1}{3}w^3 = \frac{1}{3}\tau^3, & w \leq \tau \leq u \\ &= \frac{1}{3}\tau^3 - (\tau - u)^3, & u \leq \tau \leq 2u, \\ &= 2u^3 + \frac{1}{3}(\tau - 3u)^3, & 2u \leq \tau \leq 3u. \end{aligned} \quad (2.17)$$

and

$$\begin{aligned} G_2(\tau) &= \int_w^\tau g_2(t)dt + \frac{1}{3}w^3 = \frac{1}{3}\tau^3 - 2(\tau - w)^3, & w \leq \tau \leq \frac{3}{2}w \\ &= w^3 + (\tau - 2w)^3, & \frac{3}{2}w \leq \tau \leq 2w, \\ &= w^3, & 2w \leq \tau \leq 3u. \end{aligned} \quad (2.18)$$

It is noted that $G_1(w) - G_2(w) = 0$, $G_1(3u) - G_2(3u) = 2u^3 - w^3$, and $\int_w^\tau g(t)dt = G_1(\tau) - G_2(\tau)$.

Statement 6:

$$0 \leq \Phi(v|m) \leq 1. \quad (2.19)$$

$\Phi(V|m) = 1$ shows that (under the given m) V provides the maximum degree of disagreement, i.e., $v = (m - 1)(5 - m)$. (2.20)

$\Phi(v|m) = 0$ shows that (under the given m) v provides the minimum degree of disagreement, i.e., $v = (m - [m])([m] + 1 - m)$. (2.21)

$$\begin{aligned} \Phi(v|m) &= 0, & m &= 1. & (2.22) \\ &= \frac{G_1(\tau) - G_2(\tau)}{2u^3 - w^3}, & 1 &< m \leq 3 \\ &= \Phi(v|6 - m), & 3 &< m \leq 5. \end{aligned}$$

It should be noted that variance is an unconditional measure of disagreement; "unconditional" in the sense that two variances can be compared without regard to the underlying mean values. On the other hand, being a conditional probability distribution, $\Phi(v|m)$ is a *relative* measure of disagreement computed at a given mean value m . It is the thesis of this paper that disagreement should be measured using both v and $\Phi(v|m)$ as is demonstrated in the example cases to follow.

Case 1: When the magnitudes of v and $\Phi(v|m)$ agree.

More precisely, the following is the situation under consideration.

$$v_1 > v_2, \text{ and } \Phi(v_1|m_1) > \Phi(v_2|m_2).$$

Here, v_1 represents a higher level of dispersion than v_2 . Further, v_1 represents a higher level of disagreement (i.e., a higher cumulative value in terms of the conditional distribution) than v_2 at their respective locations, i.e., at m_1 and m_2 . Therefore, the disagreement is considered higher for the combination (v_1, m_1) than for (v_2, m_2) .

Example: Consider the situation where $(v_1, m_1) = (1.9, 4.2)$ and $(v_2, m_2) = (1.1, 4.6)$. $v_1 > v_2$ since v_1 is 1.9 and v_2 is 1.1. By computation we get $\Phi(v_1|m_1) = \Phi(1.9|4.2) = \Phi(1.9|1.8) = 0.91$ and $\Phi(v_2|m_2) = \Phi(1.1|4.6) = \Phi(1.1|1.4) = 0.78$; i.e., $\Phi(v_1|m_1) > \Phi(v_2|m_2)$. Therefore, v_1 represents a larger range of opinions than v_2 . One can initially suspect that the higher value of $v_1 = 1.9$ is due to its association with the lower mean $m_1 = 4.2$, that tends to allow more room for the range of opinions than the higher mean $m_2 = 4.6$ and therefore that the higher value $v_1 = 1.9$ is a indication of a higher level of disagreement. In this case the notion is maintained, as the inequality $\Phi(v_1|m_1) > \Phi(v_2|m_2)$ does support it.

Case 2: When the magnitudes of v and $\Phi(v|m)$ are different.

This implies the following situation:

$$v_1 > v_2, \text{ but } \Phi(v_1|m_1) < \Phi(v_2|m_2).$$

In this case, v_1 represents a higher level of dispersion in opinions than v_2 . However, the conditional distributions of the variance at m_1 and m_2 indicate that v_1 at m_1 indicates a lower percentage of the overall available range than v_2 at m_2 . Because the inequality $v_1 > v_2$ is a result of different mean values, v_1 does not necessarily indicate a higher level of disagreement than v_2 . In interpreting survey data, we should express this uncertainty. The example below serves to further illustrate this point.

Example: Slightly changing the above example, let us consider the case where $(v_1|m_1) = (1.9, 4.2)$ is the same as before, but we modify $(v_2|m_2)$ in this example to a new pair $(v_2|m_2) = (1.2, 4.6)$. It is noted that the value of v_2 is now changed from 1.1 to 1.2. The inequality $v_1 > v_2$ still holds since v_1 is 1.9 and v_2 is 1.2. However, in this case we have the following relationship:

- $\Phi(v_1|m_1) = \Phi(1.9|4.2) = \Phi(1.9|1.8) = 0.91$ and
- $\Phi(v_2|m_2) = \Phi(1.2|4.6) = \Phi(1.2|1.4) = 0.94$;

i.e., $\Phi(v_1|m_1) < \Phi(v_2|m_2)$. In this example we can see that even though $v_1 > v_2$ this does not necessarily indicate a higher level of disagreement for v_1 than v_2 , because a higher variance can be reached at $m_1 = 4.2$ than at $m_2 = 4.6$ (See Table 1). $v_1 = 1.9$ belongs to an upper 9 percent ($= 1 - 0.91$) of the admissible range at the location $m_1 = 4.2$, while $v_2 = 1.2$ at the location $m_2 = 4.6$ belongs to a more limited upper 6 percent ($= 1 - 0.94$) of the range. In interpreting the survey data, therefore, we should express the uncertainty arising from this type of a situation.

More examples will be given later in Section 4, when we compare the Φ measurement with variance and r_{WG} for climate survey data taken from a study of faculty members in various departments and colleges at a university.

2.2.3 Computing the Disagreement Index Φ and Consensus Measure Ψ

In order to compute Φ we follow the steps below:

1. If $m = 1$, then $\Phi(V|m) = 0$.
2. If $1 < m \leq 3$. For all V , $Min(V|m) \leq V \leq Max(V|m)$.

Compute:

$$u = \frac{m-1}{2}, w = Max\{m-1, 0\}, \tau = \frac{1}{2}(V+m^2-3m+2), \text{ and } D = 2u^3 - w^3.$$

At this juncture, it should be observed that

$$\begin{array}{lll}
 1 < m \leq 2 & \text{implies} & w = 0. \\
 2 < m \leq \frac{7}{3} & \text{implies} & w < \frac{3}{2}w \leq 2w < u < 2u < 3u. \\
 \frac{7}{3} < m \leq \frac{5}{2} & \text{implies} & w < \frac{3}{2}w \leq u < 2w < 2u < 3u. \\
 \frac{5}{2} < m \leq 3 & \text{implies} & w < u \leq \frac{3}{2}w < 2w < 2u < 3u.
 \end{array}$$

3. Compute $G_1(\tau)$ by (17).
4. Compute $G_2(\tau)$ by (18).
5. Compute $\Phi(v|m) = \frac{G_1(\tau) - G_2(\tau)}{D}$.
6. If $3 < m \leq 5$, compute $\Phi(v|m) = \Phi(v|6 - m)$.
7. Compute $\Psi = 1 - \Phi$.

Table (2.2) shows the Cumulative Conditional Distributions of Φ and Ψ for selected values of m and v .

2.3 Comparing Ψ with r_{WG} and $Cns(X)$

This section considers Likert data from a survey measuring organizational climate in an academic department. The examples below are not used to draw conclusions about the data, but merely to show the usefulness of Ψ compared to other measures of consensus. The post-survey data were analyzed to determine if there was any evidence that Ψ could provide more or different information about consensus of a group than variance, r_{WG} or $Cns(X)$. The different measures of consensus ($\Psi, r_{WG}, Cns(X)$) can agree or disagree and

Table 2.2: Sample Cumulative Conditional Distributions for Index of Disagreement $\Phi(v|m)$ and Measure of Consensus Ψ

Mean	Variance (v)	$\Phi(v m)$	Ψ
1.1	0.09	0.000000	1.000000
	0.1	0.000167	0.999833
	0.2	0.221333	0.778667
	0.3	0.878500	0.121500
	0.39	1.000000	0.000000
1.2	0.16	0.000000	1.000000
	0.2	0.001333	0.998667
	0.3	0.057167	0.942833
	0.4	0.284000	0.716000
	0.5	0.647333	0.352667
	0.6	0.914667	0.085333
	0.7	0.995500	0.004500
	0.76	1.000000	0.000000

below we consider various cases to compare the measures. Recall that r_{WG} is inversely related to the variance, so when the variance increases, r_{WG} decreases. $Cns(X)$ uses the probability distribution and the distance between categories to compute the consensus measure and the newly introduced Ψ is computed by considering both the mean and the range of variance at that specific mean.

Table (2.3) offers an example where the variance of Q38 is larger than for Q39. In this case all three measures of consensus agree; the smaller variance goes along with a greater consensus. Looking at this more closely, one notices that a mean score of 3.8696 is .8696 away from a mean of 3, where the range of the variance is largest, and the mean 2.2899 is only slightly closer (.7101) to the mean of 3. In this case when the variance for one question is smaller than the variance for another question, all measures of consensus go in the opposite direction and get larger. When the distance from the middle of the scale is similar for the mean values, then the range of the variance is also

Table 2.3: Comparison of $Cns(X)$ and r_{WG} with Ψ for Q38 and Q39

Question	Mean	Variance	r_{WG}	$Cns(X)$	Ψ
Q38	2.2899	0.8265	0.5867	0.6867	0.9068
Q39	3.8696	0.7621	0.6189	0.7267	0.9188

Table 2.4: Comparison of $Cns(X)$ and r_{WG} with Ψ for Q1 and Q2

Question	Mean	Variance	r_{WG}	$Cns(X)$	Ψ
Q1	3.7887	0.7976	0.6012	0.7342	0.9097
Q2	3.3803	0.8390	0.5805	0.6721	0.9351

very similar, so Ψ will agree with the other measures of consensus.

Next, we consider a case Table (2.4) where variance of Q1 is smaller than the variance of Q2. For this example, both r_{WG} and $Cns(X)$ for Q1 are larger than for Q2, while Ψ goes in the opposite direction. Since variance for Q1 is smaller than for Q2 it would seem that consensus should be larger for Q1 with the lower variance, but the reason that Ψ does not agree with this (or with r_{WG} and $Cns(X)$) is due to the different mean scores that provide different possible ranges for the variance. In this example, the mean score for Q2 (3.3803) is closer to 3, where the range of the variance is the largest, than the mean of Q1 at 3.7887 (almost 4). This implies that there is a greater range for variance for the mean of Q2. Noting that the variance of Q2 is larger than Q1 does not in itself indicate consensus is smaller as one might assume. As Ψ considers both the mean and the range of the variance, this nuance is captured; however, this is not captured by either r_{WG} or $Cns(X)$.

In the final example, we consider a case Table (2.5) where variance of Q23 is smaller than the variance of Q34. For this example, both r_{WG} and $Cns(X)$ for Q23 are larger than Q34, while Ψ goes in the opposite direction. The reason that Ψ does not agree with r_{WG} and $Cns(X)$ again is due to the different mean scores that provide different possible ranges for the variance.

Table 2.5: Comparison of $Cns(X)$ and r_{WG} with Ψ for Q23 and Q34

Question	Mean	Variance	r_{WG}	$Cns(X)$	Ψ
Q23	1.7429	0.9184	0.5408	0.6735	0.8435
Q34	3.5915	0.9308	0.5346	0.6628	0.8821

In this example, the mean score for Q23 (1.7429) is 1.2571 from the mean of 3, where the range of the variance is the largest, and the mean of Q34 (3.5915) is only .5915 away from 3. This implies that there is a greater range for variance for the mean of Q34. So noting that the variance of Q34 is larger than Q23 again does not indicate consensus is smaller as one might assume. As previously stated, Ψ considers both the mean and the range of the variance, and the mean of 3.5915 will have a much larger range for the variance, than the mean of 1.7429. So even though the variances look very similar, the one at mean of 1.7429 really implies a much higher variance and thus a lower consensus.

This section has given a case where Ψ agrees with r_{WG} and $Cns(X)$ and two cases where Ψ disagrees with the other two. The latter two examples show that Ψ can indeed provide additional information about the consensus of a groups' perceptions on Likert items.

2.4 Conclusions and Future Work

A new index of disagreement (or measure of consensus) has been developed for the comparison of data collected using Likert items. This paper shows that variance alone does not tell the entire story about the disagreement or the consensus of a group on a Likert item. This new index of disagreement exploits the conditional distribution of the variance for a given mean and provides more information than the mean and variance alone or even other measures that are based solely on the variance. The index introduced in this

paper takes into account both the mean and the variance and can be used to compare questions with different means or the same question for different groups.

Although the computation for this new measure may seem daunting at first, it can easily be computed using an Excel spreadsheet with just a few columns for calculations or developed as a function in SPSS or your favorite software package. The authors have developed a simple spreadsheet for computing Ψ and have plans to develop an interactive website for users to enter the mean and variance to easily compute this measure of consensus.

Future work includes proving that this index can be generalized to larger values of n and demonstrating with Likert data from several disciplines that Ψ can add to the interpretation of the consensus within groups, between groups, or the change in consensus of a group across time.

Table 2.6: Climate Survey Results.

#	SURVEY STATEMENT	Mean	Variance	Ψ
1	My opinions are respected at work.	3.7798	0.8399	0.8960
2	There is a feeling of unity and cohesion in my workplace.	3.3303	0.8529	0.9310
3	I plan to remain in academia for the rest of my career.	4.3099	0.4740	0.9911
4	I have heard denigrating remarks by faculty, staff, or students about minority religious groups.	1.6571	0.9005	0.8192
5	At WVU, people use inclusive language (e.g., he/she, chairperson).	3.5701	0.6436	0.9757
6	There is little collegiality among people at WVU.	2.4811	1.1092	0.8063
7	I feel safe in and around my workplace.	4.2130	0.9542	0.8366
8	I am treated by administration and colleagues with courtesy and respect.	3.9541	1.0627	0.8008
9	There is a feeling of positive morale.	3.4444	1.165109	0.7864
10	People who make complaints of sexual harassment are protected from harmful consequences.	3.5050	0.7627	0.9528
11	I have been asked intrusive questions about my personal life at work.	1.8952	0.8832	0.8857
12	Mentoring at WVU has been beneficial to my career.	3.1495	1.0340	0.8637
13	WVU is a friendly workplace.	3.7156	0.9091	0.8752
14	I have experienced discrimination in the promotion process because of my gender.	1.9811	0.8949	0.8812
15	I'd recommend WVU to a prospective faculty member of my gender.	3.8056	1.0553	0.8054
16	Administrative officials, faculty, or staff tell sexist jokes.	1.5701	0.4927	0.9863
17	I have heard denigrating remarks by faculty, staff, or students about race/ethnicity.	1.6698	0.7566	0.9149

Continuation of Table 2.6				
#	SURVEY STATEMENT	Mean	Variance	Ψ
18	Work is fairly distributed among colleagues in my workplace.	2.8224	1.2984	0.7337
19	I have heard denigrating remarks about persons of my gender.	1.7075	0.7232	0.9353
20	I am expected to work an unhealthy and unreasonable number of hours to succeed.	2.6944	1.3170	0.7160
21	Sexual harassment complaints are investigated adequately.	3.3402	0.7060	0.9678
22	I have the opportunity to collaborate with other faculty at WVU.	4.1743	0.4971	.9870
23	Colleagues make me uncomfortable by commenting on my physical appearance.	1.7358	0.8058	0.9044
24	I believe qualified men are given more career opportunities than qualified women.	2.4579	1.3826	0.6507
25	Women are adequately represented among the WVU faculty.	2.5189	1.1091	0.8115
26	When I speak out on issues of concern to me, I am labeled a troublemaker.	2.4857	1.2907	0.7070
27	My work schedule is sensitive to my social and family commitments.	3.4860	1.0446	0.8436
28	I am proud to be a member of this faculty.	3.9252	0.8434	0.8954
29	I have heard denigrating remarks by faculty, staff, or students about gay, lesbian, bisexual, or transgender persons.	1.8131	0.9837	0.8217
30	There is gender equity in salaries.	2.8738	1.3271	0.7189
31	Faculty are accepted and respected by their peers regardless of gender.	3.9174	0.7802	0.9153

Continuation of Table 2.6				
#	SURVEY STATEMENT	Mean	Variance	Ψ
32	Taking time off for family indicates to administration or colleagues a lack of commitment to my job.	2.5094	1.0523	0.8392
33	Administration respects my religious and cultural holidays when scheduling work commitments (e.g., clinical days, important meetings or events).	3.6916	0.9134	0.8759
34	My talents are recognized by administration or colleagues.	3.6389	1.0366	0.8268
35	Sexual harassment incidents occur at WVU.	3.2596	0.7766	0.9450
36	Appropriate accommodations are made for faculty with physical disabilities.	3.5294	0.7070	0.9653
37	Women's views are represented fairly on major committees.	3.5377	0.8986	0.9030
38	Women are interrupted at meetings more often than men.	2.3558	0.8528	0.9062
39	Contributions made by male and female faculty are equally valued.	3.8762	0.8945	0.8945
40	I have the same opportunities for informal networking as my peers.	3.7143	1.0330	0.8211
41	I have equal access to departmental resources (e.g., space and laboratory support staff) as my peers.	3.8426	1.0871	0.7885
End of Table				

Chapter 3

Paper 2: A Geometric Approach for Computing a Measure of Consensus for Groups

This paper published in October 2016 in the International Mathematical Forum in HIKARI Ltd. See references for the full citation.

Abstract

Akiyama et al. in [3] present a new important method to determine the index of disagreement, Φ , and measure consensus, Ψ , for a group of individuals responding to a Likert scale question. Their new measure exploits the conditional distribution of the variance for a given mean. The index allows for the comparison of consensus values of different questions for the same group or the same question for different groups, even though the questions may have different means. However, in [3] the complicated details make the new index very difficult to compute. This paper presents a simpler and more straight forward method to determine the Akiyama et al. measure of consensus, Ψ , by using computational geometry and numerical concepts. This geometric method is much easier to understand and computes the same values that Akiyama et al. get using their method. Moreover, this new algorithmic method is much easier to generalize to values of n larger than $n = 5$, since there are many studies that use Likert scales with more than five answer choices. The algorithms presented in this paper can easily be applied using any software package with just a few steps for calculations. The authors have developed a simple spreadsheet for computing Φ and Ψ that allows the users to enter the mean and variance to get the index of disagreement and measure of consensus values.

3.1 Introduction

The term “consensus” has been frequently used in literature in recent years in many different areas including, politics, science, justice, and social choice. K. Lehrer and C. Wagner in [42] have a good explanation why many different areas do not just use, but actually need the term “consensus.” They explain that various atomistic and individualistic theories have griped philosophical thought from time to time. However, some of these theories have not proven to be enough. Therefore, some factor of agreement (or consensus) is needed to give an explanation of some central visualization or way of thinking. For example, in politics, there is a tradition that seeks to justify government in term of consensus. Additionally, the term “consensus” is widely used in the area of group decision making (GDM) and it can have various meanings, such as, the full agreement of all the specialist concerning all the feasible alternatives or it can mean judgments arrived at by ‘most of’ those concerned [12] [16].

Since “consensus” is used in a variety of fields, we can find more than one formal definition. First, Cambridge dictionaries define it as “a generally accepted opinion or decision among a group of people” [19]. Second, it may be defined as “associated with the state of agreement in a group, or may impute to a position reached by a group of individuals acting as a whole.” [60]. Also, simply put, consensus is a priority [33], [59].

In this paper, we use the term consensus to mean an opinion, belief or understanding reached by a group of persons who can agree on Likert scale items. Disagreement refers to a difference of belief, opinion, or perception. Therefore, we can say basically that disagreement is the inverse or complement of consensus.

When looking at the recent literature in the field of consensus, we can easily notice that most of the ideas are focused on how we can find more consensus, or how we can build consensus among a group. Justice and Thomas in [36] did a good job of giving a step by step process in many different cases

to get the best agreement in a group. While S. Alonso et al. [7], try to use the occurrence of new electronic technologies to make the agreement among a large number of individuals easier and possible.

According to the literature above, in many instances it is important to find or build more consensus. How can we know how much consensus we have and how much we need? How can we measure consensus without rules or principles to calculate the percentage of consensus? Therefore, it is very important to find or build general rules and methods to measure the consensus and then we can judge if we need more consensus or not.

There are two simple measures that are commonly used to determine if there is consensus: the percentage agreement measure and the variance. The percentage agreement measure, which is also the easiest measure, works perfectly for binary responses and it has been used in a small group consensus [23] [49]. There are some drawbacks for using the variance as a measure of disagreement or consensus. It is less accurate or invalid if it is used to compare groups that have different sizes or different means [17].

Recently, more sophisticated methods for measuring consensus have been introduced. In the information theory field, there is an expression called “Entropy” or the measure of the degree of unrest of a system. Shannon’s [54] vastly used formula for Entropy is $\sum p_i \log p_i$ (where p_i is the probability of i^{th} event’s outcome). Shannon’s formula has been recently applied to Likert scale responses by Tastle and Weirman [58] [59]. Additionally, there are other approaches for measuring consensus like Dong and Zhang [21] who proposed a direct consensus framework for multi-person decision making with different preference representation structures. Further, one of the easier ways to determine consensus is interrater reliability, r_{WG} , that is represented in [35] by James et al.

In this article, we consider the Akiyama et al.[3] method for measuring consensus. These researchers have developed an index of disagreement, Φ , and measure of consensus, Ψ , that exploit the conditional distribution of the

variance for a given mean. This method has been compared to other measures of consensus, such as r_{WG} , and shown to give additional information [3]. Moreover, the index allows for the comparison of consensus values of different questions for the same group or the same question for different groups, even though the groups are different sizes and the questions may have different means. While the measure is valuable, the method to find Φ and Ψ in [3] is quite difficult to understand and hard to compute. The work outline in the article provides a straightforward algorithmic method to compute this very valuable measure.

This article will first discuss preliminary concepts of computational geometry and conditional probability. The outline of the mathematical foundations and discussion of the index in Akiyama et al. [3] will be in section 2. In section 3, we present new algorithms for finding the required area and computing the consensus values. Also in the same section, we present a comparison of the Akiyama et al. [3] calculations of the area by using calculus methods and our calculations using the algorithms to find the area. Finally, the conclusion and future work will be the last section.

3.2 Preliminaries

In this section, we present basic facts from the computational geometry, conditional probability and consensus.

3.2.1 Computational Geometry

Computational geometry is broadly construed as the study of algorithms for solving geometric problems on a computer. It stood out from the fields of algorithms design and analysis at the end of 1970s [18]. The early algorithmic solutions for numerous geometric problems were slow or difficult to understand and apply, but in recent years, many of the new algorithmic techniques have progressed and many of the previous approaches have been simplified.

The importance of computational geometry is a combination of two main factors: sound connections between classical mathematics and the theory of computer science on one side, and many ties with applications on the other [29] [53].

In subsequent sections, ideas from computational geometry will be used to simplify methods for finding areas previously found using integral calculus. One important method that will be employed for this work is the technique for finding the area of any convex polygon. There are several methods for doing this and all the methods are easily programmed [18] [45].

It is not difficult to find the area of any polygon by triangulating, and then do the summing of the triangular areas [18]. The way to triangulate any convex polygon is by choosing any vertex to be a common vertex with all diagonals incident to this common vertex. Notice that the common vertex can be any vertex serving as the center of the convex polygon. Therefore, the area of a polygon with vertices v_0, v_1, \dots, v_{n-1} categorized counterclockwise can be determined by:

$$A(P) = A(v_0, v_1, v_2) + A(v_0, v_2, v_3) + \dots + A(v_0, v_{n-2}, v_{n-1}) \quad (3.1)$$

where P is the polygon and v_0 is the center vertex.

Although the polygons in this work are convex polygons, the theorem below generalizes the equation (3.1) to convex and non-convex cases.

Theorem Let a polygon (convex or non-convex) P have vertices v_0, v_1, \dots, v_{n-1} labeled counterclockwise, and let p be any point in the plane. Then

$$A(P) = A(p, v_0, v_1) + A(p, v_1, v_2) + A(p, v_2, v_3) + \dots + A(p, v_{n-2}, v_{n-1}) + A(p, v_{n-1}, v_0) \quad (3.2)$$

If $v_i = (x_i, y_i)$, this expression is equivalent to the equations

$$2A(P) = \sum_{i=0}^{n-1} (x_i y_{i+1} - y_i x_{i+1}) \quad (3.3)$$

$$= \sum_{i=0}^{n-1} (x_i + x_{i+1})(y_{i+1} - y_i). \quad (3.4)$$

Proof The proof of this theorem is done by induction on the number of vertices n of P . For the entire proof see [45].

3.2.2 Conditional Probability

In statistics, there is no difference when determining the probability for one event or for more than one independent events, like the probability to pass the test of class A and to get a full grade for the homework of class B . However, there are many dependent events in life for which we many want to obtain the probability, such parking in a no-parking zone and getting a parking ticket or having a high level of education with a wide experience and getting a job. These kind of probabilities, when the probability of the second event is dependent on the probability of the first event, are called a conditional probabilities. The formal notation for the conditional probability is given by $P(B|A)$, which means the event B happens given that event A has already occurred [11]. To see the difference between the “regular” probability and the conditional probability, let’s take a look at the following simple example.

Example Assume we roll two fair six-sided dice, say A is the first die and B is the second, and we want to predict the outcome of $A = 3$. In other words, what is the probability that the first die will be a three?

Basically, $A = 3$ in exactly 6 of the 36 outcomes, and hence $P(A = 3) = 6/36 = 1/6 = 0.167$. What about the prediction of $A + B \leq 6$, we mean, what is the probability of $A + B \leq 6$? This sum would occur exactly 15 out of 36 possible roles. Then $P(A + B \leq 6) = 15/36 = 0.417$. Now what if we

want to find the probability of $A = 3$ given that $A + B \leq 6$, so the conditional probability $P(A = 3|A + B \leq 6) = 3/15$.

In previous work [3] it was shown the range of the variance is dependent on the mean. In other words if you are given a particular mean, m , then the variance can only take on certain values in a particular range. Thus, there is a conditional probability relationship between the mean and variance. Later in the paper we will continue to discuss the importance of this relationship in finding the measure of consensus.

3.3 Theoretical Foundations

Since this paper seeks to present a different easier method to find the measure of consensus first presented by Akiyama et al. [3], it has almost the same theoretical foundations. However, we need to look at the main idea for the index of disagreement and discuss the new easier and more generalizable method of finding areas than what is presented in [3].

Imagine there is a survey that asked respondents to choose precisely one answer from the numbers 1 to n such that the scale is defined to allow computation of the mean and variance. This means the scale does not represent a categorical variable, but rather an interval variable. There is no theoretical limitation in choosing the number n ($n > 1$) of allowable responses. However, an n that is too large would make the survey senseless. On the other hand, if n is too small, it would make the situation trivial. For instance, the number $n = 2$ leaves nothing for us to discuss, because our discussion is to address the range of variance (as an index of disagreement) at a given mean. Thus, we add the constraint that $n > 2$. In this paper, we will focus on $n = 5$, because this work depends on [3] and $n = 5$ is one of the most prevalent modes of questionnaire design.

In [3] it was established that for a given mean m , the variance v , ranges between a minimum and a maximum value that are determined by m . We

may believe that if the variance v , calculated from responses to a question from the survey defined above, for question one is higher than the variance for question two that this suggests there is a higher level of disagreement or lower level of consensus for question one than there is for question two. However, in general this is not necessarily the case because the variance ranges in an interval determined by the mean, variances arising from the two survey question for the same group that do not have the same mean cannot be reasonably compared and do not provide a viable way of comparing consensus on the two question. This also applies in case that the variance obtained from different groups for the same question. Thus, any method of measuring disagreement or consensus should take into consideration the fact that the range of the variance is a function of the mean.

In this paper, we remind the reader of Akiyama et al. method for finding an index of disagreement, denoted by $\Phi(v; m)$, that takes into account the relationship between the mean and range for the variance. The measure of consensus is defined as the complement of disagreement, $\Psi(v; m) = 1 - \Phi(v; m)$. The way the index of disagreement and measure of consensus is defined allows for the comparison among questions with different means or of groups across time.

3.3.1 Defining the Index of Disagreement

In order to present the easier method for computing the index of disagreement, we will present selected sections of details from [3] and then explain the steps for the new ways to compute these same values. Here we should also note that we are always dealing with probabilities in our discussions so all points p_i will first satisfy, $p_i \geq 0$ and $p_i \leq 1$.

One of the steps necessary to build the index of disagreement in [3] is finding the area $A(v) = A_1(v) - A_2(v)$ that satisfy the following equations

$$\begin{aligned}
p_1 &= \frac{1}{2}(v + m^2 - 5m + 6) - (x + 3y) \\
p_2 &= -(v + m^2 - 4m + 3) + (3x + 8y) \\
p_3 &= \frac{1}{2}(v + m^2 - 3m + 2) - (3x + 6y).
\end{aligned} \tag{3.5}$$

We let $t = \frac{1}{2}(v + m^2 - 3m + 2)$ and since each $p_i \geq 0, i = 1, 2, 3$, then the equations in (3.5) will be equivalent to the following

$$\begin{aligned}
x + 3y &\leq t - (m - 2) \\
3x + 8y &\geq 2t - (m - 1) \\
3x + 6y &\leq t \\
0 &\leq x, y \leq 1.
\end{aligned} \tag{3.6}$$

Thus, the area of $A_1(v)$ is defined to be points that satisfy the first and third inequalities in (3.6), or points that lie between the two lines. While $A_2(v)$ is defined to be points that satisfy the second and third inequalities in (3.6).

Determining this area is very important because the conditional probability density function of $v|m$ is represented by the ratio $\frac{A(v)}{\int A(v)dv}$.

Akiyama et al. in [3] then state three or four different equations to find the area of each $A_1(v)$ and $A_2(v)$. These equations are determined by the different values of t . For example, the area of $A_1(v)$ was determined by:

For $1 \leq m \leq 3$, we have

$$\begin{aligned}
A_1(v) &= \frac{t^2}{36}, & w \leq t \leq u \\
&= \frac{t^2 - 3(t - u)^2}{36}, & u \leq t \leq 2u, \\
&= \frac{(t - 3u)^2}{36}, & 2u \leq t \leq 3u.
\end{aligned} \tag{3.7}$$

where $u = 3(\frac{m-1}{2})$ and $w = \max\{m - 2, 0\}$

The way of finding the area of $A_2(v)$ is almost equally as complicated as

$A_1(v)$, see [3]. At this point we determined that using some simple computational geometry techniques could provide an easier and clearer method for finding the required area.

3.4 New Algorithms for Computing the Index of Disagreement

The first step toward computing the index of disagreement is finding the area of $A(v)$. To determine this area we now apply algorithmic approach for each $A_1(v)$ and $A_2(v)$ with the consideration that our area is bounded by lines $x = 0, y = 0, x = 1$ and $y = 1$.

Algorithm

1. Graph each linear equation from (3.6) in the (x, y) plane.
2. Determine the intersection points between any two lines or with x -axis and y -axis, say s_i .
3. Avoid any intersection point that lies outside the region $[0, 1] \times [0, 1]$. i.e., $x, y < 0$ or $x, y > 1$.
4. Add $\pm\epsilon$ to all intersection points which are inside the square interval of $(0, 1)$. ϵ is a small positive arbitrary constant.
5. Check $s_i \pm \epsilon$ and keep the points that satisfy all numbers of line equations and avoid others.
6. Once we have the intersection points of the wanted area, apply any computational geometry methods to find the area of this polygon.

The method for finding the area using computational geometry is explained in Section 2.1.

Now that we have determined an easy way to compute the area $A(v)$, let's take a look at the definition of the index of disagreement and then determine an easier way that it can be calculated.

The index of disagreement as defined by Akiyama et al. [3] is stated as following:

Definition:

The index of disagreement $\Phi(v|m)$ is defined as below:

(i) If $m = 1$, then we define $\Phi(v|m) = 0$.

(ii) If $1 < m \leq 3$, then we define

$$\Phi(v|m) = \frac{\int_w^\tau g(t)dt}{2u^3 - w^3} \text{ where } \tau = \frac{1}{2}(V + m^2 - 3m + 2).$$

(iii) If $3 \leq m \leq 5$, then we define

$$\Phi(v|m) = \Phi(v|6 - m).$$

where $g(t)$ here is represent $\frac{A(v)}{36}$.

Since, as we see above, the way of finding $g(t)$ (or $A(v)$) in [3] has different cases and depend on different conditions, then finding the integration of $g(t)$ will have different cases as well. For example, to find the integration of $g_1(t)$ (or $\frac{A(v)}{36}$) Akiyama et al. [3] defined the following function.

$$\begin{aligned} G_1(\tau) &= \int_w^\tau g_1(t)dt + \frac{1}{3}w^3 = \frac{1}{3}\tau^3, & w \leq \tau \leq u \\ &= \frac{1}{3}\tau^3 - (\tau - u)^3, & u \leq \tau \leq 2u, \\ &= 2u^3 + \frac{1}{3}(\tau - 3u)^3, & 2u \leq \tau \leq 3u. \end{aligned} \quad (3.8)$$

Since we found the area $A(v)$ by couple steps using the computational geometry techniques, we will not need many different cases to find the integration of $A(v)$. In fact, you can use any numerical integration methods to determine the integration. The denominator $2u^3 - w^3$ that stated in the definition above is basically the integration of $g(t)$ from minimum to maximum value of t . It can be determined by using the numerical integration methods.

The following algorithm used to find the disagreement and consensus values for a given mean and variance.

Algorithm

Input: Mean m and Variance v . **Output:** Disagreement and Consensus.

1. If $m > 3$ then $m = 6 - m$.
2. Set N . {N is any large number}.
3. Determined t, w and u
4. $h = \frac{(t-w)}{N}$
5. For $j : 1$ to $N - 1$
 $Count = w + j * h$
 If j is odd Then $S_1 = S_1 + D(Count)$ Else $S_2 = S_2 + D(Count)$
6. $Int = (h/3) * (D(w) + 4 * S_1 + 2 * S_2 + D(t))$ { D(t) is the area that calculated by last algorithm}
7. To find the Denom apply steps 4 – 6 with $t = 3u$
8. Disagreement= $\frac{Int}{Denom}$ and Consensus= $1 - Disagreement$.

When implementing this algorithm we used the Simpson Method to find the numerical integration and we can use any other methods by editing steps 5 and 6 only.

Table 3.1 shows the comparison of Akiyama et al. index of disagreement, Φ , and measure of consensus, Ψ , and our results for Φ_{CG} and Ψ_{CG} for selected values of m and v .

Table 3.1: Examples of Comparison of Akiyama et al. indices Φ and Ψ and our result Φ_{CG} and Ψ_{CG} for selected values of m and v

Mean	Variance (v)	Φ	Ψ	Φ_{CG}	Ψ_{CG}
1.1	0.09	0.0000	1.0000	0.0000	1.0000
	0.1	0.0001	0.9998	0.0001	0.9998
	0.2	0.2213	0.7786	0.2213	0.7786
	0.3	0.8785	0.1215	0.8785	0.1215
	0.39	1.0000	0.0000	1.0000	0.0000
1.2	0.16	0.0000	1.0000	0.0000	1.0000
	0.2	0.0013	0.9986	0.0013	0.9986
	0.3	0.0571	0.9428	0.0571	0.9428
	0.4	0.2840	0.7160	0.2840	0.7160
	0.5	0.6473	0.3526	0.6473	0.3526
	0.6	0.9146	0.0853	0.9146	0.0853
	0.7	0.9955	0.0045	0.9955	0.0045
	0.76	1.0000	0.0000	1.0000	0.0000

3.5 Conclusions and Future Work

The index of disagreement (or measure of consensus) in this paper is important because it takes into account the relationship between the mean and the variance and therefore can meaningfully be used to compare questions with different means. By using the computational geometry techniques, a new way to determine the index of disagreement $\Phi_{CG}(v|m)$ (or measure of the consensus $\Psi_{CG}(v|m)$) has been developed. This paper shows that we can get the same values of disagreement and then the consensus by using simple geometric and numerical steps without using many cases and conditions.

These new algorithms can easily be computed using your favorite software package with just a few steps for calculations. The authors have developed a simple spreadsheet for computing Ψ that allow the users to enter the mean and variance to easily compute this measure of consensus.

Future work includes proving that this index can be generalized to larger

values of n and demonstrating with Likert data from several disciplines that Ψ can add to the interpretation of the consensus within groups, between groups, or the change in the consensus of a group across time.

Chapter 4

Paper 3: Using a-Multidimensional Approach for Generalizing a Consensus Measure to Likert Scales of Any Size n

This paper was submitted in March 2017 in the Knowledge-Based
Systems Journal in ELSEVIER Inc.

Abstract

Many consensus measures use Likert data for comparison. Although these measures should work with any number n of choices on the Likert scale, the measurements have been most widely studied and demonstrated for $n = 5$. One measure of consensus introduced by Akiyama et al. that depends on both the mean and variance, gives results that can differentiate between some group consensus behaviors better than measures that rely on either just the mean or just the variance separately. However, this measure is more complicated and not easy to apply and understand. This paper addresses these two common problems by introducing a new way for calculating the measure of consensus and the new method presented in this article works for any number of Likert item choices. Because the Akiyama et al. measure depends on both the mean and the variance it is more complicated to compute, so to avoid this, this paper presents new algorithms that are easy to apply and to understand. The novelty is using a computational approach in n -Dimensional space. Numerical examples in three-dimensional (for $n=6$) and four-dimensional (for $n=7$) spaces are provided in this paper to assure the accuracy of results.

4.1 Introduction

The significance of reaching consensus in a group of persons or among groups can easily be appreciated by anyone who has ever been involved in a group decision-making process. Indeed, some researchers believe that coming to consensus plays the key role in group decision-making. [26], [61].

The importance of consensus can be further appreciated when we consider the vast number of fields, other than just group decision making, that use consensus. For example, politics, economics, social choice and science all use the idea of consensus [8], [6], [36]. When you consider the work of Lehrer and Wagner [42], you can see how some factor of agreement (or consensus) is necessary for some central conceptions of thinking.

Consensus takes a significant place in many major areas, therefore there is a demand for an accurate gauge to measure consensus. The central idea of many researchers working in this field of study concerns how to build consensus (or diminish the disagreement) among all people in one group or in more than one group. Many of these measures have an iterative process to try to come to agreement or build consensus [31], [32], [46], [61]. If your task is to build consensus by using a multistage process within a group as in [21] or even for each person individually as in [60], you will need at the end of each stage a useful instrument or measurement to measure consensus.

There are several different meanings or definitions in the literature for the term “consensus” or “disagreement”. However, these two concepts are always antonyms, or mathematically each one as the complement of the other. In this paper, we use the term consensus to mean an opinion or belief reached by a group of persons who can agree on Likert scale items. While disagreement refers to a difference of opinion or perception.

The research to find a mathematical or statistical measure of consensus (or index of disagreement) began with the researchers working to find a way to build consensus within a group. The simplest and most widely used measures of consensus are the percent agreement measure and the vari-

ance. The percent agreement measure has been used in different cases and has been applied to small group consensus. This measure gives only a percentage of team members who accept a particular opinion [23], [49]. The variance measure, evaluated from Likert items, is usually used to talk about the disagreement (or lack of consensus) [44]. This statistical measurement is significant in precise comparison situations. However, the variance can be useless in cases of comparing a different size of groups or for groups that have different means[17].

Another common approach to calculating the consensus is r_{WG} , that represented in [35] by James et al. This measure is based on the variance and is sometimes called inter-rater reliability. Two other measures of consensus are presented by Kendall [38], and Alcalde–Unzu and Vorsatz [5]. The measure Kendall proposes, usually known as Kendall’s tau (τ), is limited to the status of two individuals and directly calculates the ratio of pairwise comparisons of how two people agree. The second measures, symbolically known as (σ), introduced by Alcalde–Unzu and Vorsatz. This measure determines for any pair of alternatives the absolute value of the difference between the ratio of persons who choose one alternative and the proportion of individuals who choose the other option and then takes the average of these numbers over all possible pairs of alternatives[5]. The concepts, comparisons, and properties of these two measures well introduced in [4].

There are many other more advanced methods presented to measure the consensus within a group of individuals. The method of Beliakov et al. [10] concentrates on structuring a function that can measure the degree of consensus from a set of inputs provided as numbers from the unit interval. In the information theory field, Shannon’s [54] formula for Entropy, or the measure of the degree of disorder of a system [9], is widely used. The formula is $\sum p_i \log p_i$ (where p_i is the probability of i^{th} event’s outcome). Tastle and Weirman [58] [59] apply Shannon’s formula to Likert scales responses. González-Arteaga [26] provides an approach to consensus measurement based

on the Pearson correlation coefficient.

Based on what is presented above, there are several ways to define consensus and various ways to measure consensus within and among groups. However, as far as we can tell, there is no approach using computational geometry concepts in n -Dimensional space as introduced in this paper. In this article, we consider a new method to determine the measure of consensus (or conversely the index of disagreement) presented by Akiyama et al. [3] with the key to this new method using computational geometry concepts presented in Abdal Rahem and Darrah [2]. The researchers in [3] present the sophisticated index of disagreement, called Φ , and the measure of consensus, Ψ , that used the conditional distribution of the variance for a given mean[11]. The distinction of this index is it allows for the comparison of consensus values of various questions for the same group or the same question for diverse groups, even comparisons of groups with different sizes and the questions with different means. However, in [3] this index is proven to work for $n = 5$ Likert scales only and is difficult to understand and not easy to apply.

Consequently, Abdal Rahem and Darrah in [2] provide a straightforward algorithmic method to compute the reliable measure introduced in [3]. The idea of this new approach is much easier to use and understand than the way of calculating the index of disagreement presented in [3]. However, in [2], the approach is shown for only two-dimensional space, which relates to $n = 5$ Likert scale. The aim of the research presented in this article is to break the limitation of working with $n = 5$ Likert scale only. Likert items are used in many disciplines to measure attitudes, preferences, and subjective reactions[24], [44], [57]. For this discussion, a global Likert scale is used, with the integers 1 through n corresponding to the words strongly disagree through strongly agree or any other words the researcher prefers or question suggests.

The rest of the paper is organized as follows. Section 4.2 presents the theoretical foundation of our work and defines the index of disagreement

for any number of Likert scale and provide a visual to illustrate how the equations in section 4.2 look for any dimension. In Section 4.3 the techniques are applied to $n = 6$. Section 4.4 provides an examples of special cases of sets of probabilities as well as a specific means with different variances of each fixed mean. Finally, some concluding remarks and future works are discussed in Section 4.5.

4.2 Theoretical Foundations

The main theoretical foundations of the consensus measurement are introduced in Akiyama et al. [3]. There we introduce the idea that variance is always a function of the mean. For any fixed mean, there will be a range of variances that will be possible. Then given a mean m and a variance v , we find the consensus measure for this two numbers by using conditional probability to determine the “ratio between a part of the range of the variance for that mean to the total range of variance for that mean”.

In the first paper [3], this measurement is computed by a series of analytical steps for $n = 5$. We first algebraically reduce the computation to a set of three equations in two dimensions. Then by finding the areas between pairs of these equations and subtracting these two area, we can find a number that represents a measurement for area given a particular mean, m , and variance, v . We use this method to define a function $A(v)$ and integrate this function from the $h(\text{minimum } v)$ to $h(v)$ for the given v . Finally, we divide this number by the integration over the total range of variance, $h(\text{minimum } v)$ to $h(\text{maximum } v)$, to get the index of disagreement Φ . The measure of consensus, Ψ , will then be $1-\Phi$.

In the second paper, Abdal Rahem and Darrah [2], simplify the calculation of these areas and make it possible to generalize to any n , two algorithms using computational geometry are presented to replace the complicated analytical steps involving calculus. The first algorithm focuses on computing

$A(v)$ and the second algorithm is for using numerical methods to compute the ratio of the integration. Also, all the background from computational geometry and conditional probability are located in [2]. Furthermore, Abdal Rahem and Darrah [2] provides a basis for the generalization work to $n > 5$ that follows.

4.2.1 Generalization of Calculations to $n > 5$ for Index of Disagreement (conversely Consensus Measure)

Let us start by introducing the notation for the two main variables we will work with. Suppose m and v refer to the mean and the variance computed from the survey wherein the respondents are asked to pick out exactly one answer from the items: $1, 2, \dots, n$.

Other common notations used throughout are the probabilities $p_i, i = 1, 2, \dots, n$. Note that, as usual, all probabilities in our discussions are values p_i that satisfy: $0 \leq p_i \leq 1$. Moreover, $A(v)$ refer to an area with points $(x_1, x_2, \dots, x_{n-3})$ in R^{n-3} space.

As utilized in [3], the basic equations for the foundation to compute the index of disagreement are the sum of all probabilities is equal to one, the equation for computing the mean, and the equation derived from computing the variance of random variables. Mathematically, we can write the system of equations as follows:

$$\begin{aligned}
 p_1, p_2, p_3 \dots, p_n &\geq 0. \\
 p_1 + p_2 + p_3 \dots + p_n &= 1. \\
 p_1 + 2p_2 + 3p_3 \dots + np_n &= m. \\
 p_1 + 4p_2 + 9p_3 \dots + n^2p_n &= m^2 + v.
 \end{aligned} \tag{4.1}$$

Using the notion $x_1 = p_4, x_2 = p_5, \dots, x_{n-3} = p_n$, we can solve (4.1) for p_1, p_2, p_3 in terms of x_1, x_2, \dots, x_{n-3} as below:

$$\begin{aligned}
p_1, p_2, p_3, \dots, p_n &\geq 0. \\
p_1 &= \frac{1}{2}(v + m^2 - 5m + 6) - f_1(x_1, x_2, \dots, x_{n-3}). \\
p_2 &= -(v + m^2 - 4m + 3) + f_2(x_1, x_2, \dots, x_{n-3}). \\
p_3 &= \frac{1}{2}(v + m^2 - 3m + 2) - f_3(x_1, x_2, \dots, x_{n-3}).
\end{aligned} \tag{4.2}$$

Where f_1, f_2 and f_3 are functions of $n - 3$ variables. Since we have all the probabilities greater than zero ($p_1 \geq 0, p_2 \geq 0, p_3 \geq 0$), we can rewrite the system above as following:

$$\begin{aligned}
p_1, p_2, p_3, \dots, p_{n-3} &\geq 0. \\
\frac{1}{2}(v + m^2 - 5m + 6) - f_1(x_1, x_2, \dots, x_{n-3}) &\geq 0. \\
-(v + m^2 - 4m + 3) + f_2(x_1, x_2, \dots, x_{n-3}) &\geq 0. \\
\frac{1}{2}(v + m^2 - 3m + 2) - f_3(x_1, x_2, \dots, x_{n-3}) &\geq 0.
\end{aligned} \tag{4.3}$$

or equivalently

$$\begin{aligned}
x_1, x_2, \dots, x_{n-3} &\geq 0. \\
f_1(x_1, x_2, \dots, x_{n-3}) &\leq \frac{1}{2}(v + m^2 - 5m + 6). \\
f_2(x_1, x_2, \dots, x_{n-3}) &\geq (v + m^2 - 4m + 3). \\
f_3(x_1, x_2, \dots, x_{n-3}) &\leq \frac{1}{2}(v + m^2 - 3m + 2).
\end{aligned} \tag{4.4}$$

In order to simplify (4.4), let we define t and a function $h_m(v)$ as below:

$$t = h_m(v) = \frac{1}{2}(v + m^2 - 3m + 2) \tag{4.5}$$

where m is our fixed mean and v is any variance in the range with respect to that mean. Therefore, (4.4) will be:

$$\begin{aligned}
x_1, x_2, \dots, x_{n-3} &\geq 0. \\
f_1(x_1, x_2, \dots, x_{n-3}) &\leq t - m + 2. \\
f_2(x_1, x_2, \dots, x_{n-3}) &\geq 2t - m + 1. \\
f_3(x_1, x_2, \dots, x_{n-3}) &\leq t.
\end{aligned} \tag{4.6}$$

$A(v)$ can be defined as the set of all $(n - 3)$ -tuples, $(x_1, x_2, \dots, x_{n-3})$, in the $(n - 3)$ -Dimensional space that satisfy (4.6). Due to the symmetry of the mean with respect to the midpoint, in the subsequent discussions we will restrict the range of the mean to $1 \leq m \leq \frac{n+1}{2}$. The remaining portion $\frac{n+1}{2} \leq m \leq n$ can be treated as the symmetric reflection of $1 \leq m \leq \frac{n+1}{2}$.

In order to find the minimum t ($\min t$) and the maximum t ($\max t$), substitute the minimum v ($\min v$) and maximum v ($\max v$) respectively in h_m . Consequently, $t = h_m(v)$ is a linear (one-to-one) mapping of the interval $[\min v, \max v]$ onto $[\min t, \max t]$. The inverse of $t = h_m(v)$ is also a linear mapping of $[\min t, \max t]$ onto $[\min v, \max v]$ given by $v = h_m^{-1}(t) = 2t - m^2 + 3m - 2$.

For purposes of finding the area $A(v)$ for fixed m , we derive three equations from (4.6) to get three lines, planes or hyperplanes in R^{n-3} space, say $g_i(x_1, x_2, \dots, x_{n-4})$ where $i = 1, 2, 3$ that subdivided the R^{n-3} space. For example, in R^2 we have:

$$\begin{aligned}
g_1(x) &= \frac{t - m + 2}{3} - \frac{x}{3} \\
g_2(x) &= \frac{2t - m + 1}{8} - \frac{3x}{8}. \\
g_3(x) &= \frac{t}{6} - \frac{x}{2}.
\end{aligned} \tag{4.7}$$

and in R^3 we get:

$$\begin{aligned}
g_1(x, y) &= \frac{t - m + 2}{6} - \left(\frac{x + 3y}{6}\right) \\
g_2(x, y) &= \frac{2t - m + 1}{15} - \left(\frac{3x + 8y}{15}\right). \\
g_3(x, y) &= \frac{t}{10} - \left(\frac{3x + 6y}{10}\right).
\end{aligned} \tag{4.8}$$

We set $f_3(x_1, x_2, \dots, x_{n-3}) \leq t$ which leads to defining $g_3(x_1, x_2, \dots, x_{n-4})$, so we compare $g_1(x_1, x_2, \dots, x_{n-4})$ and $g_2(x_1, x_2, \dots, x_{n-4})$ to $g_3(x_1, x_2, \dots, x_{n-4})$. If we only compare $g_1(x_1, x_2, \dots, x_{n-4})$ and $g_2(x_1, x_2, \dots, x_{n-4})$ to find an area between them (if its exist) this will not give us any information related to t . Hence, we define $A(v)$ as follow:

$$A(v) = A_1(v) - A_2(v)$$

where

$$A_1(v) = \{(x_1, \dots, x_{n-3}) \mid g_2(x_1, \dots, x_{n-4}) \leq x_{n-3} \leq g_3(x_1, \dots, x_{n-4})\} \tag{4.9}$$

and

$$A_2(v) = \{(x_1, \dots, x_{n-3}) \mid g_1(x_1, \dots, x_{n-4}) \leq x_{n-3} \leq g_3(x_1, \dots, x_{n-4})\} \tag{4.10}$$

Note that all x_i are probabilities, so we restrict, $0 \leq x_1, x_2, \dots, x_{n-4} \leq 1$.

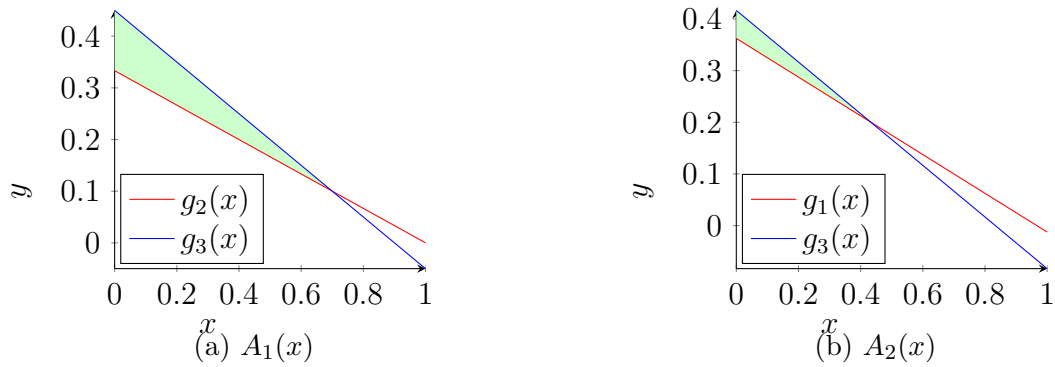


Figure 4.1: Example for The Area $A_1(v)$ and $A_2(v)$ in R^2

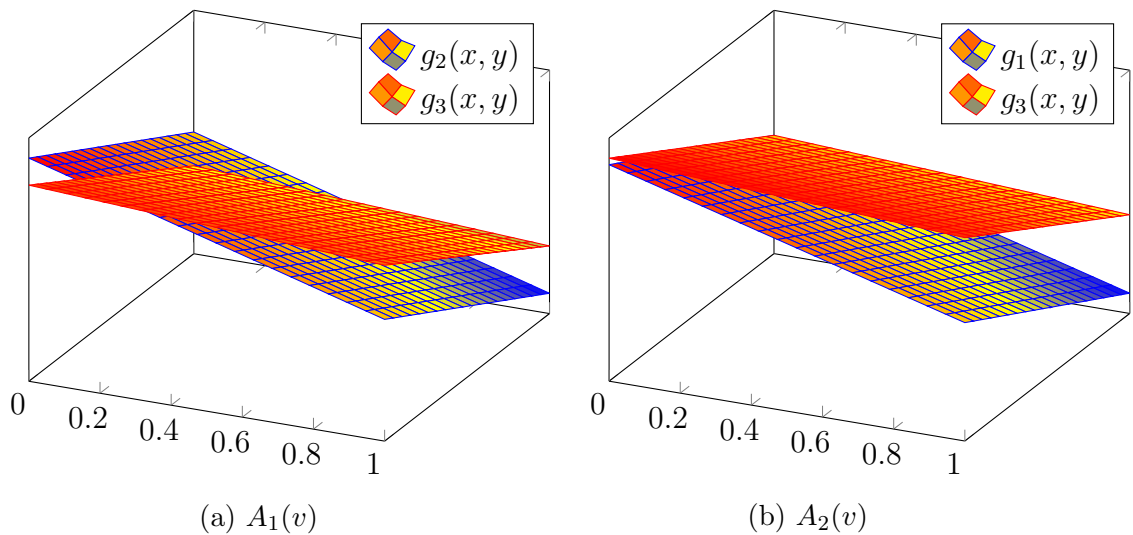


Figure 4.2: Example for of Planes used to Compute Areas $A_1(v)$ and $A_2(v)$ in R^3

Now that we have presented all the equations required for restricting the target area, the next step is to compute these areas in any dimension.

4.2.2 Determine the Index of Disagreement

Although mathematicians may prefer an analytical process to get an exact solution, in many cases it is extremely hard, or even impossible to find one.

The way of finding the area $A(v)$ introduced in Akiyama et al. [3] for $n = 5$ uses an analytical method to find the exact solution; however, it has many cases that must be considered. Therefore, the new computational approach presented in this paper finds a very good solution (as good as is required) and can be generalized to any $n \geq 5$.

To examine what it means to determine these areas or volumes in n -dimensional space, we can first look at the problem in 2-dimensional space. The equations $g_i(x)$ are line segments and so the intersections are points. To find the desired area, we simply find the area bounded between two lines that satisfies $0 \leq x \leq 1$ and $0 \leq y \leq 1$, since x and y are probabilities. In this case, there are two ways to compute the area, analytically [3] and by using the computational geometry concepts [2].

By adding one more probability, the volume then is in three dimensions. In this case, the equations $g_i(x, y)$ are planes. This means the intersections are lines instead of points. Moreover, for $n = 7$, we must work in four dimensions. Thus, the equations $g_i(x, y, z)$ are hyperplanes and the intersections of any two hyperplanes is a 3-dimensional object, and so on.

Consequently, since the region has such a strange shape in general, calculating its area or volume proves to be very difficult with analytical methods, especially as we go to higher dimensions. But calculating the area of rectangles or cubes is simple. We will use this method to simplify our calculations by subdividing the region we want to measure into small squares, cubes, or hypercubes as is a common method for approximating an area or volume. One of the popular approaches to finding the area under a curve numerically is by using the *Riemann Sum*. Employing the same idea as the Riemann sum, with some modification to make it more accurate and to better fit with our problem, we can determine the area or volume of the required region.

Before presenting the steps of the algorithms, recall that the area for any n is bounded by $[0, 1]$ (i.e. $0 \leq x_i \leq 1$ for any $x_i \in R^{n-3}$).

Algorithm

Divide the interval $[0, 1]$ for every x_i -axis into N subintervals of length Δ , ($\Delta = \frac{1}{N}$), where N is a positive integer

For $i_1 = 1$ to N

For $i_2 = 1$ to N

\vdots

For $i_{n-3} = 1$ to N

$$s_j = s_{i_1, i_2, \dots, i_{n-3}}$$

If s_j satisfy (4.9) then $A_1 = A_1 + \Delta^{n-3}$ (add one hypercube to the area)

If s_j satisfy (4.10) then $A_2 = A_2 + \Delta^{n-3}$ (add one hypercube to the area)

$$A(v) = A_1 - A_2.$$

Note that in this approach of finding the area, the larger the N (i.e. the smaller Δ), the closer the estimate gets to the exact area.

Once you get $A(v)$, you can use the same algorithm for determining the integration as we presented in [2] to compute the index of disagreement and the consensus value.

4.3 Index of Disagreement in 3D

In order to ensure that the methods described above produce acceptable results they are applied to different cases. For $n = 5$, you can find the examples with details in Akiyama et al. [3] or Abdal Rahem and Darrah [2].

For $n = 6$, our example is in 3-dimensional space, and the equations $f_i(x, y, z), i = 1, 2, 3$ are as follows:

$$\begin{aligned} f_1(x, y, z) &= x + 3y + 6z. \\ f_2(x, y, z) &= 3x + 8y + 15z \\ f_3(x, y, z) &= 3x + 6y + 10z \end{aligned} \quad (4.11)$$

Which means the system of inequities in (4.6) becomes:

$$\begin{aligned} x, y, z &\geq 0. \\ x + 3y + 6z &\leq t - m + 2. \\ 3x + 8y + 15z &\geq 2t - m + 1. \\ 3x + 6y + 10z &\leq t. \end{aligned} \quad (4.12)$$

For determining the volume restricted by two planes and the z -axis, apply the algorithm above. This will find the volume by adding each small cube when s_j satisfies one of the systems of equations to get A_1 and A_2 and then use those to find $A(v)$.

4.4 Numerical Example

Now we can look at some example to see if the method produces the desired results. First, we consider some special cases of sets of probabilities. We use the probabilities to determine the mean and the variance. We apply the algorithms above to compute the index of disagreement and then consensus values for these special cases to show that they make sense.

The first we consider the case when all the respondents have chosen the same response. In other words, when one of the probabilities, say p_6 , is one and all others are zeros. This case gives us $m = 6$ and $v = 0$, and therefore the ratio of the index is $\Phi_{n=6} = \frac{0}{0.000443} = 0$ which is gives the consensus value of one, $\Psi = 1 - 0$.

The second case is when the responses are evenly distributed at opposite ends of the scale. For example, probabilities are 0.5 for p_1 and p_6 , and all others are zeros. This time will get the mean $m = 3.5$ and the variance $v = 6.25$. These values of the mean and variance lead to full disagreement (i.e. $\Phi = 1$) since the ratio is $\Phi = \frac{0.0052}{0.0052} = 1$, which then means the consensus is zero. Notice that the results in these two cases are exactly what we would expect and also what we computed for $n = 5$ in these situations using the original method.

Before looking at several different values for the mean and variance, we will look at one more special case. Now, suppose all probabilities are equal. That means $p_i = \frac{1}{6}, i = 1, 2, \dots, 6$. The mean in this case is $m = 3.5$ and the variance is $v = 2.92$. These two values for mean and variance imply $\Phi = \frac{0.0034}{0.0052} = 0.6538$. Subsequently, the consensus $\Psi = 1 - 0.6538 = 0.3462$ compared to $\Phi = 0.6666$ and $\Psi = 0.3334$ for $n = 5$ in the similar situation and again computing these values with the original method.

For selected values of m and v , table (1) shows the different cases of index of disagreement, for $n = 6$, Φ and the consensus Ψ .

Table 1. The Index of Disagreement Φ and The Consensus Ψ for selected values of m and v .

Mean	Variance (v)	$\Phi_{n=6}$	$\Psi_{n=6}$
1.1	0.09	0.0000	1.0000
	0.1	0.0238	0.9762
	0.2	0.3562	0.6435
	0.3	0.755	0.245
	0.39	0.9595	0.0405
1.2	0.49	0.0000	1.0000
	0.16	0.0000	1.0000
	0.2	0.0134	0.9866
	0.3	0.0874	0.9126
	0.4	0.2609	0.7391
	0.5	0.5275	0.4725
	0.6	0.7736	0.2264
	0.7	0.9374	0.0626
	0.76	0.9899	0.0101
	0.96	1.0000	0.0000

The results of all examples above are reasonable and acceptable especially if we compare these results with similar cases for, $n = 5$, in [2].

4.5 Conclusions and Future Work

A new approach for generalizing the consensus measure (or index of disagreement) presented in Akiyama et al. [3] has been developed to work for Likert scales with any number of choices. By using computational geometry and n -dimensional space concepts, this paper not only has been able to generalize all the work constructed in [3] and [2] but show comparable results to those

obtained in last two articles. The difference of this approach compared to the other consensus measures, which also works for any Likert items, is that this measure considers the conditional distribution of the variance for a given mean. In other words, this method provides more information than others that are only based on just the mean or just the variance.

Another distinguishing feature of this new method is that it is easy to understand and apply. In fact, we can summarize this work as easy as A, B, C. (A) Set up the equations g_i , (B) Determine the areas, and (C) calculate the integrations and then the index of disagreement. Additionally, by using these simple steps in an Excel spreadsheet, you can get all the variable defined above, and for (B) and (C), all you need to do is enter the algorithms into any software you prefer. For instance, all the examples in this paper were computed using Microsoft Excel spreadsheet and Matlab.

In future studies, we plan to develop the same ideas to compute the consensus measure for continuous scales. We also plan to continue looking for easier and faster methods to find the areas and volumes described in the paper by using advanced computation geometry and high-dimensional statistics concepts [53], [51].

Chapter 5

Discussions and Open Problems

Since the Likert scale is one of the most popular scales used in different areas, the consensus measurement presented in this work also be very useful in many disciplines. Although there are various methods of measuring consensus that rely on the Likert scale, there remains a need for a measure that gives accurate results or more information as well as works for any numbers of scales not just for small n . We have dealt with more than one way to build a new consensus measurement. The ways of computing the measure of consensus presented in this thesis relied on calculus, statistical and computational geometry concepts in two and more dimensional space.

The mean and variance are the standard measures of consensus that work for any number on the Likert scale and other common scales. However, Tastle in [59] shows that neither mean nor standard deviation (a square root of variance) alone are adequate to get an accurate measure of consensus. Moreover, the second chapter presents a comparison between the new index of disagreement Φ (or measure of the consensus Ψ) and two other measures, r_{WG} by James et al.[35] and $Cns(X)$ by Tastle and Wierman [59]. The comparison shows that the new measure Ψ provides additional or varying information because it relies on the mean and the variance at the same time. The outcome of the comparison given initially shows there is a relationship

between the mean and the variance, for each fixed mean there is a range of variance values that depend on that mean. This relationship along with the conditional probability concepts led us to build the new index of disagreement (or the measure of consensus) with consideration of both the mean and the variance.

Notice that the focus of the Chapter 2 (or first paper) ensures that the new index of disagreement was useful and provided additional information, but it had some complicated steps of calculus. The utility of the measure is demonstrated in a real life example of Likert data from a survey measuring organizational climate in an academic department at West Virginia University. After introduction of the measure, the primary objective was to provide an easier way to apply and work with the new measure with any number of Likert items.

In order to generalize of the new consensus measurements Ψ presented in the second chapter for any size of Likert scale, a new way to compute the measure was introduced in the Chapter 3 (second paper) by using computational geometry and numerical analysis concepts. This new approach provides a way to get the same useful result of Ψ by using only two easy algorithmic steps. While the first algorithm focuses on replacing the analytical steps with computational geometric procedures, the concentration of the second algorithm was on finding the required integrals by using a numerical methods.

Any researcher seeking to deal with multidimensional space is able to notice that working in more than two or three dimensional space has both advantages and disadvantages. Multidimensional space opens up opportunities to work with various problems, but it may cause some difficulties in setting up operations or methods to process steps. Two important ideas are presented in Chapter 4 (or third paper). First, the ability to generalize the new index of disagreement to work with any number of Likert scale. Second, introducing a new method for getting accurate results for the consensus

values by using simple steps of an algorithmic approach that applies in any dimensional space.

In addition to all the features that are mentioned above, this work opens more avenues for further studies and discussions in future. The first perspective that can be brought into focus is generalizing the index of disagreement Φ by using the statistical concepts of expectation value $E(X)$. For example, when $n = 5$, the last two equations in (2.8) are equal to $E(X)$ and $E(X^2)$ respectively. For higher n you can just add more equations depend on the definition of r^{th} moment, $E(X^r)$ [28]. In fact, we investigated this direction and made satisfactory progress of all requirement and logical steps using the idea of expectation values. We added the equations of $E(X^3)$ and $E(X^4)$ for $n = 7$ which resulted in five lines of equations instead of three in (2.12) that give $A(v)$. We implemented all these steps using Excel to examine different cases of the mean and the variance values using this method.

The advanced computation geometry properties or high-dimensional statistical methods were utilized to develop the algorithms in Chapter 3 for easier and faster ways can also be done it different ways. The advance computational geometry concepts of convex hull could be used to improve the first algorithm for finding the area/volume of $A(v)$ [53]. While using the high-dimensional statistics methods such as the Monte Carlo method would be a practical improvement to the second algorithm [52].

Another gap for further studies in the future is developing the multidimensional algorithm that is introduced in Chapter 4. This could be done by the utilizing the idea of finding the area of any polygons in the complex plane [55], with advanced linear algebra technique [62], or by developing some advanced optimization methods like a “Hook and Jeeves” method [14].

Finally, expanding the domain of the index of disagreement Φ , presented in Chapter 2, to work with any continuous or partial scale. Furthermore, we will work to develop an interactive website for users to enter the mean and variance that can be used to compute such measures of consensus easily. We

have already built two programs for finding the index of disagreement and then the measure of consensus for $n = 5, 6$, and 7 , in two, three and, four dimensions. The first program is written in Excel spreadsheet by using Visual Basic (VB) language, while the second program is written in MATLAB.

Bibliography

- [1] M. Abd Al-Rahem, Using a-Multidimensional Approach for Generalizing a Consensus Measure to Likert Scales of Any Size n , Knowledge-Based Systems (2017) (Submitted).
- [2] M. Abd Al-Rahem, M. Darrah, A Geometric Approach for Computing a Measure of Consensus for Groups, International Mathematical Forum, 11 (2016) 961-973.
- [3] Y. Akiyama, J. Nolan, M. Darrah, M. Abd Al-Rahem, L. Wang, A Method for Measuring Consensus Within Groups: An Index of Disagreement Via Conditional Probability, Information Sciences, 345 (2016) 116–128.
- [4] J. Alcalde-Unzu, M. Vorsatz, Measuring consensus: Concepts, comparisons, and properties, Consensual Processes, Fuzziness and Soft Computing, 267 (2011) 195–211.
- [5] J. Alcalde-Unzu, M. Vorsatz, The measurement of consensus: an axiomatic analysis. FEDEA Working Paper 28 (2008).
- [6] S. Alonso, E. Herrera-Viedma, F. Chiclana, A. Herrera, A web based consensus support system for group decision making problems and incomplete preferences, Information Sciences, 180 (2010) 4477-4495.

- [7] S. Alonso, I. J. Pérez, F. J. Cabrerizo, E. Herrera-Viedma, A linguistic consensus model for Web 2.0 communities, *Appl. Soft Comput. J.*, 13 (1) (2013) 149–157.
- [8] R. M. Alston, J. R. Kearl, M. B. Vaughan, Is there consensus among economists in the 1990's, *Paper and Proceedings American Economic Review*, 82 (1992) 203-209.
- [9] K. D. Bailey, *Social Entropy Theory*, State University of New York, (1990).
- [10] G. Beliakov, T. Calvo, S. James, Consensus measures constructed from aggregation functions and fuzzy implications, *Knowledge-Based Systems*, 55 (2014) 1-8.
- [11] A. G. Bluman, *Elementary Statistics: A Step by Step Approach*, Eighth edition, The McGraw-Hill, USA, (2012).
- [12] S. Boroushaki and J. Malczewski, Measuring consensus for collaborative decision-making: A GIS-based approach, *Comput. Environ. Urban Syst.*, 34 (4) (2010) 322–332.
- [13] R. Bosch, *Characterizations of Voting Rules and Consensus Measures*, Tilburg University, (2005) Ph.D. thesis.
- [14] D. B. Brian, *Basic Optimization Method*, Edward Arnold, London (1984).
- [15] J. Carifio, R. J. Perla, Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends about Likert Scales and Likert Res ponse Formats and their Antidotes, *Journal of Social Sciences*, 3 (3) (2007) 106-116.

- [16] F. Chiclana, J. M. Tapia García, D. Moral, M. Jose, and E. Herrera-Viedma, A Statistical Comparative Study of Different Similarity Measures of Consensus in Group Decision Making, *Information Sciences*, 221 (2013) 110–123.
- [17] L. G. Conway, M. Schaller, Methods of Measurement of Consensual Beliefs Within Groups, *Group Dynamics: Theory, Research, and Practice* 2 (4) (1998) 241-252.
- [18] M. DeBerg, O. Cheong, M. Kreveld, and M. Overmars, *Computational Geometry: Algorithms and Applications*, Third Edition, Springer, USA, 2008.
- [19] <http://dictionary.cambridge.org/us/dictionary/english/consensus>, 2016.
- [20] Y. Dong, Guiqing Zhang, Wei-Chiang Hong, Yinfeng Xu, Consensus models for AHP group decision making under row geometric mean prioritization method, *Decision Support Systems* 49 (2010) 281-289.
- [21] Y. Dong, H. Zhang, Multiperson decision making with different preference representation structures: A direct consensus framework and its properties, *Knowledge-Based Systems* 58 (2014) 45-57.
- [22] R. Frigg, C. Werndl, *Entropy – A Guide for the Perplexed*. Oxford University Press (2010).
- [23] R. D. Gailbreath, S. L. Wagner, R. G. Moffett III, M. B. Hein, Homogeneity in behavioural preferences among U.S. army leaders, *Group Dynamics: Theory, Research, and Practice* 1 (1997) 222-230.
- [24] P. L. Garder, Scales and Statistics, *Review of Educational Research* 45 (1) (1975) 43-57.

- [25] J. W. Gibbs, *The scientific papers of J. Willard Gibbs*, Longmans, Green, and Co. (1906) 11-12.
- [26] T. González-Arteaga, R. A. Calle, F. Chiclana, A new measure of consensus with reciprocal preference relations: The correlation consensus degree, *Knowl. Based Syst.* 107 (2016) 104–116.
- [27] S. Grant, T. Aitchison, E. Henderson, J. Christie, S. Zare, J. McMurray, H. Dargie, A Comparison of the Reproducibility and the Sensitivity to Change of Visual Analogue Scales, Borg Scales, and Likert Scales in Normal Subjects During Submaximal Exercise, *Chest.* 116 (5) (1999) 1208–1217.
- [28] C. M. Grinstead, *Grinstead and Snell’s Introduction to Probability*, University Press of Florida (2009).
- [29] L. Guibas, *Basic algorithms and combinatorics in computational geometry*, Stanford university, USA.
- [30] A. P. Hare, A Study of Interaction and Consensus in Different Sized Groups, *American Sociological Review*, 56 (1952) 261-267.
- [31] E. Herrera-Viedma, R. Herrera, F. Chiclana, A consensus model for multiperson decision making with different preference structures, *IEEE Transactions on Systems, Man and Cybernetics – Part A: Systems and Humans*, 32 (2002) 394-402.
- [32] E. Herrera-Viedma, L. Martinez, F. Mata, F. Chiclana, A consensus support system model for group decision-making problems with multi-granular linguistic preference relations, *IEEE Transactions on Fuzzy Systems* 13 (2005) 644-658.
- [33] E. Herrera-Viedma, F. J. Cabrerizo, J. Kacprzyk, and W. Pedrycz, “A review of soft consensus models in a fuzzy environment,” *Inf. Fusion*, 17 (2014), 4–13.

- [34] T. A. Hill, An Experimental Study of the Relationship Between Opinionated Leadership and Small Group Consensus, *CM*, 43 (1976) 246-257.
- [35] L. R. James, R. G. Demaree, G. Wolf, r_{WG} : An Assessment of Within-Group Interrater Agreement, *Journal of Applied Psychology* 78 (2) (1993) 306-309.
- [36] T. Justice, D.W. Jamieson, T. Justice, D.W. Jamieson, *The Facilitator's Fieldbook: Step-by-Step Procedures*. AMACON, a division of the American Management Association, (2012).
- [37] R. M. Kaplan, D. P. Saccuzzo, *Psychological Testing: Principles, Applications, and Issues*. (8th ed.). Belmont, CA (2010).
- [38] M. Kendall, *Rank correlation methods*. 3rd edition, Hafner Publishing Company, New York, (1962).
- [39] D. A. Kenny, D. C. Rubin, Estimating chance reproducibility in Guttman scaling, *Social Science Research*, 6 (1977) 188-196.
- [40] J.A. Kline, Orientation and Group Consensus, *Central States Speech Journal*, 23 (1972) 44-47.
- [41] T. J. Knutson, An Experimental Study of the Effects of Orientation Behavior on small group of consensus, *Speech Monographs*, 39 (1972) 159-165.
- [42] K. Lehrer, C. Wagner, *Rational consensus in science and society : a philosophical and mathematical study*, Philosophical studies series in philosophy, (1981).
- [43] J. Michell, Quantitative science and the definition of measurement in psychology, *British Journal of Psychology*, 88 (3) (1997) 355-383.
- [44] G. Norman, Likert scales, levels of measurement and the 'laws' of statistics, *Advances in Health Science Education* 15 (2010) 625-632.

- [45] J. O'Rourke, computational geometry in C, Second edition, Cambridge University Press, (1998).
- [46] R. Parreiras, P. Ekel, J. S. C. Martini, R. M Palhares, A flexible consensus scheme for multicriteria group decision making under linguistic assessments, *Information Sciences*, 180 (2010) 1075-1089.
- [47] R. Parreiras, P. Ekel, F. Bernardes Jr., A dynamical consensus scheme based on a nonreciprocal fuzzy preference relation modeling, *Information Sciences*, 211 (2012) 1-17.
- [48] W. Pedrycz, P. Ekel, R. Parreiras, *Fuzzy Multicriteria Decision-Making: Models, Methods, and Applications*, Wiley, Chichester, (2011).
- [49] H. Prapavessis, A.V. Carron, Sacrifice, cohesion, and conformity to norms in sport teams, *Group Dynamics: Theory, Research, and Practice* 1 (1997) 231-240.
- [50] J. R. Quinlan, Induction of Decision Trees, *Machine Learning*, 1 (1986) 81-106.
- [51] P. Rigollet, *High-Dimensional Statistics*, Massachusetts Institute of Technology: MIT OpenCourseWare, <https://ocw.mit.edu>, (2015).
- [52] C. P. Robert, G. Casella, *Monte Carlo Statistical Methods*, Springer Texts in Statistics, (2004).
- [53] M. Shabbir, *Some results in computational and combinatorial geometry*, The State University of New Jersey, USA, (2014).
- [54] C. E. Shannon, The Mathematical Theory of Communication, *Bell Systems Technical Journal*, (1948).
- [55] G. C. Shephard, Regular complex polytopes, *Proc. London Math. Soc.* (3) 2 (1952) 82-97.

- [56] B. Spillman, J. Bezdek and R. Spillman, Development of an instrument for the dynamic measurement of consensus, *Communication Monographs*, 46 (1979) 1-11.
- [57] S. S. Stevens, On the Theory of Scales of Measurement, *Science*, 103 (1946) 677-680.
- [58] W. J. Tastle, M. J. Wierman, Consensus and Dissention: A New Measure of Agreement, in Proc. 2005 Annual Meeting of North American Fuzzy Information Processing Society (NAFIPS 2005), Michigan, (2005) 385-388.
- [59] W. J. Tastle, M. J. Wierman, Consensus and dissention: A measure of ordinal dispersion, *International Journal of Approximate Reasoning* 45 (3) (2007) 531-545.
- [60] A. K. Wisecup, Do We Have Consensus? : Examining the Sources of Systematic Variation in Cultural Identity Meanings, Duke University, (2011).
- [61] B. Zhang, Y. Dong, Y. Xu, Multiple attribute consensus rules with minimum adjustments to support consensus reaching, *Knowledge-Based Systems*, 67 (2014) 35-48.
- [62] J. Zhong, K. Tang, A. K. Qin, Finding Convex Hull Vertices in Metric Space Jinhong, *International Joint Conference on Neural Networks*, Beijing, China, (2014) 6-11.