

# Extended Voting Measures

Andrew Heard and Tim Swartz \*

## Abstract

This paper extends a number of voting measures used to quantify voting power. The extension is based on the recognition that individuals sometimes vote in coalitions. This observation gives rise to a statistical model which considers past voting patterns of subsets of eligible voters. The model is then used to obtain estimates of the probabilities of all voting combinations from which empirical measures are calculated. The calculation of the estimated probabilities may involve high-dimensional integrations. An example is given based on past decisions arising from the Supreme Court of Canada.

**Key words and phrases :** Bayesian modelling, voting power, Gibbs sampling, importance sampling.

**AMS 1991 subject classifications :** 62F10, 62P25, 65C05.

---

\*A. Heard is Associate Professor, Department of Political Science, Simon Fraser University, 8888 University Drive, Burnaby BC, Canada V5A1S6. T. Swartz is Associate Professor, Department of Mathematics and Statistics, Simon Fraser University, 8888 University Drive, Burnaby BC, Canada V5A1S6. T. Swartz was partially supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

## 1. INTRODUCTION

With complex voting rules, it is not always the case that voting members have equal strength. As a result, a number of indices have been introduced to quantify relative voting power. An incomplete list of such indices includes those proposed by Shapley and Shubik (1954), Banzhaf (1965), Deegan and Packel (1978), Johnston (1978) and Straffin (1978). An appraisal of the merits of the first four of these indices is given by Felsenthal and Machover (1995).

Whereas voting measures are widely used in the social sciences, their theoretical development has arisen primarily from the game theory literature. Consequently, the various measures are formal. In other words, the various measures are based on counting arguments and ignore statistical perspectives.

In this paper, we consider the use of historical voting data to extend voting measures. When voting members explicitly agree to vote in coalitions, the expressed relationships are used to determine traditional voting power indices. For example, if two members always decide to vote the same, traditional methods consider the two members as a single member with a double vote. It seems natural then to use historic data to capture the stochastic coalitions and anti-coalitions that exist between voting members.

The voting measures that we consider are the class of Banzhaf indices (Banzhaf (1965)) and the class of satisfaction indices (Straffin (1978)). These measures are most readily extended to “empirical” versions through the consideration of historical voting data. Not only do our empirical indices use the information available from historical voting data, but they also avoid

an unrealistic assumption used in the calculation of the traditional Banzhaf and satisfaction indices. The traditional Banzhaf and satisfaction indices are calculated under the assumption of equal probabilities for all voting combinations. Heard and Swartz (1997a) consider a simple statistical model based on the Dirichlet distribution to provide empirical Banzhaf indices. In this paper, we consider a more complex model which expresses the random historical data in terms of unknown voting probabilities. The proposed model permits historical data that may involve only subsets of the eligible voting members. The computational challenge involves the estimation of the unknown voting probabilities. This translates to an integration problem which may possibly be high-dimensional.

In Section 2, we describe the class of Banzhaf indices and the class of satisfaction indices used to quantify voting power. Through the consideration of historical voting data, these measures are extended to give empirical Banzhaf and empirical satisfaction indices. Although the Banzhaf indices have received greater attention in the social science literature, we argue that the satisfaction indices might be preferred. In Section 3, we describe the statistical model used in the development of both the empirical Banzhaf and empirical satisfaction indices. We also look at the implications of the model. Gibbs sampling and importance sampling approaches are considered in Section 4 in an attempt to solve the associated integration problem. A critical analysis of the various integration methods is provided. In Section 5 we look at an example based on data arising from the Supreme Court of Canada.

## 2. BANZHAF AND SATISFACTION INDICES

Along with the Shapley-Shubik index, Banzhaf indices have enjoyed widespread use in the social sciences. Banzhaf indices have received particular attention in courts of law (see Johnson (1969) and Lucas (1982)) and in discussions involving Canada's constitutional amending formula (see Kilgour (1983), Levesque and Moore (1984) and Heard and Swartz (1997b)).

The Banzhaf index for a voter can be defined as the probability that the voter's decision is pivotal in determining the voting outcome where it is assumed that each combination of "Yes" and "No" votes by the voting members is equally likely. For example, consider a referendum involving 3 voters where a majority (ie. 2 or more votes) is required for the referendum to pass. Then using obvious notation, the first voter's decision is pivotal in the outcomes (YYN), (YNY), (NNY) and (NYN) since a change by the first voter causes a reversal in the voting decision. Therefore the Banzhaf index for the first voter is  $4(1/2)^3 = .5$ .

A variation of the Banzhaf index is the power to initiate index PII for an individual. It can be defined as the conditional probability that the voter's decision is pivotal in determining the voting outcome given that a negative decision has been reached where it is assumed that each combination of "Yes" and "No" votes by the voting members is equally likely. Using the example above, the outcomes (NNN), (YNN), (NYN) and (NNY) each lead to a negative decision and the first voter's decision is pivotal in the cases (NYN) and (NNY). Therefore the PII for the first voter is  $2(1/2)^3/4(1/2)^3 = .5$ . Similarly, the power to prevent index (PPI) for an individual can be defined as the conditional probability that the voter's decision is pivotal in determining

the voting outcome given that a positive decision has been reached where it is again assumed that each combination of “Yes” and “No” votes by the voting members is equally likely. The PII and the PPI can be seen as refinements of the Banzhaf index. They provide information on the power of an individual to initiate and to block a voting decision. The Banzhaf index can be expressed as  $2(PII)(PPI)/(PII + PPI)$ .

Sometimes the Banzhaf index, the PII and the PPI are scaled such that the sum of each index over all voting members is equal to 100. However, in these cases, the attractive probabilistic interpretation is lost. The indices of Coleman (1971) are also seen to be proportional to the Banzhaf index.

The class of satisfaction indices are voting measures used to describe how often the decision of a voting member agrees with the group decision. The satisfaction index for a voter can be defined as the probability that the voter is satisfied where it is assumed that each combination of “Yes” and “No” votes by the voting members is equally likely. For example, consider again a referendum involving 3 voters where a majority is required for the referendum to pass. Then the first voter is satisfied in the outcomes (YYY), (YNY), (YYN), (NNN), (NYN) and (NNY). Therefore the satisfaction index for the first voter is  $6(1/2)^3 = .75$ .

Analogous to the PII and the PPI, we can define the negatively-oriented and positively-oriented satisfaction indices. The negatively-oriented satisfaction index for a voter is the conditional probability that the voter is satisfied given that the voting decision is negative. Similarly, the positively-oriented satisfaction index for a voter is the conditional probability that the voter

is satisfied given that the voting decision is positive. Both of these indices are calculated under the assumption that each combination of “Yes” and “No” votes by the voting members is equally likely.

Now whether one uses the Banzhaf indices, the satisfaction indices or any of the scaled versions, all of these indices share the often unrealistic assumption of equal probability for all voting combinations. This is equivalent to assuming independence amongst the voters and equal probability (ie.,  $1/2$ ) of a “Yes” and “No” vote for each voting member. We relax this assumption allowing the possibility that certain voting members tend to align in their voting patterns. The idea is simple. We estimate the probabilities of each of the  $2^m$  voting combinations. We then calculate empirical versions of the Banzhaf and satisfaction indices by summing over the appropriate voting combinations. In Section 3, we describe the statistical model used to estimate the voting probabilities.

Although the satisfaction indices do not have as wide usage as the Banzhaf indices, they do not suffer from a certain grouping paradox as described in Brams (1975). A difficulty with the empirical Banzhaf indices is that they can sometimes mask power when members form coalitions. For example, consider the case of 5 individuals where a majority vote is required for a decision to pass. Suppose further that the first 4 members are of one mind and agree on almost every issue. In this extreme situation, members 1-4 are omnipotent since their voting outcome determines the decision regardless of the vote of member 5. With a huge number of historical votes, we would have empirical probabilities  $\text{Prob}(YYYYY) + \text{Prob}(YYYYN) +$

$\text{Prob}(NNNNN) + \text{Prob}(NNNNY) \approx 1$ . Hence, individually, members 1-4 are rarely pivotal in the voting decision. Consequently, their empirical Banzhaf indices would be nearly zero even though they are extremely powerful.

As mentioned by a referee, all of the traditional voting measures consider only the possibility of a “Yes” and “No” vote for each voting member and do not include the possibility of abstentions. If a new measure could be developed which takes into account all  $3^m$  possibilities, it may also be possible to estimate these parameters and likewise construct empirical measures.

### 3. THE MODEL

Consider a statistical model involving  $m$  voters. Let the subscript  $i_j = 2$  denote that member  $j$  votes “Yes”, let  $i_j = 1$  denote that member  $j$  votes “No” and let  $p_{i_1 \dots i_m}$  be the corresponding probability for all  $m$  voters. The calculation of the various voting measures from Section 2 assumes that  $p_{i_1 \dots i_m} = 1/2^m$  for all voting combinations  $i_1 \dots i_m$ . We develop a statistical model whereby historical voting data provides an estimate  $\hat{p}_{i_1 \dots i_m}$  of  $p_{i_1 \dots i_m}$ . We then use the estimated probabilities to calculate the various “empirical” voting measures. This provides a framework for informal coalitions that may exist between certain voting members. For example, consider again the case of  $m = 3$  voters where a majority vote is required to pass a referendum. Then the empirical Banzhaf index for the first voter is  $\hat{p}_{221} + \hat{p}_{212} + \hat{p}_{112} + \hat{p}_{121}$  and the empirical satisfaction index for the first voter is  $\hat{p}_{222} + \hat{p}_{212} + \hat{p}_{221} + \hat{p}_{111} + \hat{p}_{121} + \hat{p}_{112}$ .

The data consists of votes taken on  $n$  previous occasions. Let  $\mathbf{x}_i = (x_{i1}, \dots, x_{im})$  denote the

voting result on occasion  $i$  where

$$x_{ij} = \begin{cases} 0 & \text{if individual } j \text{ did not vote on occasion } i \\ 1 & \text{if individual } j \text{ voted "No" on occasion } i \\ 2 & \text{if individual } j \text{ voted "Yes" on occasion } i \end{cases} .$$

We therefore allow historical data to include occasions where all members vote and occasions where only a subset of the members vote. The provision for the latter situation typically results in the consideration of much larger data sets. For example, we consider occasions where certain members are unable to vote and also occasions where members intentionally abstain from voting. We also consider situations where the voting body changes over time. For example, we may begin with  $m_1$  voting members who vote over a period of time. After a while,  $m_2$  of the voting members leave and are replaced by an additional  $m_3$  voting members who then vote over a period of time. In this case,  $m = m_1 + m_3$  as the  $m_2$  voters are still considered in the analysis.

Since the  $2^m$  voting probabilities  $p_{i_1 \dots i_m}$  are defined on the simplex, it is natural to assign a Dirichlet(1, ..., 1) distribution for the prior density  $[\mathbf{p}]$ . This flat proper prior describes our apriori view of the distribution of voting probabilities. It is a widely accepted reference prior and is often used to express ignorance (see Berger (1985)). Note that the use of the prior means  $\hat{p}_{i_1 \dots i_m} = E(p_{i_1 \dots i_m}) = 1/2^m$  for all  $i_1 \dots i_m$  provide estimates that give the original assumptions used in the calculation of the traditional Banzhaf and satisfaction indices.



The conditional distribution of the data  $\mathbf{x}$  given  $\mathbf{p}$  (known as the likelihood) is given by

$$[\mathbf{x} | \mathbf{p}] \propto \prod_{i=1}^n \sum_{S_i} p_{\mathbf{y}} \quad (1)$$

where  $S_i = \{\mathbf{y} = (y_1, \dots, y_m) : y_j = 1, 2 \forall j \text{ where } y_j = x_{ij} \text{ if } x_{ij} \neq 0\}$ .

To better understand the specification of the likelihood, observe first that the product in (1) is taken over the  $n$  voting occasions. In the case of complete data (ie. all members always vote) the likelihood reduces to  $[\mathbf{x} | \mathbf{p}] \propto \prod_{i=1}^n p_{\mathbf{x}_i}$  which is simply a multinomial probability. For incomplete data (ie. some of the  $x_{ij} = 0$ ) we have modelled the outcome of the  $i^{\text{th}}$  voting occasion as the marginal probability of the actual voting members by summing over the indices of the members who did not vote.

Although seemingly intuitive, the specification of the likelihood (1) has interesting consequences. To simplify the presentation, consider the case of  $m = 2$  voters although the discussion readily extends to the general case. We first emphasize that  $p_{ij}$  is the joint probability that member 1 votes  $i$  and member 2 votes  $j$  where members 1 and 2 are possibly sitting in the same room and are possibly exerting some influence on one another. Using the notation  $p_{i*}$  and  $p_{*j}$  we refer to the probabilities that member 1 votes  $i$  and member 2 votes  $j$  respectively where the voting decisions are made in privacy. We therefore think of the subscript  $*$  as an indicator of absence from the voting table. Under independence (ie.  $p_{ij} = p_{i*}p_{*j}$ ), it is immediate that the probability of a vote made in privacy is equal to the sum of the relevant joint probabilities (for

example,  $p_{i*} = p_{i1} + p_{i2}$ ). Therefore, in the case of members who vote separately or in the case of truly independent thinkers, there is no difficulty with the specification of the likelihood (1). When members exert influence on one another, the modelling assumptions  $p_{i*} = p_{i1} + p_{i2}$  and  $p_{*j} = p_{1j} + p_{2j}$  serve as an approximation.

Given the likelihood  $[\mathbf{x} \mid \mathbf{p}]$  in (1) and the prior density  $[\mathbf{p}] \propto 1$ , the Bayesian paradigm states that the posterior density is given by

$$[\mathbf{p} \mid \mathbf{x}] \propto \prod_{i=1}^n \sum_{S_i} p_{\mathbf{y}} \tag{2}$$

where  $\mathbf{p}$  is defined on the simplex of dimension  $2^m - 1$ . From the posterior distribution, we then obtain the posterior means  $\hat{p}_{i_1 \dots i_m} = E(p_{i_1 \dots i_m} \mid x)$  to calculate the various empirical voting measures. Note that in the case of complete data, the posterior reduces to a Dirichlet distribution and it follows that

$$\hat{p}_{i_1 \dots i_m} = \frac{1 + \sum_{j=1}^n I_{i_1 \dots i_m}(\mathbf{x}_j)}{2^m + n}$$

where  $I_{i_1 \dots i_m}$  is an indicator function. Note further that in the absence of historical data, we have  $n = 0$  and therefore  $p_{i_1 \dots i_m} = 1/2^m$  which corresponds to the values used in the traditional calculations.

More generally, the posterior distribution (2) is a mixture of Dirichlets. Letting  $v_i$  denote the number of missing voters on the  $i^{\text{th}}$  voting occasion, the mixture distribution consists of  $2^{\sum_{i=1}^n v_i}$

mixture components. Although the mean vector for a mixture of Dirichlets can be calculated analytically, in many cases the number of mixture components would be too great for even a computer to handle. We therefore turn to approximate solutions to the integration problem. These are taken up in Section 4.

Note that rather than calculating posterior means, one may instead prefer to estimate voting probabilities using posterior modes. With the flat Dirichlet prior, this is equivalent to maximum likelihood estimation. Note that in the case of complete data, maximum likelihood estimation results in voting probabilities that are given by proportions of votes in the respective voting directions. This in turn reduces empirical satisfaction indices to the proportion of time that individuals are satisfied.

In some problems, the nature of the voting issues may suggest a modification to the statistical model. In particular, we may want to impose the following restriction on the voting probabilities:

$$p_{i_1 \dots i_m} = p_{i'_1 \dots i'_m} \quad \text{where } i'_j = \begin{cases} 2 & i_j = 1 \\ 1 & i_j = 2 \end{cases} \quad j = 1, \dots, m. \quad (3)$$

In assumption (3), note that  $i_1 \dots i_m$  and  $i'_1 \dots i'_m$  have the same alignment (ie. a specified subset of voting members agree and the complementary subset disagree). This is a reasonable assumption from the point of view that voting issues are random. For if issue  $A$  is random, then so is  $\bar{A}$ , and we would require that  $p_{i_1 \dots i_m}(A) = p_{i_1 \dots i_m}(\bar{A})$  for all  $i_1 \dots i_m$  which in turn implies (3). Assumption (3) has the effect of reducing the dimension of the parameter space

by half. With the restriction, the posterior distribution retains the same form in (2) except that  $\sum p_{2i_2 \dots i_m} = 1/2$  where the summation is taken over  $i_j = 1, 2$ , and  $j = 2, \dots, m$ . The only time when we do not impose assumption (3) is when the voting outcomes “Yes” and “No” have directional meaning. For example, a vote of “Yes” might always be a vote in favour of conservatism.

Therefore we have developed a method to adapt the traditional Banzhaf and satisfaction indices to empirical indices. Furthermore, our methods permit historical voting data that include cases where only subsets of the voters participate.

#### 4. COMPUTATIONS

We now consider the calculation of the mean vector of the posterior distribution (2). It is well-known that high-dimensional integration can suffer from many pitfalls. We therefore consider various integration techniques as this provides a check on the accuracy of computations.

The technique of importance sampling proceeds by expressing an integral  $I(v)$  as

$$I(v) = \int_S v(\mathbf{p})[\mathbf{p} \mid \mathbf{x}] d\mathbf{p} = \int_S \frac{v(\mathbf{p})[\mathbf{p} \mid \mathbf{x}]}{g(\mathbf{p})} g(\mathbf{p}) d\mathbf{p}$$

where  $S$  is the simplex of dimension  $2^m - 1$ ,  $v = v(\mathbf{p})$  denotes a posterior characteristic of interest and  $g(\mathbf{p})$  is an importance sampling density. We then generate independent variates

$\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(N)}$  from  $g(\mathbf{p})$  and estimate  $I(v)$  by

$$\hat{I}(v) = \frac{1}{N} \sum_{i=1}^N \frac{v(\mathbf{p}^{(i)})[\mathbf{p}^{(i)} | \mathbf{x}]}{g(\mathbf{p}^{(i)})}.$$

It is clear that  $\hat{I}(v)$  is an unbiased and consistent estimate of  $I(v)$  provided that the integral  $\int_S v(\mathbf{p})[\mathbf{p} | \mathbf{x}]^2/g(\mathbf{p}) d\mathbf{p}$  is finite. The strategy in importance sampling is to find an importance sampling density  $g(\mathbf{p})$  which mimics  $[\mathbf{p} | \mathbf{x}]$  and allows efficient generation of variates. The reason that we want  $g(\mathbf{p})$  to mimic  $[\mathbf{p} | \mathbf{x}]$  is so that the variance of  $\hat{I}(v)$  is not too large and convergence to  $I(v)$  is rapid. We then estimate the mean vector of the posterior distribution by  $\hat{I}(\mathbf{p})/\hat{I}(1)$ .

We consider 3 importance sampling schemes. The simplest uses the flat Dirichlet prior  $[\mathbf{p}] \propto 1$  as the importance sampler. Whereas generation from the Dirichlet distribution is straightforward and the approach works well in low dimensions (ie. small  $m$ ), we have found that the method is unreliable in higher dimensions. For example, importance sampling from the prior gave incorrect results in the example of Section 5 where  $m = 9$  and the dimension is  $2^{m-1} - 1 = 255$ . The problem is that the prior does not mimic the posterior surface very well, and as a result, too much sampling is done in regions of low posterior probability. This is a common problem when using priors as importance samplers.

Our second importance sampler is more sophisticated in that more of the posterior is absorbed into the importance sampler. That is, we use a Dirichlet importance sampler where the

Dirichlet parameter corresponding to  $p_{i_1 \dots i_m}$  is  $1 + \sum_{j=1}^n I_{i_1 \dots i_m}(\mathbf{x}_j)$ . Therefore, when most of the  $n$  historical voting occasions involve no absentee voters, the importance sampler will provide a good match to the posterior. Although this importance sampler performs better than the first, it may still suffer from sampling too frequently in regions of low posterior probability.

As a third attempt, we try the standard importance sampling approach based on the multivariate Student distribution. To use Student importance sampling or any of its variants (see Evans and Swartz (1995)), we must first transform the integral to  $\mathcal{R}^{2^m-1}$ . There are a number of ways of doing this including the method described by Evans, Gilula and Guttman (1989). Regardless of the transformation, one will be left with integrals of the form  $\int_{\mathcal{R}^{2^m-1}} f(\mathbf{x}) d\mathbf{x}$  where  $f(\mathbf{x})$  may be an enormously complex (and expensive to evaluate) function. To begin Student importance sampling, the calculation of the Hessian matrix of  $f(\mathbf{x})$  is first required. Since the second derivatives of  $f(\mathbf{x})$  are typically too difficult to express in closed form, it is customary to obtain approximations using first and second differences. This alone requires  $O(2^{2^m+1})$  calls of the function  $f(\mathbf{x})$ . It is because of this problem, the difficulty of high-dimensional optimization in determining the mode of  $f(\mathbf{x})$  and the problem of numerical inversion of large covariance matrices that Student importance sampling is generally considered infeasible for high-dimensional problems. In fact, Student importance sampling fails to work on the example presented in Section 5.

We make a final comment on the various importance sampling techniques which is relevant to the problem at hand. Although the following observation is known to those working in

numerical integration, to our knowledge, it has not been explicitly reported in the literature. The question arises as to when the estimate  $\hat{I}(\mathbf{p})/\hat{I}(1)$  is close enough to the posterior mean  $I(\mathbf{p})/I(1)$ . A naive approach would involve periodically looking at the value  $\hat{I}(\mathbf{p})/\hat{I}(1)$ , and then stopping the simulation when the quantity has stabilized. For this problem, it turns out that such an approach can be misleading. As reported earlier, our importance samplers may sample too frequently in regions of low posterior probability. The result of this is a poor estimate  $\hat{I}(\mathbf{p})/\hat{I}(1)$ , but moreover, this estimate may appear to converge. The reason for this is that even though some generated  $\mathbf{p}^{(i)}$  give atypical values of  $\hat{I}(\mathbf{p})$  and  $\hat{I}(1)$ , the ratio of  $\hat{I}(\mathbf{p})$  to  $\hat{I}(1)$  tends to cancel and provide typical values. As a diagnostic, it is better to look at the quantities  $\hat{I}(\mathbf{p})$  and  $\hat{I}(1)$  individually. This is roughly equivalent to the widely recommended method of looking at the importance sampling weights (see Evans and Swartz (1995)).

We now consider the approximation of the posterior mean vector via Gibbs sampling with data augmentation. Recall that the vector  $\mathbf{x}_i$  expresses the outcome for each voting member on the  $i^{\text{th}}$  voting occasion. In cases where certain members do not vote, we complete the voting record with the latent vector  $\mathbf{z}_i$  where

$$z_{ij} = \begin{cases} 0 & \text{if individual } j \text{ actually voted on occasion } i \\ 1 & \text{if individual } j \text{ would have voted "No" on occasion } i \\ 2 & \text{if individual } j \text{ would have voted "Yes" on occasion } i \end{cases} .$$

Therefore the vector  $\mathbf{x}_i + \mathbf{z}_i$  gives the complete voting record for all members on occasion  $i$

whether or not they had actually voted. The introduction of the latent vector  $\mathbf{z}_i$  in Gibbs sampling is known as data augmentation (see Tanner and Wong (1987) and Gelfand and Smith (1990)).

To obtain the relevant conditional distributions for Gibbs sampling, we write

$$[\mathbf{x}, \mathbf{z}, \mathbf{p}] \propto [\mathbf{x}, \mathbf{z} \mid \mathbf{p}][\mathbf{p}] \propto \prod_{i=1}^n p_{\mathbf{x}_i + \mathbf{z}_i} \propto \prod_{i_1 \dots i_m} p_{\sum_{j=1}^n I_{i_1 \dots i_m}(\mathbf{x}_j + \mathbf{z}_j)}. \quad (4)$$

Using (4), the conditional density  $[\mathbf{p} \mid \mathbf{x}, \mathbf{z}]$  is Dirichlet where the Dirichlet parameter corresponding to  $p_{i_1 \dots i_m}$  is  $1 + \sum_{j=1}^n I_{i_1 \dots i_m}(\mathbf{x}_j + \mathbf{z}_j)$ . The remaining conditional densities  $[\mathbf{z}_i \mid \mathbf{p}, \mathbf{x}, \mathbf{z}_{(i)}]$ ,  $i = 1, \dots, n$  are discrete where  $\mathbf{z}_{(i)}$  denotes all vectors  $\mathbf{z}_1, \dots, \mathbf{z}_n$  except  $\mathbf{z}_i$ . Specifically,

$$\text{Prob}[\mathbf{z}_i = (z_{i1}, \dots, z_{im}) \mid \mathbf{p}, \mathbf{x}, \mathbf{z}_{(i)}] = \frac{p_{z_{i1} \dots z_{im}}}{\sum_{S_i} p_{\mathbf{y}}}$$

where  $z_{ij} = x_{ij}$  if  $x_{ij} \neq 0$  and  $S_i = \{\mathbf{y} = (y_1, \dots, y_m) : y_j = 1, 2 \forall j \text{ where } y_j = x_{ij} \text{ if } x_{ij} \neq 0\}$ . Then beginning with starting values  $\mathbf{p}^{(0)}, \mathbf{z}_1^{(0)}, \dots, \mathbf{z}_n^{(0)}$ , Gibbs sampling proceeds by iteratively sampling from the conditional distributions. After a sufficient period of time, the generated  $\mathbf{p}^{(i)}$  values are averaged to estimate the posterior mean vector. To obtain the posterior standard deviations of the empirical indices, a little more computation is required. At each iteration, we need to calculate the indices based on the generated  $\mathbf{p}^{(i)}$  values. We then obtain the average of the indices  $\hat{E}(I)$  and the average of the squares of the indices  $\hat{E}(I^2)$ . The posterior standard deviations are then approximated by  $(\hat{E}(I^2) - \hat{E}(I)^2)^{1/2}$ .



The method of Gibbs sampling with data augmentation proves to be the most successful of the integration methods considered in this paper. In particular, it is easy to sample from the conditional distributions and the approach gives correct answers for the example presented in Section 5. We are confident in the Gibbs sampling solutions due to the agreement of results based on different starting values for the Gibbs sampling algorithm. We also note the agreement of the Gibbs sampling solutions with those obtained by importance sampling approaches in lower dimensional problems.

## 5. AN EXAMPLE

We now turn to an example concerning the  $m = 9$  judges who currently sit on the Supreme Court of Canada. For a decision to pass, a majority of the judges must vote in favour of the issue. Formally, each judge has equal voting power, and all traditional voting indices highlight this fact. However, it is possible that some judges may be particularly influential and it is possible that informal coalitions may exist between certain judges. From a political science perspective, these are topics of great interest. We gain some insight on these questions using the empirical voting measures developed in this paper.

Data has been collected on the  $n = 421$  Supreme Court decisions rendered in the period 1993 to 1996 where only 169 of the decisions involve complete data (ie. no absentee judges). It is interesting to note that amongst the 421 decisions, 298 of these involve a consensus decision. This indicates an extreme departure from the traditional assumption of equal probability for all voting combinations. The data is readily available from the authors upon request.

We impose restriction (3) as a “Yes” vote has no directional meaning. This requires the calculation of integrals of dimension  $2^{m-1} - 1 = 255$ . We note that for the special structure inherent in this problem (ie.  $m$  odd and a positive decision requiring majority), the symmetry is such that the satisfaction index, the positively-oriented satisfaction index and the negatively-oriented satisfaction index are all equal. Gibbs sampling is successfully carried out using 10,000 iterations.

In Table 1, we compare various measures associated with the Supreme Court data. We observe that although each judge has the same formal power (expressed by the satisfaction index), the reality is that different judges have varying degrees of satisfaction. In particular, it is noteworthy that the 2 female judges, C. L’Heurex-Dubé and B. McLachlin, have the lowest empirical probabilities of satisfaction. That is, they are in agreement with Supreme Court decisions less often than the other 7 judges. We also observe that the empirical satisfaction indices have the same ordering as the proportion of time that each judge votes in the majority. In fact, the empirical satisfaction index can be viewed as a compromise between using no data (ie. the standard satisfaction index) and relying completely on the data (ie. the proportion of time that each judge votes according to the majority). Finally, we note that consensus decisions increase the values of the empirical satisfaction indices. When these “easy” decisions are removed, and the analysis repeated on the 126 non-unanimous decisions, we see that the new empirical indices become smaller but retain their order. A more detailed analysis of the dataset is under preparation in a companion paper.

## REFERENCES

- Banzhaf, J.R. (1965), "Weighted voting doesn't work: a mathematical analysis", *Rutgers Law Review*, 19, 317-343.
- Berger, J.O. (1985), *Statistical Decision Theory and Bayesian Analysis*, Second Edition, Springer-Verlag.
- Brams, S.J. (1975), *Game Theory and Politics*, The Free Press: Macmillan.
- Coleman, J.S. (1971), "Control of collectivities and the power of a collectivity to act", *Social Choice*, eds. Lieberman, B., New York: Gordon and Breach, 269-298.
- Deegan, J. and Packel, E.W. (1978), "A new index of power for simple n-person games", *International Journal of Game Theory*, 7, 113-123.
- Evans, M., Gilula, Z. and Guttman, I. (1989), "Latent class analysis of two-way contingency tables by Bayesian methods", *Biometrika*, 76, 557-563.
- Evans, M. and Swartz, T.B. (1995), "Methods for approximating integrals in Statistics with special emphasis on Bayesian integration problems", *Statistical Science*, 10, 254-272.
- Felsenthal, D.S. and Machover, M. (1995), "Postulates and paradoxes of relative voting power - a critical re-appraisal", *Theory and Decision*, 38, 195-229.
- Gelfand, A.E. and Smith, A.F.M. (1990), "Sampling based approaches to calculating marginal densities", *Journal of the American Statistical Association*, 85, 398-409.

Heard, A.D. and Swartz, T.B. (1997a), "Empirical Banzhaf indices", *Public Choice*, To appear.

Heard, A.D. and Swartz, T.B. (1997b), "The regional veto formula and its effects on Canada's constitutional amendment process", *Canadian Journal of Political Science*, 30, 339-356.

Johnson, R.E. (1969), "An analysis of weighted voting as used in reappointment of county governments in New York state", *Albany Law Review*, 34, 1-45.

Johnston, R.J. (1978), "On the measurement of power: some reactions to Laver", *Environment and Planning A*, 10, 907-914.

Kilgour, D.M. (1983), "A formal analysis of the amending formula of Canada's Constitution Act", *Canadian Journal of Political Science*, 16, 771-777.

Levesque, T.J. and Moore, J.W. (1984), "Citizen and provincial power under alternative amending formulae: an extension of Kilgour's analysis", *Canadian Journal of Political Science*, 17, 157-166.

Lucas, W.F. (1982), "Measuring power in weighted voting systems", *Political and Related Models (Vol. 2 in series Models in Applied Mathematics)*, eds. Brams, S.J., Lucas, W.F., and Straffin, P.D., New York: Springer, 183-238.

Shapley, L.S. and Shubik, M. (1954), "A method for evaluating the distribution of power in a committee system", *American Political Science Review*, 48, 787-792.

Straffin, P.D. Jr. (1978), "Probability models for power indices", *Game Theory and Political Science*, ed. P.C. Ordeshook, New York: New York University Press, 477-510.

Tanner, M.A. and Wong, W.H. (1987), "The calculation of posterior distributions by data augmentation", *Journal of the American Statistical Association*, 82, 528-540.

Table 1: Satisfaction indices (S), empirical satisfaction indices (ES), the proportion of decisions satisfied (PS) and empirical satisfaction indices based on the restricted data set of 126 non-unanimous decisions (ESR) for the 9 Supreme Court judges. Posterior standard deviations are given in parentheses.

Judge	S	ES	PS	ESR
A. Lamer	.637	.812 (.016)	.933	.670 (.026)
G. La Forest	.637	.812 (.016)	.929	.669 (.025)
C. L'Heureux-Dubé	.637	.750 (.018)	.807	.572 (.026)
J. Sopinka	.637	.807 (.016)	.915	.657 (.025)
C. Gonthier	.637	.808 (.016)	.923	.659 (.025)
P. Cory	.637	.837 (.015)	.964	.711 (.024)
B. McLachlin	.637	.774 (.017)	.873	.600 (.027)
F. Iacobucci	.637	.833 (.015)	.955	.702 (.025)
J. Major	.637	.802 (.016)	.903	.650 (.026)