

Some Aspects of the Sampling Distribution of the Apportionment Index and Related Inference

Laura Koskela, Bikas K. Sinha and Tapio Nummi

Koskela, L., Sinha, B.K. & Nummi, T. 2007. Some aspects of the sampling distribution of the Apportionment Index and related inference. *Silva Fennica* 41(4): 699–715.

As customer-oriented production strategies have gained ground in the sawmill industry, proper measurement of the fit between the log demand and log output distributions has become of crucial importance. The prevailing means of measuring the outcome is the so-called Apportionment Index (*AI*), which is calculated from the relative proportions of the observed and required distributions. Although some statistical properties of the *AI* have recently been examined and alternative means of measuring the bucking outcome have been suggested, properties of the sampling distribution of the *AI* have not yet been widely studied. In this article we examine the asymptotic sampling distribution for the *AI* by assuming a multinomial distribution for the outcome. Our results are based mainly on large-sample normal approximations. Also some studies of the determination of the number of logs needed to obtain a specified level of accuracy of the *AI* have been carried out.

Keywords bucking, dissimilarity index, harvesting, multinomial distribution, overlapping coefficient

Addresses *Koskela & Nummi*: University of Tampere, Dept of Mathematics, Statistics and Philosophy, FI-33014 University of Tampere, Finland; *Sinha*: Indian Statistical Institute, Kolkata, India

E-mail laura.koskela@uta.fi

Received 19 March 2007 **Revised** 13 August 2007 **Accepted** 25 September 2007

Available at <http://www.metla.fi/silvafennica/full/sf41/sf414699.pdf>

1 Introduction

The prevailing Nordic harvesting system is to convert tree stems into smaller logs immediately at the harvesting site. Modern cut-to-length (CTL) harvesters are equipped with high-class measuring and computing systems which make

computer-based bucking optimization possible in harvesters. After crosscutting the harvester's computer updates the so-called cumulative outcome matrix which keeps a record of the logs produced.

The common trend in the sawmill industry, at least in Scandinavia, is towards customer-ori-

ented production of well-defined products. It has become more and more important not only to supply the sawmill with a sufficient number of logs at minimum cost, but also to ensure that the raw material meets the requirements of the sawmill as regards to length, diameter and quality distribution of logs (Kivinen 2004). This in turn, has made proper assessment of the goodness of the bucking outcome of crucial importance.

In general, as listed by Kivinen et al. (2005), there are two main situations where the agreement between the distribution of logs demanded by the sawmill (demand distribution) and the actual outcome (output) distribution of logs is of particular interest. These are 1) the standard pre-harvest planning procedure where bucking simulators are commonly used to determine the best available stands for prevailing customer orders, and 2) the postharvest analysis where it may be desirable to know, for example, how various harvesters have succeeded in meeting a certain demand distribution or to determine whether there are any significant differences between various wood suppliers. A proper measure for evaluating the bucking outcome also provides information on how to adjust the bucking instructions to meet the desired log distribution. According to Vuorenpää et al. (1997) the large sawmills should monitor the progress of the work of the harvesting machines on the daily basis to assure that the requirements of the sawmill are adequately satisfied.

The most widely used measure for assessing the agreement between the log demand and the log output distributions in Scandinavia is *Apportionment Index (AI)* or *Apportionment Degree* (see Section 2.1 for mathematical formulation). The *AI* measures the output of the actual harvesting operation by comparing the relative proportions of the outcome and target distributions. The Apportionment Index has gained ground especially by merit of its simplicity, ready interpretability and ease of use. The measure has been criticized mainly as being too crude, since, for example, it attributes the same weight to all log classes. The price-weighted version of the *AI* was proposed by Kivinen et al. (2005) and Nummi et al. (2005). Penalty-based variants of the traditional measure are suggested in papers by Kirkkala et al. (2000), Weijo (2000) and Malinen and Palander (2004). Sinha et al. (2005b) generalized the measure and

introduced the family of Apportionment Indices. However, the Apportionment Index and its derivatives have been used mainly as a descriptive device and very little is known of their statistical properties. While there have been attempts to understand this index (Nummi et al. 2005 and Sinha et al. 2005a), a serious theoretical foundation is still lacking. For a more comprehensive introduction to the topic we refer to Koskela (2007).

Kivinen et al. (2005) defined four criteria for an ideal measure of bucking outcome. According to the first two criteria an ideal measure should 1) take into account the size of the stands to be compared in terms of the number of logs (i.e., given a common log demand distribution, the results from two stands with different numbers of logs should be commensurable) and 2) make it possible to compare results based on the demand matrices of different sizes (i.e., a different number of diameter or length classes or both). These criteria address the problems often encountered by forest managers in practice when determining a suitable stand or group of stands for the given demand distribution. It was concluded by the authors that the Apportionment Index, at least in principle, takes into account the stand size expressed in terms of the number of logs. This is because the *AI* apply proportional quantities of logs in each diameter-length combination. The issue related to handle the potentially confounding effect caused by different matrix sizes is, however, much more complicated.

Kivinen et al. (2005) also conducted simulation studies to test the Apportionment Index and three other goodness-of-fit measures against the given criteria. The study showed that stands with a small number of stems or a large number of small-sized stems were likely to match the log demand distribution more poorly than stands with a large number of stems and/or a wide diameter at breast height (DBH). This behaviour can be easily understood by noting that in stands with few trees there basically are fewer opportunities to produce logs in all log categories than in stands with many trees. In some situations we may only be interested in the proportion of "correctly" located logs in the outcome distribution with respect to the demanded log distribution without taking into account any underlying conditions

such as the stand size (cf. interpretation of the *AI* in Section 2.2). However, there may also exist situations where one is not only interested in the magnitude of the value of the measure itself but also wants to know whether the performance of the harvesting operation is tolerable with respect to some underlying conditions. For example, for a smaller number of logs harvested a lower value of the *AI* may not necessarily indicate a poorly performing harvesting operation. This then raises a question of the least tolerable value of the *AI*. This information is important especially for forest managers.

In this paper we examine some aspects of the asymptotic sampling distribution of the Apportionment Index by assuming a multinomial distribution for the outcome (see also Ransom 2000). In Section 3 we derive the approximate first two moments, i.e. the mean and the variance, of the *AI* assuming multinomial distribution for the bucking outcome and using normal approximations. In Section 4 the derivated moments are utilized to construct a lower tolerance limit for the *AI* and simulations are then carried out to evaluate the goodness of the approximations. Some studies on the determination of the number of logs needed to attain high apportionment with a given accuracy are described in Section 5. The section illustrates the effect of certain underlying conditions such as the dimension and the form of the target distribution on the value of the *AI*.

2 Measuring Bucking Outcome by the Apportionment Index

2.1 Mathematical Formulation of the Apportionment Index

The output of the actual harvesting operation has been measured mainly by comparing the relative proportions of the output and target distributions. More specifically, let

$$\mathbf{T} = \begin{bmatrix} t_{11} & t_{12} & \dots & t_{1n} \\ t_{21} & t_{22} & \dots & t_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ t_{m1} & t_{m2} & \dots & t_{mn} \end{bmatrix} = (t_{ij})$$

denote the $m \times n$ demand (target) matrix (or table) for a certain log type, where each row represents a particular small end diameter (SED) class of logs, each column refers to a particular length class and t_{ij} is the number of logs in the i th diameter class and j th length class, $i = 1, \dots, m$ and $j = 1, \dots, n$. A log with an SED of d and a length of l will belong to the log class (i, j) if the log satisfies the constraints $d_i \leq d < d_{i+1}$ and $l_j \leq l < l_{j+1}$. Correspondingly, $m \times n$ matrix $\mathbf{O} = (o_{ij})$ is used for the outcome of the harvesting operation. Matrices may be converted into vectors by arranging the columns of a matrix underneath each other. In what follows we consider the vectorized forms $\mathbf{t} = (t_{11}, t_{21}, \dots, t_{mn})' = (t_1, t_2, \dots, t_k)'$ and $\mathbf{o} = (o_{11}, o_{21}, \dots, o_{mn})' = (o_1, o_2, \dots, o_k)'$ of the demand and outcome matrices, respectively.

A common practice in Scandinavia is to evaluate the fit between the demand and actual output log distributions using the Apportionment Index first introduced in forestry by Bergstrand in the mid-1980s (e.g. Bergstrand 1989). For a fixed quality class the *AI* is defined as

$$AI = 1 - 0.5 \times \sum_{i=1}^k |o_i^* - t_i^*| \tag{1}$$

where $o_i^* = o_i / \sum_{j=1}^k o_j$ and $t_i^* = t_i / \sum_{j=1}^k t_j$ are the relative proportions of the outcome and target matrices, respectively. After some simple manipulations it can be shown that the *AI* can be rewritten as

$$AI = \sum_{i=1}^k \min(o_i^*, t_i^*) \tag{2}$$

The maximum value of the *AI* is 1 (100%), which indicates a perfect match between the distributions. The minimum value of the index is $\min(t_1^*, t_2^*, \dots, t_k^*)$, i.e. the smallest relative cell target, which is reached when all the logs fall into the diameter-length class of the smallest target proportion. The *AI* may be interpreted as the proportion of “correctly” located logs in the outcome distribution with respect to the demanded log distribution. For example, if the *AI* value were 0.85, this would mean that 85% of the produced logs are in accordance with the demanded distribution while 15% are of the wrong size and should have been allotted to other log categories during the bucking process to make the outcome equal to the

target, i.e. to attain complete agreement between the two distributions.

2.2 Some Related Measures

The *overlapping coefficient (OVL)* is defined as a measure of agreement or similarity between two probability distributions or two populations represented by such distributions (Inman and Bradley 1989). In a simple univariate case the *OVL* is simply the fraction of the probability mass common to both distributions. In a case of two discrete probability distributions, the *OVL* and the *AI* have the same formula. Both measures are also related to the so-called *Dissimilarity Index (DI)* or *Index of Dissimilarity*, which is commonly used e.g. to summarize the closeness of fit of a model to the categorical sample data (e.g. Agresti 2002, p. 329–330). One version of the *DI* has gained ground especially in sociology, where it has become the most common measure of social segregation. In fact, it is easy to see that $OVL = 1 - DI = AI$ for two discrete probability distributions.

The *OVL* has been proposed as a generalized measure of agreement. Obviously, however, the *Dissimilarity Index* appeared first. One of the very first instances of the *DI* as a measure of segregation was that of Jahn et al. (1947), and after Duncan and Duncan's (1955) methodological paper the index started to attain popularity among sociologists. Research characterizing the properties of the sampling distributions of the segregation indices, including the *Dissimilarity Index*, is rare. One important exception, however, is that provided by Ransom (2000), who examines the sampling distribution of the *DI*. He studies the segregation by sex across different occupations and derives the asymptotic sampling distribution for the *Dissimilarity Index* based on the statistical model provided by the multinomial distribution. Another interesting article is that by Inman and Bradley (1991), where approximations to the mean and variance of the index are given under a multivariate hypergeometric distribution for the cell frequencies in the context of a $2 \times k$ cross-classification table subject to fixed row and column totals. However, since the target distribution is a fixed quantity, neither of these

two considerations can be directly applied to the *Apportionment Index*.

3 Mean and Variance of the Apportionment Index

Appendix A includes some preliminary results used for the mathematical derivations in this section. We assume that the observed distribution (outcome) of logs follows the multinomial distribution

$$(o_1, o_2, \dots, o_k) \sim MN(N; t_1^*, t_2^*, \dots, t_k^*) \tag{3}$$

where

(o_1, o_2, \dots, o_k) is the random outcome,

$N = \sum_{i=1}^k o_i$ is the total number of logs harvested,

t_i^* s are the relative target values with $t_i^* \in (0, 1)$ and $\sum_{i=1}^k t_i^* = 1$.

The distribution assumption may not necessarily make sense if a stand is tried to harvest by a target which is unsuitable for the particular stand. It is, for example, impossible to produce logs with large SED from trees with smaller DBH. A situation of this kind immediately leads to a poor fit between the outcome and demand distribution arguing that the outcome cannot be a realization from the given multinomial distribution. However, if the pre-harvest planning procedure is performing appropriately, it feels reasonable that the selection of the harvested stand is done such that the production of the requested logs is possible.

For the relative random outcome we use the notation $\mathbf{o}^* = (o_1^*, \dots, o_k^*)$, where $o_i^* = o_i / N \in (0, 1)$ and $\sum_{i=1}^k o_i^* = 1$. It follows from Eq. 3 that

$$o_i \sim Bin(N, t_i^*) \tag{4}$$

and further

$$E(o_i^*) = E\left(\frac{o_i}{N}\right) = t_i^*,$$

$$V(o_i^*) = V\left(\frac{o_i}{N}\right) = \frac{t_i^*(1-t_i^*)}{N} = \sigma_i^2 \tag{5}$$

and

$$\text{Corr}(o_i^*, o_j^*) = -\sqrt{\frac{t_i^* t_j^*}{(1-t_i^*)(1-t_j^*)}} = \rho_{ij} \tag{6}$$

Under certain conditions, the binomial distribution can be approximated by the normal distribution. A conservative rule to follow in the case of $X \sim \text{Bin}(n, p)$ is that $\min(np, n(1-p)) \geq 5$ (see e.g. Casella and Berger 2002, p. 104–105). In terms of target values this means that e.g. for $N = 500$ the cell target t_i^* should lie in the interval $[0.01, 0.99]$ and for $N = 1000$ it should belong to $[0.005, 0.995]$. Since in practice the total number of logs (N) is usually large, the conditions will be satisfied even for quite small target values. The normal approximation gives

$$\frac{\sqrt{N}(o_i^* - t_i^*)}{\sqrt{t_i^*(1-t_i^*)}} \sim N(0,1) \tag{7}$$

Recall the definition of the Apportionment Index in Eq. 1. Before deriving the moments of the AI , we define

$$\widetilde{AI} = 2(1 - AI) = \sum_{i=1}^k |o_i^* - t_i^*| \tag{8}$$

If we assume normal distribution (Eq. 7), upon application of Result 1 in Appendix A, $E(\widetilde{AI})$ becomes

$$\begin{aligned} E(\widetilde{AI}) &= \sum_{i=1}^k E(|o_i^* - t_i^*|) = \sum_{i=1}^k \sigma_i \sqrt{\frac{2}{\pi}} \\ &= \sqrt{\frac{2}{N\pi}} \sum_{i=1}^k \sqrt{t_i^*(1-t_i^*)} \end{aligned} \tag{9}$$

Hence,

$$E(AI) = 1 - \frac{1}{2} E(\widetilde{AI}) = 1 - \sqrt{\frac{1}{2N\pi}} \sum_{i=1}^k \sqrt{t_i^*(1-t_i^*)} \tag{10}$$

It can easily be seen that $E(AI)$ increases as N increases and $E(AI) \rightarrow 1$ as $N \rightarrow \infty$. Thus, under the assumption of multinomial distribution, index values close to one are expected for very large N . Applying Result 1 and Corollary 2 in Appendix A, $V(\widetilde{AI})$ becomes

$$\begin{aligned} V(\widetilde{AI}) &= \sum_{i=1}^k V(|o_i^* - t_i^*|) + \sum \sum_{i \neq j} \text{Cov}(|o_i^* - t_i^*|, |o_j^* - t_j^*|) \\ &= \sum_{i=1}^k \sigma_i^2 \left(1 - \frac{2}{\pi}\right) + \sum \sum_{i \neq j} \sigma_i \sigma_j \left(\psi(\rho_{ij}) - \frac{2}{\pi}\right) \\ &= \sum_{i=1}^k \sigma_i^2 - \frac{2}{\pi} \left(\sum_{i=1}^k \sigma_i\right)^2 + \sum \sum_{i \neq j} \sigma_i \sigma_j \psi(\rho_{ij}) \\ &= \sum_{i=1}^k \frac{t_i^*(1-t_i^*)}{N} - \frac{2}{\pi} \left(\sum_{i=1}^k \sqrt{\frac{t_i^*(1-t_i^*)}{N}}\right)^2 \\ &\quad + \sum \sum_{i \neq j} \frac{\sqrt{t_i^* t_j^* (1-t_i^*)(1-t_j^*)}}{N} \psi\left(\sqrt{\frac{t_i^* t_j^*}{(1-t_i^*)(1-t_j^*)}}\right) \end{aligned} \tag{11}$$

with $\psi(\rho) = \frac{2}{\pi} |\rho| \arcsin |\rho| + \frac{2}{\pi} \sqrt{1-\rho^2}$

(see Result 4 and Table 9 in Appendix A). Hence,

$$\begin{aligned} V(AI) &= \frac{1}{4} V(\widetilde{AI}) \\ &= \sum_{i=1}^k \frac{t_i^*(1-t_i^*)}{4N} - \frac{1}{2\pi} \left(\sum_{i=1}^k \sqrt{\frac{t_i^*(1-t_i^*)}{N}}\right)^2 \\ &\quad + \sum \sum_{i \neq j} \frac{\sqrt{t_i^* t_j^* (1-t_i^*)(1-t_j^*)}}{4N} \psi\left(\sqrt{\frac{t_i^* t_j^*}{(1-t_i^*)(1-t_j^*)}}\right) \end{aligned} \tag{12}$$

As seen in Eq. 12, the variance of the AI is a decreasing function of N and $V(AI) \rightarrow 0$ as $N \rightarrow \infty$, i.e. very little variation is expected for a particularly large N . In a special case of equal target proportions, i.e. $\mathbf{t}^* = (1/k, \dots, 1/k)$, $E(AI)$ and $V(AI)$ reduce to

$$E(AI) = 1 - \sqrt{\frac{1}{2\pi N}} (k-1) \tag{13}$$

and

$$\begin{aligned} V(AI) &= \frac{k-1}{4\pi kN} [(\pi-2) + (k-1)\{\pi\psi(\rho) - 2\}] \\ &= \frac{k-1}{4\pi kN} \left[(\pi-2) + (k-1) \left\{ \pi\psi\left(\frac{1}{k-1}\right) - 2 \right\} \right] \end{aligned}$$

respectively.

4 Computational Examples

4.1 The AI and Its Lower Tolerance Limit for a Real Target Matrix

The target matrix of Nummi et al. (2005) and a modification of the outcome distribution presented in the same article are used here. The target matrix **T** and the actual outcome sample matrix **O** are given in Table 1 and Table 2, respectively. For these matrices, calculation gives $AI = 0.962$, which indicates that 96.2% of the produced logs are in accordance with the target, while 3.8% should have been located in other log categories during the bucking operation. Taking a closer look at the matrices it can be seen that 13 out of 25 log categories in the outcome matrix are overloaded by a total of 38 logs if compared to the target. These 38 logs are those which should have been located during the bucking process in those 10 log categories with shortfall. There are only 10 underloaded log categories, since 2 out of 25 categories exhibit a perfect match between the target and the outcome. See Remark 1 below.

Using standard theory we may deduce that

$$\frac{AI - E(AI)}{\sqrt{V(AI)}} \sim N(0,1), \tag{14}$$

for a large number of logs, N , and for a moderate number of log categories, k . $E(AI)$ and $V(AI)$ are given in Eq. 10 and Eq. 12, respectively. From Eq. 14 it follows that

$$1 - \alpha = P\left(-z_\alpha \leq \frac{AI - E(AI)}{\sqrt{V(AI)}}\right) \tag{15}$$

$$= P\left(AI \geq E(AI) - z_\alpha \sqrt{V(AI)}\right)$$

giving the approximate $100 \times (1-\alpha)\%$ lower tolerance limit for the AI as

$$E(AI) - z_\alpha \sqrt{V(AI)} \tag{16}$$

For the output and target matrices to be considered similar, the lower tolerance limit serves as the least acceptable value of the AI . Thus, in a case where the observed outcome matrix assigns the index value which falls below Eq. 16, we would consider the outcome to indicate disagreement with the given target matrix. Overall, the larger the observed AI , the better the agreement. We hence suggest values higher than the mean ($E(AI)$) to indicate a satisfactory level of agreement. It is worth noting, especially for practical implementation, that $E(AI)$ is very easy to compute. For $N = 1000$ and the above target matrix, $E(AI) = 0.939$ and $V(AI) = 0.9025 \cdot 10^{-4}$. Hence, the 95% lower tolerance limit is 0.923. For the sample outcome

Table 1. Target matrix.

Top diam (mm)	Length (cm)					Total
	430	460	490	520	550	
160	28	16	58	45	45	192
200	37	17	65	45	37	201
240	17	49	37	44	55	202
280	22	39	39	44	59	203
340	19	30	47	54	52	202
Total	123	151	246	232	248	1000

Table 2. Outcome matrix.

Top diam (mm)	Length (cm)					Total
	430	460	490	520	550	
160	27	14	62	46	48	197
200	39	13	72	46	32	202
240	14	46	35	45	58	198
280	16	40	39	42	66	203
340	9	35	49	55	52	200
Total	105	148	257	234	256	1000

Table 3. The frequency distribution of the amount of overloads and underloads.

x_i	overload	f_i	underload
1	5	1	
2	2	3	
3	2	2	
4	1	1	
5	1	1	
6	–	1	
7	2	–	
8	–	–	
9	–	–	
10	–	1	
	13	10	

matrix, however, $AI = 0.962$. This far exceeds the mean 0.939 and we may hence conclude that the agreement in this case is highly satisfactory. Note that both the expected value and the lower tolerance limit take into account the effect of the number of logs. This may give valuable information e.g. when results of different harvesting operations with different stand sizes and targets are compared.

Remark 1. We propose to take a closer look at the two matrices in Tables 1 and 2. As noted above, there are 13 (respective 10) log categories which indicate overload (respective underload) while 2 are “neutral”. We now examine the frequency distribution of the amount of overloads and underloads in this example. Both these distributions are shown in Table 3. As expected, $\sum x_i f_i = 38$ for both frequency distributions.

The computation of the AI simply depends on the total amount of overloads (or underloads) and not on their specific frequency distribution. However, from a practical point of view, the distortion or deviation from the target matrix in terms of the overload or underload distribution will have an impact on the overall cost-effectiveness of the sample output matrix. This calls for a price-weighted index of agreement (see Kivinen et al. 2005 and Nummi et al. 2005). For uniform price policy (only of theoretical interest), possibly concentration of overload or underload in a few categories would be desirable. It may, for example, be preferable to have overload in categories of larger diameter-length combinations, as large logs

may possibly be converted into logs of smaller dimensions, if required.

4.2 A Simulation Study for the Tolerance Limit

A simulation study was conducted to investigate the accuracy of the confidence level $(1-\alpha)$ of the approximate tolerance limit (TL) given in Eq. 16. However, we carried out this study with reference to both-sided tolerance limits mainly to understand the nature of normal approximation. The approximate $100(1-\alpha)\%$ tolerance limits used in this simulation study are defined as

$$E(AI) \pm z_{\alpha/2} \sqrt{V(AI)} \quad (17)$$

We simulated 10000 outcome matrices from $MN(N; t_1^*, \dots, t_k^*)$, with the given target probabilities (Table 1) and $N = 50, 100, 500, 1000, 5000, 10000, 50000, 100000$. The index value was computed for each simulated outcome (with respect to the given target matrix) and $100(1-\alpha)\%$ TLs were constructed using Eq. 17. The confidence levels (CL) of the approximate TLs were evaluated by calculating the proportion of the times AI falling within the respective tolerance limits. We also report the proportions of times the AI takes a value below and above the lower and upper tolerance limits. These are denoted in Table 4 by LP and RP, respectively. In Table 5 $E(AI)$ and $V(AI)$ are computed for all N and the given target using Eq. 10 and Eq. 12, respectively. We also report the sample point estimates of $E(AI)$ and $V(AI)$ calculated from the simulations.

Tables 4 and 5 show that the approximation works reasonably well. The CL values are in accordance with the respective confidence levels and the sample point estimates match nicely with $E(AI)$ and $V(AI)$. However, comparison of the LP and RP values in Table 4 reveals that in most of the cases LP is slightly greater than RP, which indicates some degree of non-symmetry in the distribution of the AI . Table 5 also illustrates the performance of $E(AI)$ and $V(AI)$ as a function of N . It can be seen that a relatively large N (≥ 500) is needed to obtain $E(AI) \geq 0.914$. With respect to the given 5×5 target distribution, $N \geq 368$ seems to be sufficient for $E(AI) > 0.90$.

Table 4. The approximated confidence levels for the approximative tolerance limits based on 10 000 simulations and $N = 50, 100, 500, 1000, 5000, 10000, 50000, 100000$.

N	$1-\alpha$						$1-\alpha$					
	CL	0.900 LP	RP	CL	0.950 LP	RP	CL	0.990 LP	RP	CL	0.999 LP	RP
50	0.9109	0.0495	0.0396	0.9575	0.0263	0.0162	0.9935	0.0049	0.0016	0.9992	0.0008	0.0000
100	0.9132	0.0487	0.0381	0.9589	0.0274	0.0137	0.9920	0.0061	0.0019	0.9993	0.0006	0.0001
500	0.8974	0.0606	0.0420	0.9511	0.0300	0.0189	0.9906	0.0071	0.0023	0.9987	0.0011	0.0002
1000	0.8958	0.0511	0.0531	0.9492	0.0271	0.0237	0.9914	0.0061	0.0025	0.9988	0.0012	0.0000
5000	0.9050	0.0526	0.0424	0.9542	0.0287	0.0171	0.9900	0.0073	0.0027	0.9986	0.0013	0.0001
10000	0.9035	0.0515	0.0450	0.9464	0.0321	0.0215	0.9905	0.0063	0.0032	0.9990	0.0010	0.0000
50000	0.9013	0.0590	0.0397	0.9498	0.0341	0.0161	0.9894	0.0090	0.0016	0.9980	0.0019	0.0001
100000	0.8963	0.0578	0.0459	0.9491	0.0306	0.0203	0.9894	0.0083	0.0023	0.9989	0.0011	0.0000

Table 5. $E(AI)$, $V(AI)$ and their sample point estimates $avg(AI)$ and $s^2(AI)$ based on 10 000 simulations and $N = 50, 100, 500, 1000, 5000, 10000, 50000, 100000$.

N	$E(AI)$	$avg(AI)$	$V(AI)$	$s^2(AI)$
50	0.729	0.729	$1.805 \cdot 10^{-3}$	$1.684 \cdot 10^{-3}$
100	0.808	0.808	$9.025 \cdot 10^{-4}$	$8.362 \cdot 10^{-4}$
500	0.914	0.914	$1.805 \cdot 10^{-4}$	$1.829 \cdot 10^{-4}$
1000	0.939	0.940	$9.025 \cdot 10^{-5}$	$9.049 \cdot 10^{-5}$
5000	0.972	0.973	$1.805 \cdot 10^{-5}$	$1.798 \cdot 10^{-5}$
10000	0.981	0.981	$9.025 \cdot 10^{-6}$	$8.980 \cdot 10^{-6}$
50000	0.991	0.991	$1.806 \cdot 10^{-6}$	$1.837 \cdot 10^{-6}$
100000	0.994	0.994	$9.025 \cdot 10^{-7}$	$9.133 \cdot 10^{-7}$

5 Determination of N to Attain High Apportionment with a Given Accuracy

For very large N , index values close to 1 are expected. In this section we illustrate the effect of certain underlying conditions such as the dimension and the form of the target distribution by determining the size of N required to attain a high value of the index with a chosen level of confidence. In other words, we determine N , say N_0 , such that for a given $\epsilon (> 0)$, relative target matrix and α ,

$$P(AI > 1 - \epsilon) \geq 1 - \alpha \tag{18}$$

which is equivalent to

$$P(\widetilde{AI} < 2\epsilon) \geq 1 - \alpha \tag{19}$$

where \widetilde{AI} is as defined in Eq. 8. Assuming Eq. 14 and a given relative demand distribution, N_0

is solved from $z = z_\alpha$, where

$$z = \frac{2\epsilon - E(\widetilde{AI})}{\sqrt{V(\widetilde{AI})}} \sim N(0,1) \tag{20}$$

z_α is the upper α 100% point of $N(0,1)$ distribution, and $E(\widetilde{AI})$ and $V(\widetilde{AI})$ are as defined in Eq. 9 and Eq. 11, respectively. This yields

$$N_0 = \frac{\left(z_\alpha \sqrt{\Delta(\mathbf{t}^*)} + \sqrt{\frac{2}{\pi}} \Gamma(\mathbf{t}^*) \right)^2}{4\epsilon^2} \tag{21}$$

where $\mathbf{t}^* = (t_1^*, t_2^*, \dots, t_k^*)$ is the relative target,

$$\begin{aligned} \Delta(\mathbf{t}^*) = & \frac{1}{4} \sum_{i=1}^k t_i^* (1 - t_i^*) - \frac{1}{2\pi} \left(\sum_{i=1}^k \sqrt{t_i^* (1 - t_i^*)} \right)^2 \\ & + \frac{1}{4} \sum_{i \neq j} \sqrt{t_i^* t_j^* (1 - t_i^*) (1 - t_j^*)} \psi \left(\sqrt{\frac{t_i^* t_j^*}{(1 - t_i^*) (1 - t_j^*)}} \right) \end{aligned} \tag{22}$$

and

$$\Gamma(\mathbf{t}^*) = \sum_{i=1}^k \sqrt{t_i^* (1-t_i^*)} \tag{23}$$

Note that $\Delta(\mathbf{t}^*)$ and $\Gamma(\mathbf{t}^*)$ are proportional to $V(\widetilde{AI})$ and $E(\widetilde{AI})$, respectively.

We investigated three illustrative examples. In the first N_0 is solved for the 5×5 target distribution given in Table 1. In the sequel we call this target distribution the referred target matrix. In the second case the dependence of N_0 on the dimension of the target matrix is examined by solving N_0 for a k -dimensional relative target matrix $\mathbf{t}^* = (1/k, 1/k, \dots, 1/k)$ as k varies from 20 to 30. The third example investigates the effect of the form of the target matrix on the magnitude of N_0 .

Example 1: The referred 5×5 target distribution

Assume the target distribution in Table 1. Table 6 displays N_0 solved for some given values of ϵ and α .

The above results and Eq. 21 show that the number of logs N_0 depends on both the specified accuracy ϵ and the specified confidence level α . Table 6 shows for $\alpha = 0.05$, for example, that while reducing ϵ from 0.1 to 0.01, N_0 must be increased from 582 to 58186. For $\alpha = 0.01$ and the same decrease in ϵ , the number of logs needs to be increased from 685 to 68482. For both confidence levels a more than 100 times larger N is required for $\epsilon = 0.01$ than for $\epsilon = 0.1$. Hence, a decrease in confidence level α needs to be followed by

an increase in N to attain the specified accuracy. Similarly, a decrease in ϵ , which determines the accuracy, requires an increase in N to attain the specified confidence.

Example 2: Target with equal probabilities

Assume the relative demand distribution $\mathbf{t}^* = (1/k, 1/k, \dots, 1/k)$. Table 7 displays N_0 solved for $\alpha = 0.05$ and some selected values of ϵ as k varies from 20 to 30. The bolded row corresponds to the 5×5 matrix. The results with $\alpha = 0.01$ are shown in Table 10 in Appendix B.

Table 7 and also Table 10 in Appendix B confirm the intuitively obvious result that higher dimensional targets require a larger N to maintain the specified accuracy and/or the specified

Table 6. N_0 solved for the referred 5×5 target matrix and some given values of ϵ and α .

ϵ	$1-\epsilon$	α	
		0.01	0.05
		N_0	
0.01	0.99	68482	58186
0.02	0.98	17121	14547
0.03	0.97	7609	6465
0.04	0.96	4280	3637
0.05	0.95	2739	2327
0.06	0.94	1902	1616
0.07	0.93	1398	1187
0.08	0.92	1070	909
0.09	0.91	845	718
0.10	0.90	685	582

Table 7. N_0 solved for $\alpha = 0.05$ and some selected values of ϵ as k varies from 20 to 30.

k	ϵ										
	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	
20	49871	12468	5541	3117	1995	1385	1018	779	616	499	
21	51912	12978	5768	3245	2076	1442	1059	811	641	519	
22	53943	13486	5994	3371	2158	1498	1101	843	666	539	
23	55962	13991	6218	3498	2238	1555	1142	874	691	560	
24	57972	14493	6441	3623	2319	1610	1183	906	716	580	
25	59973	14993	6664	3748	2399	1666	1224	937	740	600	
26	61966	15491	6885	3873	2479	1721	1265	968	765	620	
27	63950	15988	7106	3997	2558	1776	1305	999	790	640	
28	65927	16482	7325	4120	2637	1831	1345	1030	814	659	
29	67897	16974	7544	4244	2716	1886	1386	1061	838	679	
30	69860	17465	7762	4366	2794	1941	1426	1092	862	699	

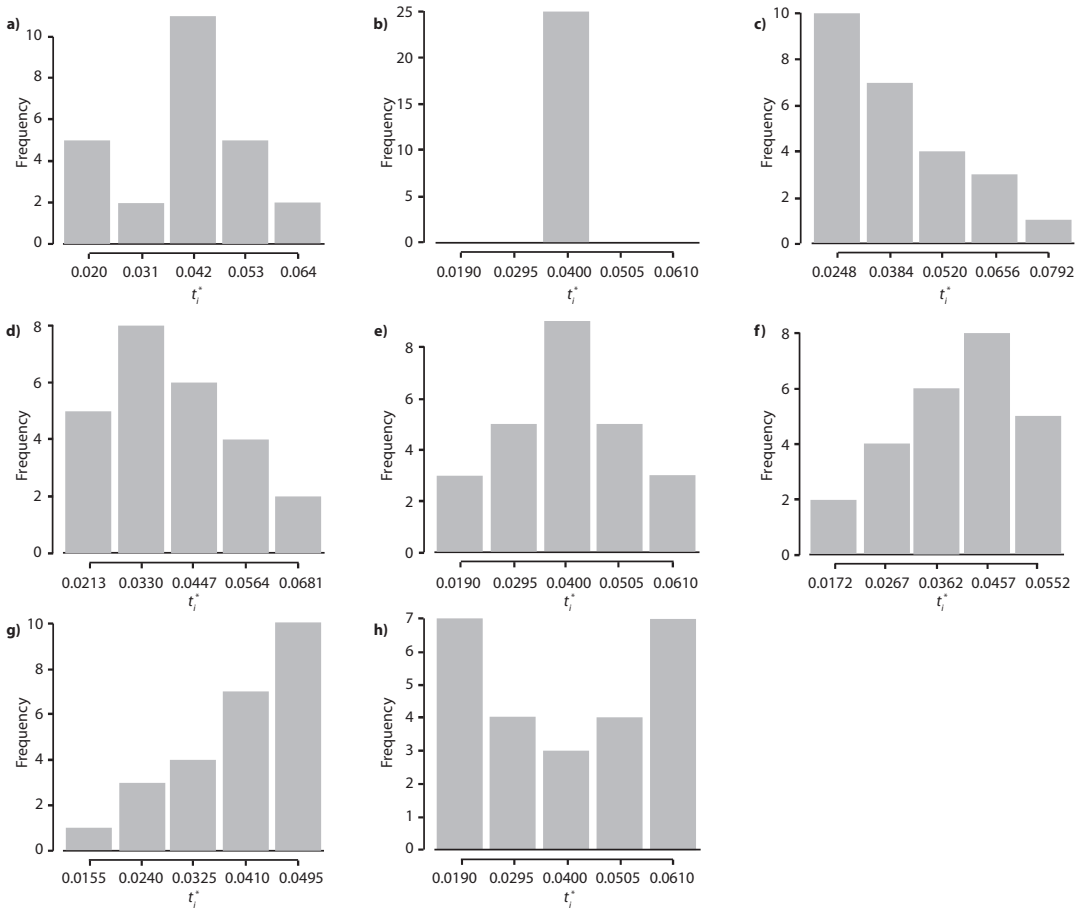


Fig. 1. a) The referred target matrix b) Target matrix 1 c) Target matrix 2 d) Target matrix 3 e) Target matrix 4 f) Target matrix 5 g) Target matrix 6 h) Target matrix 7.

confidence level. In other words, for a fixed N , low dimensional targets tend to give higher index values than targets of higher dimension.

Example 3: Targets with a variety of forms

The results in the above examples indicate that the form of the target distribution has a demonstrable effect on the size of N_0 required. For example, for the referred 5×5 target distribution and the choices $\varepsilon = 0.05$ and $\alpha = 0.05$, $N_0 = 2327$ is required (Table 6). For the 5×5 equal probability distribution, however, $N_0 = 2399$ is needed (Table 7). In this example we examine the effect of the form of the target distribution.

The relative target distributions to be compared are constructed by changing the frequency pattern of the relative values of the referred 5×5 target given in Table 1. The idea is to cover most of the commonly observed standard unimodal frequency curves. A detailed description of the construction procedure is left to Appendix C. The frequency histogram of the relative target values of the referred demand distribution and of those constructed (Target matrices 1–7) are displayed in Fig. 1. Table 11 in Appendix C reports some summary statistics related to the targets presented.

Table 8 shows N_0 solved for the target matrices 1–7, $\alpha = 0.05$ and a variety of ε ($= 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.10$). The bolded row corresponds to the 5×5 target

Table 8. N_0 solved for the target matrices 1–7, $\alpha = 0.05$ and a variety of ϵ .

Target	Ref. fig.	ϵ									
		0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
1	1b	59973	14993	6664	3748	2399	1666	1224	937	740	600
2	1c	58006	14502	6445	3625	2320	1611	1184	906	716	580
3	1d	58346	14586	6483	3647	2334	1621	1191	912	720	583
4	1e	58664	14666	6518	3667	2347	1630	1197	917	724	587
5	1f	58781	14695	6531	3674	2351	1633	1200	918	726	588
6	1g	59013	14753	6557	3688	2361	1639	1204	922	729	590
7	1h	57494	14373	6388	3593	2300	1597	1173	898	710	575

matrix with equal probabilities. The corresponding results for $\alpha = 0.01$ are displayed in Table 12 in Appendix C.

Table 8 above and Table 12 in Appendix C show that N_0 depends on the form of the given target matrix. For all pairs of ϵ and α examined in the example, we observe the largest N_0 for the target with equal cell probabilities. This behaviour can partly be explained by observing that for a fixed N , minimum $E(AI)$ is attained for $\mathbf{t}^* = (1/k, 1/k, \dots, 1/k)$. The above result is easy to prove by denoting $a_i = \sqrt{t_i^*}$ and $b_i = \sqrt{1 - t_i^*}$, and applying the Cauchy-Schwarz inequality, i.e. $(\sum_i a_i b_i)^2 \leq (\sum_i a_i^2)(\sum_i b_i^2)$.

6 Discussion

Although the Apportionment Index is the most widely used measure for evaluating the bucking outcome in Scandinavia, relatively little is known regarding its statistical properties. In this article we have examined the asymptotic sampling distribution of the AI assuming multinomial distribution for the outcome and utilizing large sample normal approximations. The assumed distribution of the random outcome should be reasonable when the pre-harvest planning procedure is performing appropriately. Under the assumption of the multinomial outcome distribution and for the given target matrix we have then derived the approximate first two moments, i.e. the mean and the variance. A simulation study is conducted to confirm that the approximations work reasonably well.

It can be easily seen from the derived formulas that the expected value of the AI increases as

the number of harvested logs increases and the expectation is almost one for very large number of logs. The derivations also show that the variance of the AI decreases towards zero as the number of logs increases, i.e. very little variation is expected for highly large number of logs. The findings are important and should, especially, be taken into account e.g. when comparisons are made between two different harvesting operations.

Since, for a given target distribution, the expected value of the Apportionment Index seems to strongly depend on the number of logs harvested, there may exist situations where not only the magnitude of the measure itself is of special interest but one also wants to know whether the performance of the harvesting operation is tolerable with respect to the number of stems harvested. This kind of a situation may arise, for example, when adjusting the bucking instructions according to the harvesting information gathered. In order to answer the question of a tolerable value of the AI we have derived a lower tolerance limit for the AI . It is then suggested that the lower tolerance limit should serve as the least acceptable value of the AI while index values higher than the mean would indicate a satisfactory level of agreement.

Studies are conducted on the determination of N (the number of harvested logs) needed to obtain a high apportionment with a given accuracy under some special kinds of distributions. Although our considerations are purely theoretical, our opinion is that they also have some useful practical implications. The conclusions that can be drawn from the studies in Section 5 are that not only the number of harvested logs, as noted earlier, affect the magnitude of the AI but also the dimension of the target matrix as well as the form of the target distribution. It is hence important

to also study how the values of the *AI* between target matrices with different dimensions (and varying number of logs) can be made commensurable. However, more thorough investigations are needed to exclusively determine how the form and size of the target matrix are actually affecting the *AI*. Such knowledge would be important, for example, when interpreting the values of the *AI* with certain other possible measures, as discussed by Kivinen et al. (2005).

Instead of using the Apportionment Index (or the *OVL*) alone to evaluate the similarity of two distributions, standard statistical tests (e.g. χ^2 -test) can also be applied (see e.g. Nummi et al. 2005). For a large sample size the standard statistical tests often possess high power to detect small differences. Thus virtually unimportant deviations from the target may in practice lead to rejection of the null hypothesis. As discussed by Inman and Bradley (1989) regarding the *OVL*, the *AI* should also be seen as supplementary to statistical tests rather than as an alternative method.

The connection of the Apportionment Index to the *OVL* and some other related measures is taken up in the paper to encourage those interested in the topic to do some literature research in other fields of science such as sociology. The authors strongly believe that the work done e.g. on the Dissimilarity Index and other measures of segregation may give valuable ideas for the research of the Apportionment Index as well as for the development of new measures of apportionment.

Acknowledgements

The study was completed under the project entitled “Forest-level bucking optimization including transportation cost, product demand and stand characteristics” financed by the Academy of Finland (Project 1104405). A major part of the work for this article was done while the first author was visiting the Division of Theoretical Statistics & Mathematics in the Indian Statistical Institute in Kolkata during July, 2005. The visit was financed by the Tampere Graduate School in Information Science and Engineering (TISE) and the Indian Statistical Institute. The authors thank Dr. V.-P. Kivinen for his contributions to this paper. The

authors would also like to thank the referees and the Assisting scientific editor for the comments that led to improvements of the paper.

References

- Agresti, A. 2002. Categorical data analysis, 2nd edition. New York: Wiley.
- Bergstrand, K.-G. 1989. Fördelningsapatering med näroptimalmetoden – Reviderad version [Bucking to order with a close-to-optimal method – Revised version]. Forskningsstiftelsen Skogsarbeten. 1989-12-11. 11 p. (In Swedish.)
- Casella, G. & Berger, R.L. 2002. Statistical inference. 2nd edition. Duxbury Press.
- Duncan, O.D. & Duncan, B. 1955. A methodological analysis of segregation indices. *American Sociological Review* 20: 210–217.
- Inman, H.F. & Bradley, E.L. 1989. The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities. *Communications in Statistics: Theory and Methods* 18: 3851–3874.
- & Bradley, E.L. 1991. Approximation to the mean and variance of the index of dissimilarity in $2 \times C$ tables under random allocation model. *Sociological Methods & Research* 20(2): 242–255.
- Jahn, J., Schmidt, C.F. & Schrag, C. 1947. The measurement of ecological segregation. *American Sociological Review* 12: 293–303.
- Kirkkala, A., Sikanen, L., Harstela, P., Ruha, T. & Tarnanen, T. 2000. Sakkosegmentoitu tavoitejakauma apteeraustuloksen arvioinnissa ja jakauma-asteen laskennassa [Assessing the goodness of the bucking outcome by penalty-segmented log demand distribution]. *Metsätieteen aikakauskirja* 1(2000): 59–61. (In Finnish.)
- Kivinen, V.-P. 2004. A genetic algorithm approach to tree bucking optimization. *Forest Science* 50(5): 696–710.
- , Uusitalo, J. & Nummi, T. 2005. Comparison of four measures designed between the demand and output distributions of logs. *Canadian Journal of Forest Research* 35: 693–702.
- Koskela, L. 2007. Contributions to statistical aspects of computerized forest harvesting. Academic dissertation, Acta Universitatis Tamperensis 1237.

Tampere University Press, Tampere, Finland.

Malinen, J. & Palander, T. 2004. Metrics for distribution similarity applied to the bucking to demand procedure. *International Journal of Forest Engineering* 15: 33–40.

Nummi, T., Sinha, B.K. & Koskela, L. 2005. Statistical properties of the apportionment degree and alternative measures in bucking outcome. *Revista Investigación Operacional* 26(3): 259–267.

Ransom, M.R. 2000. Sampling distributions of segregation indexes. *Sociological Methods & Research* 28(4): 454–475.

Sinha, B.K., Koskela, L. & Nummi, T. 2005a. On some statistical properties of the apportionment index. *Revista Investigación Operacional* 26(2): 169–179.

—, Koskela, L. & Nummi, T. 2005b. On a family of apportionment indices and its limiting properties. *IAPQR Transactions* 30(2): 65–87.

Vuorenpää, T., Aaltonen, A., Imponen, V. & Lukkarinen, E. 1997. Tukkijakauman ohjaus [Control of output log distributions]. *Metsätehon raportti* 38, Metsäteho Oy, Helsinki, Finland, 24 p. (In Finnish.)

Weijo, A. 2000. Runkopankki Metsähallituksen puunkannan ohjauksen apuvälineenä [The stem bank as a control tool in the wood procurement of the Metsähallitus corporation]. Master's thesis, Faculty of Forestry, University of Joensuu, Joensuu, Finland. (In Finnish.)

Total of 18 references

Appendix A

In this Appendix we present some results and corollaries to be used in Section 3.

Result 1. Suppose $X \sim N(\theta, \sigma^2)$.

Then

$$E(|X - \theta|) = \sigma \sqrt{\frac{2}{\pi}} \tag{24a}$$

$$V(|X - \theta|) = \sigma^2 \left(1 - \frac{2}{\pi}\right) \tag{24b}$$

The proof is omitted.

Result 2. Suppose $(X, Y) \sim BVN(0, 0, 1, 1, \rho)$.

Then

$$P(X > 0, Y > 0) = \int_0^\infty \int_0^\infty \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(x^2+y^2-2\rho xy)} dx dy = \frac{1}{4} + \frac{\arcsin \rho}{2\pi}$$

In other words,

$$\begin{aligned} I(\rho) &= \int_0^\infty \int_0^\infty e^{-\frac{1}{2(1-\rho^2)}(x^2+y^2-2\rho xy)} dx dy \\ &= 2\pi\sqrt{1-\rho^2} \left[\frac{1}{4} + \frac{\arcsin \rho}{2\pi} \right] \\ &= \frac{\pi}{2}\sqrt{1-\rho^2} + \sqrt{1-\rho^2} \arcsin \rho \end{aligned} \tag{25}$$

The proof is omitted.

Result 3.

$$\begin{aligned}
 J(\rho) &= \int_0^\infty \int_0^\infty uve^{-\frac{1}{2}(u^2+v^2-2\rho uv)} dudv \\
 &= \frac{\rho}{(1-\rho^2)^{3/2}} \left\{ \frac{\pi}{2} + \arcsin \rho \right\} + \frac{1}{1-\rho^2}, \quad -1 < \rho < 1, \tag{26} \\
 J(0) &= 1
 \end{aligned}$$

Proof.

Set

$$K(\rho) = \int_0^\infty \int_0^\infty e^{-\frac{1}{2}\{u^2+v^2-2\rho uv\}} dudv \tag{27}$$

Upon substituting $u = \frac{x}{\sqrt{1-\rho^2}}$ and $v = \frac{y}{\sqrt{1-\rho^2}}$ into Eq. 27 and using Eq. 25, we obtain

$$\begin{aligned}
 K(\rho) &= \frac{1}{(1-\rho^2)} \int_0^\infty \int_0^\infty e^{-\frac{1}{2(1-\rho^2)}\{x^2+y^2-2\rho xy\}} dx dy \\
 &= \frac{I(\rho)}{1-\rho^2} = \frac{1}{\sqrt{1-\rho^2}} \left[\frac{\pi}{2} + \arcsin \rho \right] \tag{28}
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 J(\rho) &= \int_0^\infty \int_0^\infty uve^{-\frac{1}{2}\{u^2+v^2-2\rho uv\}} dudv \tag{29} \\
 &= \frac{\partial K(\rho)}{\partial \rho} = \frac{\partial}{\partial \rho} \left[\frac{1}{\sqrt{1-\rho^2}} \left\{ \frac{\pi}{2} + \arcsin \rho \right\} \right] \\
 &= \frac{\rho}{(1-\rho^2)^{3/2}} \left\{ \frac{\pi}{2} + \arcsin \rho \right\} + \frac{1}{1-\rho^2}, \quad -1 < \rho < 1
 \end{aligned}$$

From Eq. 29 it is evident that $J(0) = 1$.

Result 4. Suppose $(X,Y) \sim BVN(0,0,1,1,\rho)$.

Then

$$\psi(\rho) = \psi(-\rho) = E(|XY|) = \frac{2}{\pi} |\rho| \arcsin |\rho| + \frac{2}{\pi} \sqrt{1-\rho^2} \tag{30}$$

Table 9 displays $\psi(\rho)$ for some selected ρ . Note that $\psi(0) = \frac{2}{\pi}$.

Proof.

$$\begin{aligned}
 E(|XY|) &= \varphi_{++}(\rho) + \varphi_{+-}(\rho) + \varphi_{-+}(\rho) + \varphi_{--}(\rho) = 2[\varphi_{++}(\rho) + \varphi_{+-}(\rho)] \\
 &= 2[\varphi_{++}(\rho) + \varphi_{++}(-\rho)],
 \end{aligned}$$

where $\varphi_{++}(\rho) = \int_0^\infty \int_0^\infty xyBVN(0,0,1,1,\rho) dx dy$, $\varphi_{+-}(\rho) = \int_0^\infty \int_{-\infty}^0 -xyBVN(0,0,1,1,\rho) dx dy$ and other φ 's are defined similarly.

Next observe that

$$\begin{aligned} \varphi_{++}(\rho) &= \int_0^\infty \int_0^\infty \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\{x^2+y^2-2\rho xy\}} xy dx dy && \left. \begin{array}{l} \text{upon substituting} \\ x = u\sqrt{1-\rho^2} \text{ and} \\ y = v\sqrt{1-\rho^2} \end{array} \right\} \\ &= \frac{(1-\rho^2)^{3/2}}{2\pi} J(\rho) \\ &= \frac{\rho}{4} + \frac{\rho}{2\pi} \arcsin \rho + \frac{\sqrt{1-\rho^2}}{2\pi} \end{aligned} \tag{31}$$

(See Result 3 for the definition of $J(\rho)$.)

Likewise,

$$\varphi_{+-}(\rho) = \varphi_{-+}(\rho) = -\frac{\rho}{4} + \frac{\rho}{2\pi} \arcsin \rho + \frac{\sqrt{1-\rho^2}}{2\pi} \tag{32}$$

Therefore,

$$\begin{aligned} E(|XY|) &= 2 \left[\left\{ \frac{\rho}{4} + \frac{\rho}{2\pi} \arcsin \rho + \frac{\sqrt{1-\rho^2}}{2\pi} \right\} + \left\{ -\frac{\rho}{4} + \frac{\rho}{2\pi} \arcsin \rho + \frac{\sqrt{1-\rho^2}}{2\pi} \right\} \right] \\ &= \frac{2\rho}{\pi} \arcsin \rho + \frac{2}{\pi} \sqrt{1-\rho^2} = \psi(\rho) = (\psi(-\rho)) \end{aligned}$$

Table 9. $\psi(\rho)$ for some selected ρ .

ρ	0.00	0.05	0.10	0.15	0.20	0.25	0.30
$\psi(\rho)$	0.6366	0.6374	0.6398	0.6438	0.6494	0.6566	0.6655
ρ	0.35	0.40	0.45	0.50	0.55	0.60	0.65
$\psi(\rho)$	0.6760	0.6883	0.7022	0.7180	0.7356	0.7551	0.7766
ρ	0.70	0.75	0.80	0.85	0.90	0.95	1.00
$\psi(\rho)$	0.8002	0.8260	0.8542	0.8851	0.9191	0.9567	1.0000

Corollary 1. Suppose $(X, Y) \sim BVN(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$.

Then

$$E(|X - \mu_1| |Y - \mu_2|) = \sigma_1 \sigma_2 \psi(\rho) \tag{33}$$

Corollary 2. Suppose $(X, Y) \sim BVN(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$.

Then

$$\begin{aligned} Cov(|X - \mu_1| |Y - \mu_2|) &= E(|(X - \mu_1)(Y - \mu_2)|) - E(|X - \mu_1|)E(|Y - \mu_2|) \\ &= \sigma_1 \sigma_2 \psi(\rho) - \sigma_1 \sigma_2 \frac{2}{\pi} = \sigma_1 \sigma_2 \left[\psi(\rho) - \frac{2}{\pi} \right] \end{aligned} \tag{34}$$

Appendix B

Assume the relative demand distribution $\mathbf{t}^* = (1/k, 1/k, \dots, 1/k)$. Table 10 displays N_0 for $\varepsilon = 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.10$ and $\alpha = 0.01$ as k varies from 20 to 30. The bolded row corresponds to the 5×5 target with equal probabilities.

Table 10. N_0 solved for $\alpha = 0.01$ and some selected values of ε as k varies from 20 to 30.

	ε									
	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
20	59436	14859	6604	3715	2377	1651	1213	929	734	594
21	61664	15416	6852	3854	2467	1713	1258	964	761	617
22	63876	15969	7097	3992	2555	1774	1304	998	789	639
23	66074	16519	7342	4130	2643	1835	1348	1032	816	661
24	68258	17064	7584	4266	2730	1896	1393	1067	843	683
k 25	70429	17607	7825	4402	2817	1956	1437	1100	869	704
26	72588	18147	8065	4537	2904	2016	1481	1134	896	726
27	74735	18684	8304	4671	2989	2076	1525	1168	923	747
28	76872	19218	8541	4805	3075	2135	1569	1201	949	769
29	78999	19750	8778	4937	3160	2194	1612	1234	975	790
30	81116	20279	9013	5070	3245	2253	1655	1267	1001	811

Appendix C

Let us define the summary statistics related to a target matrix as

$$Mean = \frac{1}{k} \tag{35a}$$

$$Var = \frac{\sum_{i=1}^k \left(t_i^* - \frac{1}{k} \right)^2}{k - 1} \tag{35b}$$

$$CV = \sqrt{\frac{\sum_{i=1}^k t_i^{*2}}{k} - \left(\frac{1}{k} \right)^2} / \left(\frac{1}{k} \right) \tag{35c}$$

where *Mean*, *Var* and *CV* are the sample mean, sample variance and the coefficient of variation calculated for the target values. We also include in the summary statistics the functions $\Delta(\mathbf{t}^*)$ and $\Gamma(\mathbf{t}^*)$, which are defined in Eq. 22 and Eq. 23 in Section 5, respectively. For the referred 5×5 target in Table 1 in Section 4 $Mean = 0.0400$, $Var = 0.0002$, $CV = 0.3458$, $\Delta(\mathbf{t}^*) = 0.3610$ and $\Gamma(\mathbf{t}^*) = 4.8078$.

The relative target distributions to be compared in Example 3 in Section 5 are constructed by changing the frequency pattern of the relative target values of the referred 5×5 target given in Table 1. (See their summary statistics in Table 11 below.) First, all 25 relative proportions of the referred target matrix are classified into 5 intervals of equal length ($= 0.011$) so that $(f_1, f_2, \dots, f_5) = (5, 2, 11, 5, 2)$ and $(\tilde{t}_1, \tilde{t}_2, \dots, \tilde{t}_5) = (0.020, 0.031, 0.042, 0.053, 0.064)$, where \tilde{t}_i denotes the mid-value of the i th interval and f_i is the corresponding frequency. For a new frequency pattern, say $(f'_1, f'_2, \dots, f'_5)$, we define the i th adjusted mid-value as $\tilde{t}_i^{**} = \tilde{t}_i / \left(\sum_{j=1}^5 f_j \tilde{t}_j \right)$, $i = 1, \dots, 5$. The corresponding relative target matrix is then constructed by setting the relative target values in the first f'_1 log categories equal to \tilde{t}_1^{**} , in the next f'_2 log categories equal to \tilde{t}_2^{**} etc.

Table 12 shows N_0 solved for the target matrices 1–7, $\alpha = 0.01$ and a variety of ε ($= 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.10$). The bolded row corresponds to the 5×5 target matrix with equal probabilities.

Table 11. The summary statistics of the referred 5 × 5 target matrix and of Targets 1–7.

Target	Ref. figure	$\tilde{f}_1^{**}, \dots, \tilde{f}_5^{**}$	f_1, \dots, f_5	Mean	Var	CV	$\Gamma(t^*)$	$\Delta(t^*)$
Referred	1a	0.0197,0.0305,0.0413, 0.0521,0.0629	5,2,11,5,2	0.0400	0.0002	0.3183	4.8238	0.3611
1	1b	0.0190,0.0295,0.0400, 0.0505,0.0610	0,0,25,0,0	0.0400	0.0000	0.0000	4.8990	0.3616
2	1c	0.0248,0.0384,0.0520, 0.0656,0.0792	10,7,4,3,1	0.0400	0.0003	0.4006	4.7990	0.3607
3	1d	0.0213,0.0330,0.0447, 0.0564,0.0681	5,8,6,4,2	0.0400	0.0002	0.3511	4.8162	0.3609
4	1e	0.0190,0.0295,0.0400, 0.0505,0.0610	3,5,9,5,3	0.0400	0.0002	0.3054	4.8324	0.3611
5	1f	0.0172,0.0267,0.0362, 0.0457,0.0552	2,4,6,8,5	0.0400	0.0001	0.2845	4.8383	0.3612
6	1g	0.0155,0.0240,0.0325, 0.0410,0.0495	1,3,4,7,10	0.0400	0.0001	0.2505	4.8501	0.3613
7	1h	0.0190,0.0295,0.0400, 0.0505,0.0610	7,4,3,4,7	0.0400	0.0003	0.4190	4.7722	0.3607

Table 12. N_0 solved for the target matrices 1–7, $\alpha = 0.01$ and a variety of ϵ .

Target	Ref. figure	ϵ									
		0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
1	1b	70429	17607	7825	4402	2817	1956	1437	1100	869	704
2	1c	68282	17071	7587	4268	2731	1897	1394	1067	843	683
3	1d	68654	17164	7628	4291	2746	1907	1401	1073	848	687
4	1e	69002	17251	7667	4313	2760	1917	1408	1078	852	690
5	1f	69131	17283	7681	4321	2765	1920	1411	1080	853	691
6	1g	69383	17346	7709	4336	2775	1927	1416	1084	857	694
7	1h	67727	16932	7525	4233	2709	1881	1382	1058	836	677