Conversion of ordinal attitudinal scales: an inferential Bayesian approach

by

Michael Evans
Department of Statistics
University of Toronto

and

Zvi Gilula
Department of Statistics
Hebrew University

and

Irwin Guttman
Department of Statistics
SUNY Buffalo and the University of Toronto

Technical Report No. 1008 November 11, 2010

TECHNICAL REPORT SERIES

University of Toronto

Department of Statistics

Conversion of ordinal attitudinal scales: an inferential Bayesian approach

Evans, M. Gilula, Z. and Guttman, I.*

Abstract: The need for scale conversion may arise whenever an attitude of individuals is measured by independent entrepreneurs each using an ordinal scale of its own with possibly different numbers of (arbitrary) ordinal categories. Such situations are quite common in the marketing realm. The conversion of a score of an individual measured on one scale into an estimated score of a similar scale with a different range is the concern of this paper. An inferential Bayesian approach is adopted to analyze the situation where we believe the scale with fewer categories can be obtained by collapsing the finer scale. This leads to inferences concerning rules for the conversion of scales. Further, we propose a method for testing the validity of such a model. The use of the proposed methodology is exemplified on real data from surveys concerning performance evaluation and satisfaction.

Keywords: ordinal scales, collapsing scales, scale conversions, Bayesian inference.

1 Introduction

The need for scale conversion may arise whenever an attitude of individuals is measured by independent entrepreneurs each using an ordinal scale of his/her own with possibly different number of (arbitrary) ordinal categories. In the marketing realm, it is not uncommon to encounter situations where satisfaction of customers is measured independently by several competing marketing research companies. Each company has its own satisfaction scale and they all use (non-identical) independent samples of customers from the same population.

Although scales may differ by the number of ordinal categories used, they may all share the same structure, where the lowest category refers to "very dissatisfied" and the last category refers to "most satisfied". Another, not uncommon situation arises when a company that used to measure the satisfaction of its own customers in the past by a R-category scale, is now employing a scale

*Michael Evans is Professor of Statistics (email: mevans@utstat.utoronto.ca) at the University of Toronto, Zvi Gilula is Professor of Statistics (email: msgilula@mscc.huji.ac.il) at the Hebrew University and Irwin Guttman (email: sttirwin@acsu.buffalo.edu) is Emeritus Professor of Statistics at SUNY Buffalo and the University of Toronto.

of only $K < R$ categories for a variety of reasons (this has been the case with several universities measuring student satisfaction from courses attended).

The problem of how many ordinal categories an attitudinal scale should have is an old one. See for instance (Miller 1956), and the classical paper (Green and Rao 1970). Although some compelling scientific arguments are mentioned in the literature as to some optimal aspects of ordinal scales, there is no universal consensus as to the number of categories to use. As a result, situations as described above occur quite frequently.

It might be of interest to be able to answer the question: "can we systematically convert a score of an individual measured on a given scale into an estimated score on a similar-in-nature scale but with a different number of categories?". An attempt to answer this question is the concern of this paper.

The methodology proposed in this paper assumes the following:

(a) All the underlying scales measure the same kind of attitude (satisfaction, say). We do NOT propose to convert a "satisfaction" score measured by one scale to an "agreement" score measured by another scale.

(b) All the underlying scales are ordinal-categorical scales (and not interval or ratio scales). Ordinality of categories reflects the ordinal nature of an attitude as a concept ("more", "most", better", "best", etc.).

(c) All samples measured on these scales come from the same population and no individual is sampled more than once.

Suppose then that we have two scales, labelled I and II, that are presumably measuring the same phenomenon. Scale I classifies responses into R ordered values and scale II classifies responses into K ordered values and suppose that $R > K$.

Let $X \in \{1, \ldots, R\}$ and $Y \in \{1, \ldots, K\}$ denote the values taken by a randomly selected population element on scale I and scale II respectively. Let $p_i = P(X = i)$, for $i = 1, \ldots, R$, and $q_j = P(Y = j)$, for $j = 1, \ldots, K$, denote the distributions of these variables over the population in question. Of course, these distributions are unknown and we suppose that data have been collected from sampling studies to make inferences about these distributions. Let $(f_1, \ldots, f_R)$, with $\sum_{i=1}^{R} f_i = N_R$, denote the counts from the sampling study involving scale I and $(g_1, \ldots, g_K)$, with $\sum_{j=1}^{K} g_j = N_K$, denote the counts from the sampling study involving scale II.

Our problem is concerned with determining how we should combine the results from the two studies. We take the point-of-view that our aim should be collapsing, in some fashion, scale I into scale II and then combining our inferences. In the typical application it is not at all clear how this collapsing should take place. We would like this collapsing to depend on the data and proceed according to sound inferential principles rather than follow some ad hoc procedure.

In Section 2 we formulate a Bayesian model for the collapsing. We find it most convenient to place this discussion in a Bayesian context and include discussion on appropriate priors for this problem. Also in Section 2 we discuss a computational approach for implementing a Bayesian analysis and consider some simulated data. In Section 3 we discuss the basic Bayesian inferences that follow from the model. In Section 4 we show how inferences about the true collapsing lead to rules for converting one scale to another and an assessment of the uncertainty associated with such rules. In Section 5 we consider methods for checking the model specified in Section 2. In Section 6 we apply these results to a problem of some practical importance and draw some conclusions in Section 7.

2 The Model, the Prior, and the Posterior

Given that we have observed $(f_1, \ldots, f_R)$ and $(g_1, \ldots, g_K)$, the likelihood for the unknown p and q is given by

$$L(p, q \mid f, g) = \prod_{i=1}^{R} p_i^{f_i} \prod_{j=1}^{K} q_j^{g_j}, \tag{1}$$

where $p \in S_R, q \in S_K$ and S_N denotes the $(N-1)$-dimensional simplex. Let π_R denote a prior on p. Typically, we will take this prior to be a uniform prior, i.e., $\text{Dirichlet}_R(1, \ldots, 1)$, to reflect noninformativity, but we will consider a general $\text{Dirichlet}_R(a_1, \ldots, a_R)$.

Now suppose that p is given. Of course, we do not know these values but now we will effectively put a prior on the set of all possible collapsings assuming that we know p. In other words we will specify the prior hierarchically by specifying a prior for the collapsings given the value of p, and then place a marginal prior on p as we have already specified.

First we develop a model for the collapsing. Let $(\mathcal{P}_1, \ldots, \mathcal{P}_K)$ denote an ordered partition of $\{1, \ldots, R\}$ into K disjoint subsets with $\cup_{i=1}^{K} \mathcal{P}_i = \{1, \ldots, R\}$ such that $\mathcal{P}_i \neq \phi$ for all i and, if $1 \leq i < j \leq K$, then $\mathcal{P}_i = \{s, s+1 \ldots, t\}, \mathcal{P}_j = \{u, u+1 \ldots, v\}$ with $1 \leq s \leq t < u \leq v \leq R$. The set $\mathcal{P}_i$ consists of those categories on scale I that are collapsed to form category i on scale II. Note that we have restricted to those collapsings for which no null sets are allowed in the partition, because we know that we can observe observations in each category of scale II. Therefore, proceeding as if the categories in scale II arise from a collapsing of scale I, the likelihood (1) is no longer correct, rather the likelihood for $(p, (\mathcal{P}_1, \ldots, \mathcal{P}_K))$ is given by

$$L(p, (\mathcal{P}_1, \ldots, \mathcal{P}_K) \mid f, g) = \prod_{i=1}^{R} p_i^{f_i} \prod_{j=1}^{K} \left(\sum_{l \in \mathcal{P}_j} p_l \right)^{g_j}. \tag{2}$$

The likelihood expresses the information in the data concerning the possible collapsings and the value of p.

If we denote the conditional prior on the set of all ordered partitions given p by $\pi((\mathcal{P}_1, \ldots, \mathcal{P}_K) \,|\, p)$, then the marginal prior on $(\mathcal{P}_1, \ldots, \mathcal{P}_K)$ is

$$\pi(\mathcal{P}_1, \ldots, \mathcal{P}_K) = \int_{S_R} \pi(\mathcal{P}_1, \ldots, \mathcal{P}_K \,|\, p)\pi_R(p)\, dp. \tag{3}$$

As we will see, we need to compute this quantity to implement some of our inferences. Of course, we will also need to compute the posterior probability of $(\mathcal{P}_1, \ldots, \mathcal{P}_K)$, namely,

$$\lambda(\mathcal{P}_1, \ldots, \mathcal{P}_K \,|\, f, g) \propto \int_{S_R} \prod_{i=1}^{R} p_i^{f_i} \prod_{j=1}^{K} \left(\sum_{l \in \mathcal{P}_j} p_l \right)^{g_j} \pi(\mathcal{P}_1, \ldots, \mathcal{P}_K \,|\, p)\pi_R(p)\, dp. \tag{4}$$

Clearly the specification of $\pi(\mathcal{P}_1, \ldots, \mathcal{P}_K \,|\, p)$ is a key step in the analysis. Note that the number of such partitions $(\mathcal{P}_1, \ldots, \mathcal{P}_K)$ equals the number of solutions $(i_1, \ldots, i_K)$ of $i_1 + \cdots + i_K = R$ in positive integers. Therefore, there are $\binom{R-1}{K-1}$ such partitions. Typically, for the applications we have in mind, this number will not be too large.

One possible specification for $\pi(\mathcal{P}_1, \ldots, \mathcal{P}_K \,|\, p)$ is the uniform prior. The uniform prior puts weight $1/\binom{R-1}{K-1}$ on each possible partition and so (3) equals this quantity no matter what prior is placed on p. With this choice, and π_R equal to the $\text{Dirichlet}_R(a_1, \ldots, a_R)$ density, we have that (4) is proportional to

$$\int_{S_R} \prod_{i=1}^{R} p_i^{f_i + a_i - 1} \prod_{j=1}^{K} \left(\sum_{l \in \mathcal{P}_j} p_l \right)^{g_j} dp.$$

Now, letting $u_j = \sum_{l \in \mathcal{P}_j} p_l$, and using

$$u \sim \text{Dirichlet}_K \left(\sum_{l \in \mathcal{P}_1} (f_l + a_l), \ldots, \sum_{l \in \mathcal{P}_K} (f_l + a_l) \right)$$

when $p \sim \text{Dirichlet}_R(f_1 + a_1, \ldots, f_R + a_R)$, we see that (4) is proportional to

$$\frac{\Gamma\left(\sum_{i=1}^{R} f_i + \sum_{i=1}^{R} a_i\right)}{\prod_{j=1}^{K} \Gamma\left(\sum_{l \in \mathcal{P}_j} (f_l + a_l)\right)} \int_{S_K} \prod_{j=1}^{K} u_j^{g_j + \sum_{l \in \mathcal{P}_j} (f_l + a_l) - 1}\, du$$
$$= \prod_{j=1}^{K} \frac{\Gamma\left(g_j + \sum_{l \in \mathcal{P}_j} (f_l + a_l)\right)}{\Gamma\left(\sum_{l \in \mathcal{P}_j} (f_l + a_l)\right)}.$$

So to obtain the posterior we need to sum this quantity over all partitions to evaluate the normalizing constant for the posterior probability function.

The same argument establishes closed-form expressions for (3) and (4) for a more general family of priors.

Theorem 1. Suppose that the prior on $(p,(\mathcal{P}_1,\ldots,\mathcal{P}_K))$ is specified by

$$\pi(\mathcal{P}_1,\ldots,\mathcal{P}_K \,|\, p) \propto \prod_{j=1}^{K}\left(\sum_{l\in\mathcal{P}_j} p_l\right)^{b_j}, \tag{5}$$

for $b_j > -1$, $j = 1,\ldots,K$, and $p \sim \text{Dirichlet}_R(a_1,\ldots,a_R)$. Then we have that

$$\pi(\mathcal{P}_1,\ldots,\mathcal{P}_K) \propto \prod_{j=1}^{K}\frac{\Gamma\left(b_j+\sum_{l\in\mathcal{P}_j} a_l\right)}{\Gamma\left(\sum_{l\in\mathcal{P}_j} a_l\right)} \tag{6}$$

and

$$\lambda(\mathcal{P}_1,\ldots,\mathcal{P}_K \,|\, f,g) \propto \prod_{j=1}^{K}\frac{\Gamma\left(b_j+g_j+\sum_{l\in\mathcal{P}_j}(f_l+a_l)\right)}{\Gamma\left(\sum_{l\in\mathcal{P}_j}(f_l+a_l)\right)}. \tag{7}$$

Note that the uniform prior on $(\mathcal{P}_1,\ldots,\mathcal{P}_K)$ corresponds to $b_1=\cdots=b_K=0$. If $a_1=\cdots=a_R=1$, i.e., we place a uniform prior on p, and $b_1>0$ while $b_2=\cdots=b_K=0$, then partitions that have more elements in $\mathcal{P}_1$ will receive more prior weight than partitions with fewer elements in $\mathcal{P}_1$. Similar interpretations can be made for other choices of the b_j, although they are difficult to describe precisely. It would seem, however, if we want to emphasize partitions that place more elements in certain cells of the partition then we should choose the corresponding b_j relatively large. We will refer to the prior given in (5) by the K-tuple $[b_1,\ldots,b_K]$ hereafter, whenever the prior on p is uniform. So, for example, $[0,\ldots,0]$ corresponds to the uniform prior on the set of partitions and the uniform prior on S_R.

The following example illustrates an approach to tabulating the posterior.

Example 1. *Tabulating the prior and the posterior*

To specify a partition we need only prescribe an ordered K-tuple of integers $(n_1,\ldots,n_K)$ such that $1\le n_1<n_2<\cdots<n_K=R$, and there are are $\binom{R-1}{K-1}$ such K-tuples. Denote this set of K-tuples by $T_{R,K}$. For example, if $R=5$ and $K=3$, then $\binom{4}{2}=6$ and $(n_1,n_2,n_3)=(2,4,5)$ specifies the partition $(\mathcal{P}_1,\mathcal{P}_2,\mathcal{P}_3)=(\{1,2\},\{3,4\},\{5\})$. We can systematically tabulate any prior and posterior via (6) and (7) by running through the K-tuples $(n_1,\ldots,R)$ starting with $(1,2,\ldots,K-1,R)$, letting n_{K-1} range from $K-1$ to $R-1$, then change n_{K-2} to $K-1$ and letting n_{K-1} range from K up to $R-1$, etc.

In Table 1 we have tabulated several priors. Note that the left column lists the elements of $T_{5,3}$, while the first row specifies several priors. From this we can see that the larger the value of $b_i>0$ is, the more prior weight is being placed on partitions that have more elements in $\mathcal{P}_i$, while the closer b_i is to -1 the more prior weight is being placed on partitions that place fewer elements in $\mathcal{P}_i$.

partition( )\prior[ ]	$[0,0,0]$	$[10,0,0]$	$[10,10,0]$	$[1,1,1]$	$[-.5,0,0]$
$(1,2,5)$	.167	.011	.004	.143	.229
$(1,3,5)$	.167	.011	.040	.190	.229
$(1,4,5)$	.167	.011	.239	.143	.229
$(2,3,5)$	.167	.121	.040	.190	.114
$(2,4,5)$	.167	.121	.438	.190	.114
$(3,4,5)$	.167	.725	.239	.143	.086

Table 1: Several priors on partitions in Example 1.

Suppose we observe $(f_1, f_2, f_3, f_4, f_5) = (4, 3, 6, 3, 7)$ and $(g_1, g_2, g_3) = (7, 9, 7)$. Note that (g_1, g_2, g_3) corresponds exactly to partition $(2, 4, 5)$. In Table 2 we have tabulated the posterior probabilities for the priors in Table 1. Notice that the correct partition $(2, 4, 5)$ is the posterior mode for each analysis except that using the prior $[10, 0, 0]$. This prior puts too much weight on $\mathcal{P}_1$ having most of the collapsing and so $(3, 4, 5)$ results as the mode. Still for a relatively small amount of data the methodology is clearly on the right track in identifying suitable collapsings. Also worth noting is that the most incorrect collapsing given by $(1, 2, 5)$ is always given low posterior probability. The benefit of having appropriate strong prior information is clearly demonstrated by the prior $[10, 10, 0]$.

partition( )\prior[ ]	$[0,0,0]$	$[10,0,0]$	$[10,10,0]$	$[1,1,1]$	$[-.5,0,0]$
$(1,2,5)$	.007	.001	.000	.006	.007
$(1,3,5)$	.189	.018	.037	.185	.208
$(1,4,5)$	.181	.017	.162	.167	.199
$(2,3,5)$	.200	.176	.059	.208	.189
$(2,4,5)$	.398	.351	.714	.409	.376
$(3,4,5)$	.025	.437	.028	.024	.020

Table 2: Posterior probabilities on partitions for priors in Example 1.

3 Inferences for Collapsings

One possibility for inferences is to use those based on the highest posterior density (hpd) principle (see Box and Tiao 1973 for a discussion of various Bayesian inference terms). For this we record a γ-credible region of the form

$$B_\gamma(f, g) = \{(\mathcal{P}_1, \ldots, \mathcal{P}_K) : \lambda(\mathcal{P}_1, \ldots, \mathcal{P}_K \mid f, g) > k_\gamma(f, g)\}$$

where $k_\gamma(f, g)$ is the largest value such that $\Lambda(B_\gamma(f, g) \mid f, g) \geq \gamma$ where $\Lambda(\cdot \mid f, g)$ denotes posterior measure. The size of such regions, say for $\gamma = .95$, then tells us something about the accuracy of our inferences. This gives a nested sequence of subsets, indexed by γ, with the smallest set containing only the posterior mode.

In general, inferences based on the highest posterior density principle suffer from a lack of invariance. For example, if we wanted to make inference about the continuous parameters p, or q as derived from p (see Section 4), then a γ-hpd region does not transform appropriately under 1-1, smooth transformations. A class of inferences that are invariant under such transformations is given by the relative surprise principle (see Evans, Guttman and Swartz 2006 and Evans and Shakhatreh 2008 for discussion of relative surprise inferences). A γ-relative surprise credible region for the true partition is of the form

$$C_\gamma(f,g) = \left\{ (\mathcal{P}_1,\ldots,\mathcal{P}_K) : \frac{\lambda(\mathcal{P}_1,\ldots,\mathcal{P}_K \,|\, f,g)}{\pi(\mathcal{P}_1,\ldots,\mathcal{P}_K)} > k_\gamma(f,g) \right\}$$

and $k_\gamma(f,g)$ is the largest value such that $\Lambda(C_\gamma(f,g)\,|\,f,g) \geq \gamma$. We see that the *relative belief ratio* (see Evans and Shakhatreh 2008) for $(\mathcal{P}_1,\ldots,\mathcal{P}_K)$, given by

$$RB(\mathcal{P}_1,\ldots,\mathcal{P}_K) = \frac{\lambda(\mathcal{P}_1,\ldots,\mathcal{P}_K \,|\, f,g)}{\pi(\mathcal{P}_1,\ldots,\mathcal{P}_K)},$$

is a measure of how belief in the validity of the partition $(\mathcal{P}_1,\ldots,\mathcal{P}_K)$ has changed from *a priori* to *a posteriori*. Note that a relative belief ratio is similar to a Bayes factor, which is the ratio of the posterior odds to the prior odds, and both are measuring the change in belief from *a priori* to *a posteriori*. The relative belief ratio measures change in belief on the probability scale while the Bayes factor measures change in belief on the odds scale.

Accordingly, $C_\gamma(f,g)$ contains all partitions whose relative belief ratios are above some cut-off. This principle also gives a nested sequence of subsets, indexed by γ, with the smallest set containing only the least relative surprise estimate (LRSE) (Evans, Guttman and Swartz 2006), i.e., the value that maximizes the relative belief ratio for $(\mathcal{P}_1,\ldots,\mathcal{P}_K)$. Note that the LRSE partition is the same as maximum a posteriori (MAP) partition, i.e., the mode of the posterior, whenever the prior $\pi(\mathcal{P}_1,\ldots,\mathcal{P}_K)$ is uniform, but otherwise these estimates can be different as demonstrated in Example 2.

In the cited references various optimality properties have been proved for relative surprise inferences in the class of all Bayesian inferences. Note, in particular, that because relative surprise inferences are based on relative belief ratios, we can expect these inferences to be less dependent on the prior than say hpd-inferences, as we are dividing the posterior by the prior. In fact it can be proven that the relative belief ratio is independent of the choice of the marginal prior $\pi(\mathcal{P}_1,\ldots,\mathcal{P}_K)$.

Consider now inferences for Example 1.

Example 2. *Estimating the true partition in Example 1*

In Table 3 we have given the relative belief ratios $RB(\mathcal{P}_1,\ldots,\mathcal{P}_K)$, for the problem discussed in Example 1. These are obtained by dividing the entries in Table 2 by the corresponding entries in Table 1. Notice that the LRSE is given by the correct partition $(2,4,5)$ in every case, while the mode failed for the prior $[10,0,0]$. This is characteristic of relative surprise inferences as the prior has less influence in these inferences than in other Bayesian inferences.

partition( )\prior[ ]	$[0,0,0]$	$[10,0,0]$	$[10,10,0]$	$[1,1,1]$	$[-.5,0,0]$
$(1,2,5)$	0.042	0.091	0.000	0.042	0.031
$(1,3,5)$	1.132	1.636	0.925	0.974	0.908
$(1,4,5)$	1.084	1.546	0.678	1.168	0.869
$(2,3,5)$	1.198	1.455	1.475	1.095	1.658
$(2,4,5)$	2.383	2.901	1.630	2.153	3.298
$(3,4,5)$	0.150	0.603	0.117	0.168	0.233

Table 3: Relative belief ratios for partitions in Example 1.

From the posterior probabilities in Table 2 and the relative belief ratios in Table 3, we can construct $C_\gamma(f,g)$. For example, when the prior is $[0,0,0]$, then

$$C_{.95}(f,g) = \{(1,3,5),(1,4,5),(2,3,5),(2,4,5),(3,4,5)\}, \tag{8}$$

i.e., every partition but $(1,2,5)$ is included and $k_{.95}(f,g) = 0.150$. This is not very informative, but then we do not have a lot of data. Actually, from Table 2, the posterior content of $C_{.95}(f,g)$ is 0.993 in this case.

Suppose we wished to assess the hypothesis that a certain partition was true say,

$$H_0 : (\mathcal{P}_1,\ldots,\mathcal{P}_K) = (\mathcal{P}_1^*,\ldots,\mathcal{P}_K^*),$$

i.e., we want to assess whether or not the actual collapsing is $(\mathcal{P}_1^*,\ldots,\mathcal{P}_K^*)$, as prescribed by some theory. This can be assessed by computing

$$\gamma^* = \inf\{\gamma : (\mathcal{P}_1^*,\ldots,\mathcal{P}_K^*) \in C_\gamma(f,g)\}$$

and concluding that we have evidence against H_0 when γ^* is near 1. Equivalently we can compute the relative surprise P-value given by

$$\Lambda\left(\frac{\lambda(\mathcal{P}_1,\ldots,\mathcal{P}_K \mid f,g)}{\pi(\mathcal{P}_1,\ldots,\mathcal{P}_K)} \le \frac{\lambda(\mathcal{P}_1^*,\ldots,\mathcal{P}_K^* \mid f,g)}{\pi(\mathcal{P}_1^*,\ldots,\mathcal{P}_K^*)} \,\middle|\, f,g \right),$$

i.e., the posterior probability of obtaining a relative belief ratio no larger than that obtained for the hypothesized value. Small values for the P-value provide evidence against the null hypothesis.

Example 3. *Testing a partition in Example 1*

Suppose we want to test the null hypothesis H_0 that the true partition is $(2,4,5)$ when the prior is $[0,0,0]$. Then it is clear that the P-value is 1 and we have no evidence against H_0. If we tested the null hypothesis H_0 that the true partition is $(1,2,5)$ when the prior is $[0,0,0]$, then the P-value is $1-0.993 = 0.007$ and we have substantial evidence against H_0.

4 Conversion of Scales

A problem of major interest is how we should convert one scale to another when we believe that one scale is a collapsing of the other. In one direction it seems quite clear how to do this.

Once we have selected an estimate of the collapsing, then this immediately provides a conversion rule from the finer to the coarser scale, i.e., from scale I to scale II. To assess the uncertainty in this rule we can look at the rules that would arise from each of the partitions in a γ-credible region for the collapsing. We illustrate this in the following example.

Example 3. *Converting from the finer to the coarser scale in Example 1.*

In Example 2 we showed that the LRSE partition is always $(2, 4, 5)$ for all the priors specified in Example 1. The conversion rule from scale I to scale II, using the LRSE partition, is as shown in Table 4. For example, individuals assigned a score of 1 or 2 on scale I are assigned a score of 1 on scale II.

scale I	scale II
1	1
2	1
3	2
4	2
5	3

Table 4: Conversion of scale I to scale II in Example 1 using LRSE collapsing.

To assess the uncertainty entailed in this conversion rule we now present, in Table 5, the full set of conversion rules obtained via the .95-relative surprise region (8) given in Example 2, when using the prior $[0, 0, 0]$. Of course, values 1 and 5 on scale I must always convert to 1 and 3 on scale II. From the table we see that there is little variation in converting 3 on scale I but more variation for both 2 and 4, for example, four of the five partitions in the .95-credible region convert 3 to a 2.

scale I	scale II	scale II	scale II	scale II(LRSE)	scale II
1	1	1	1	1	1
2	2	2	1	1	1
3	2	2	2	2	1
4	3	2	3	2	2
5	3	3	3	3	3

Table 5: A .95 credible region (based on the partitions in (8)) for conversion of scale I to scale II in Example 1 when using the uniform prior.

A conversion rule for expanding scale II to scale I is more problematic. To see this, suppose we have selected a rule to convert from the finer to the coarser scale. Then, if we wanted to convert from the coarser to the finer scale we have multiple possible assignments for any categories that correspond to a proper collapsing on the finer scale. A number of rules are possible and we consider two rules, one deterministic, called the *maximal rule*, while the other corresponds to a random assignment, called the *random rule*. These rules are based on having chosen a specific collapsing as represented by the partition $(\mathcal{P}_1, \ldots, \mathcal{P}_K)$.

For the maximal rule we look at estimates $\hat{p}_i$ of the p_i, corresponding to a collapsed category, and choose, for the conversion of a value on the coarser scale, the value in the finer scale that maximizes $\hat{p}_i$. For the random rule we use the the estimates $\hat{p}_i$ to construct the conditional distributions for each partition element and then randomly allocate a value on scale II to a value on scale I using the appropriate conditional distribution. The random rule seems somewhat more realistic as not every observation on a given category on scale II corresponding to a collapsing, would be assigned the same category on scale I.

Logically, it makes sense to base the estimates $\hat{p}_i$ on the conditional posterior of p given that the particular collapsing in question holds. Fixing the partition $(\mathcal{P}_1, \ldots, \mathcal{P}_K)$, the likelihood for p is given by (2) as a function of p. Now suppose that we have placed a $\text{Dirichlet}_R(a_1, \ldots, a_R)$ prior on p to start. The conditional prior distribution of q is then $\text{Dirichlet}_K(\sum_{l\in\mathcal{P}_1} a_l, \ldots, \sum_{l\in\mathcal{P}_K} a_l)$ and, from (2), the conditional posterior distribution of q is $\text{Dirichlet}_K(g_1 + \sum_{l\in\mathcal{P}_1}(f_l + a_l), \ldots, g_K + \sum_{l\in\mathcal{P}_K}(f_l + a_l))$. Given q, the conditional prior of $(p_{l_i}/q_i, \ldots, p_{u_i}/q_i)$, corresponding to the partition element $\mathcal{P}_i = (l_i, l_i{+}1, \ldots, u_i)$, is $\text{Dirichlet}_{u_i-l_i+1}(a_{l_i}, \ldots, a_{u_i})$ and the conditional posterior is $\text{Dirichlet}_{u_i-l_i+1}(f_{l_i} + a_{l_i}, \ldots, f_{u_i} + a_{u_i})$. So the conditional LRSE of p_l/q_i, for l satisfying $l_i \le l \le u_i$, is given by $f_l/(f_{l_i} + \cdots + f_{u_i})$ which we note does not involve the prior. The maximal rule then converts i on scale II to the value l satisfying $l_i \le l \le u_i$ which maximizes $f_l/(f_{l_i} + \cdots + f_{u_i})$. The random rule converts i on scale II to the value l on scale I, satisfying $l_i \le l \le u_i$, with probability $f_l/(f_{l_i} + \cdots + f_{u_i})$.

We illustrate this in the following example.

Example 4. *Converting from the coarser to the finer scale in Example 1.*

In Table 6 we give the results of converting scale II to scale I, using the maximal rule, for each of the partitions in (8). This again gives us a sense of the uncertainties involved in this conversion.

scale II	scale I	scale I	scale I	scale I (LRSE)	scale I
1	1	1	1	1	3
2	3	3	3	3	4
3	5	5	5	5	5

Table 6: Possible conversions of scale II to scale I in Example 1 when using the uniform prior and based on partitions in the .95-relative surprise region.

We note that Table 5 is based on the .95-credible region (8) that we obtained for the collapsing, and so can be interpreted as a .95-credible region for the true conversion from scale I to scale II. This is not the case, however, for Table 6 as these are obtained by a somewhat ad hoc, albeit intuitively reasonable rule. There is an element of nonidentifiability in converting from scale II to scale I.

In Table 7 we record the conditional distributions corresponding to the random rule for the LRSE partition in Example 1. We can see from this that the choice between 1 and 2 on scale I for a 1 on scale II is somewhat uncertain.

scale II \ scale I	1	2	3	4	5
1	4/7	3/7	0	0	0
2	0	0	2/3	1/3	0
3	0	0	0	0	1

Table 7: Conversion probabilities for converting scale II to scale I in Example 1 based on the LRSE partition.

5 Model Checking

A question of some interest is whether or not the model for collapsings that we have developed in Section 2 makes sense in light of the observed data $(f_1, \ldots, f_R)$ and $(g_1, \ldots, g_K)$. This is effectively model checking and should be carried out before we proceed to make inferences about an appropriate collapsing.

One approach to model checking is based on the minimal sufficient statistic for the sampling model given by (2). Once this statistic is determined we use the conditional distribution of a discrepancy statistic given the value of the minimal sufficient statistic to assess the model via a P-value. Unfortunately, in this problem it is very difficult to determine the minimal sufficient statistic let alone determine the conditional distribution.

An alternative approach might be to select a discrepancy statistic, such as Pearson's chi-squared test statistic to compare the observed g_i with the estimated expected values based on the observed f, for a particular partition $(\mathcal{P}_1, \ldots, \mathcal{P}_K)$ and do this for every possible partition. This, however, entails the use of $\binom{R-1}{K-1}$ dependent chi-squared tests and it is not at all clear how to combine these to compute an appropriate P-value.

We adopt an alternative strategy for model checking here. Suppose we are satisfied that samples have been correctly obtained from the relevant population so we feel confident that the underlying multinomial model leading to likelihood (1) is correct. Now suppose that we had no information that would suggest that scale II is a collapsing of scale I. In such a case it makes sense to put independent priors on p and q and then use the posterior to make inferences about these quantities. Suppose we use the same prior on p as prescribed in Section 2, namely, $p \sim \text{Dirichlet}_R(a_1, \ldots, a_R)$ independent of $q \sim \text{Dirichlet}_K(1, \ldots, 1)$. Although other choices are possible for the prior on q, a noninformative prior would seem to make sense when we are testing a model.

Now consider the null hypothesis

$$H_0 : q \text{ is a collapsing of } p,$$

and we will consider H_0 as a subset of $S_R \times S_K$. Notice that, provided we place a continuous prior on $S_R \times S_K$, then H_0 is a subset having prior probability equal to 0. The consequence of this is that we can think of the $\text{Dirichlet}_R(a_1, \ldots, a_R) \times \text{Dirichlet}_K(1, \ldots, 1)$ prior as being effectively on H_0^c while the prior on H_0 is as described in Section 2. The prior on H_0 only comes into play if we agree that H_0 makes sense and, to assess this, we use the prior on the complement H_0^c.

This shows that there is no conflict in our prior assessments as the prior on p is the same.

To assess H_0 we proceed as described in Evans, Gilula, and Guttman (1993) and Evans, Gilula, Guttman and Swartz (1997). For an arbitrary point $(p,q) \in S_R \times S_K$ let $d_{H_0}(p,q)$ be a measure of the distance (p,q) is from H_0. For example, we will use

$$d_{H_0}(p,q) = \min_{(\mathcal{P}_1,\ldots,\mathcal{P}_K)} \sum_{i=1}^{K} \left(q_i - \sum_{l \in \mathcal{P}_i} p_l \right)^2, \tag{9}$$

i.e., least squares distance, although other choices are possible. Now we measure the concentrations of the prior distribution and the posterior distributions of (p,q) about H_0 by the prior and posterior distributions of $d_{H_0}(p,q)$, respectively. Naturally, if H_0 is true then we should see that the posterior distribution of $d_{H_0}(p,q)$ concentrates much more closely about 0 than the prior distribution of $d_{H_0}(p,q)$. To assess this formally let $\pi_{d_{H_0}}$ and $\pi_{d_{H_0}}(\cdot \,|\, f,g)$ be the prior and posterior densities of $d_{H_0}(p,q)$. Then to assess H_0 we compute the relative surprise P-value

$$\Pi\left(\frac{\pi_{d_{H_0}}(d_{H_0}(p,q) \,|\, f,g)}{\pi_{d_{H_0}}(d_{H_0}(p,q))} \le \frac{\pi_{d_{H_0}}(0 \,|\, f,g)}{\pi_{d_{H_0}}(0)} \,\middle|\, f,g \right) \tag{10}$$

where $\Pi(\cdot \,|\, f,g)$ is now the posterior of (p,q) based on the

$$\text{Dirichlet}_R(a_1,\ldots,a_R) \times \text{Dirichlet}_K(1,\ldots,1)$$

prior. The relative belief ratio of H_0 is given by $\pi_{d_{H_0}}(0 \,|\, f,g)/\pi_{d_{H_0}}(0)$ and this measures how the data have changed beliefs in H_0. An alternative to H_0 is given by a nonzero value of d_{H_0} and so we see that (10) is the posterior probability that an alternative to H_0 does not have a relative belief ratio greater than that of H_0. As such, when (10) is small there are alternatives that have a larger relative belief ratio than H_0 with high posterior probability and we can view this as evidence against H_0. Note that $\Pi(\cdot \,|\, f,g)$ is the

$$\text{Dirichlet}_R(a_1+f_1,\ldots,a_R+f_R) \times \text{Dirichlet}_K(1+g_1,\ldots,1+g_R)$$

distribution.

It is straightforward to sample from both the posterior and prior distributions of (p,q) and from such Monte Carlo samples we can construct estimates of $\pi_{d_{H_0}}$ and $\pi_{d_{H_0}}(\cdot \,|\, f,g)$ and estimate (10). The only difficulty lies in evaluating (9) for each sample and we simply do this by brute force. Provided $\binom{R-1}{K-1}$ is not large, this is feasible. We illustrate this procedure in the following example.

Example 5. *Testing the model in Example 1.*

In Figure 1 we have plotted the prior and posterior densities of $d_{H_0}(p,q)$ when $p \sim \text{Dirichlet}_R(1,\ldots,1)$ and $q \sim \text{Dirichlet}_K(1,\ldots,1)$ based upon Monte Carlo samples from the prior and posterior of $N = 10^5$. Note that the prior and posterior densities are very similar reflecting the fact that we have a very small amount of data in this example. Irrespective of that, it is the difference between these distributions that is of importance as represented by the relative belief ratio $\pi_{d_{H_0}}(d_{H_0}(p,q)\,|\,f,g)/\pi_{d_{H_0}}(d_{H_0}(p,q))$ as graphed in Figure 2. From both plots it would appear that the posterior is concentrating closer to H_0 than the prior. The relative belief ratio at 0 is $\pi_{d_{H_0}}(0\,|\,f,g)/\pi_{d_{H_0}}(0) = 1.252$. To assess the model formally we compute the P-value (10) which equals 0.63 and so we have no evidence against H_0. Therefore, we have no evidence against the hypothesis that scale I has been collapsed into scale II and can now proceed to inference about the specific collapsing, as we have already discussed.

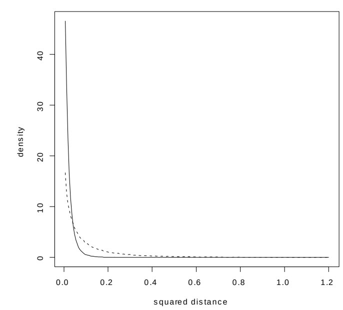

Figure 1.The prior (- -) and the posterior (–) densities of d_{H_0} in Example 5.

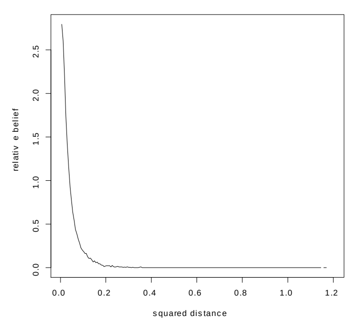

Figure 2. The relative belief ratios of d_{H_0} in Example 5.

6 Real Data Examples

We now consider some practical examples.

Example 6. *Satisfaction with cars survey.*

This example concerns a pan-European survey on driver satisfaction with their cars. The survey was done in 2001 by TNS, a well-known European marketing analytics company. Data were given to Zvi Gilula on condition that if used for publication, no disclosure of the make of the underlying cars is permitted. The survey was carried out on 10 different compact cars made in Europe and Japan and was done independently in 5 countries (Italy, Germany, Spain, UK, and France). The same 10- category scale was originally used and then was adapted for each country to allow ranking of cars by their distribution of satisfaction following the methodology reported by Yakir and Gilula (1999).

In France, a 5-category scale was used while in Germany a 3-category scale was used. The following satisfaction data were obtained based on samples of 627 in France and 501 in Germany on a car that is not made in any of these countries.

1	2	3	4	5
19	19	107	225	257

Table 8: Car satisfaction data for scale I.

1	2	3
5	381	115

Table 9: Car satisfaction data for scale II.

In Figure 3 we have plotted the prior and posterior densities for the squared distance used in model checking. In this case the P-value is effectively 0 and we have strong evidence against collapsing. This suggests that French and German respondents have fundamentally different attitudes (possibly cultural) towards this make of car.

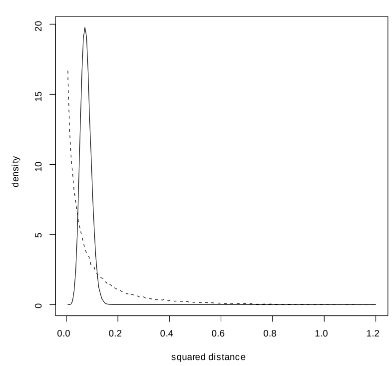

Figure 3. The prior (- -) and the posterior (–) densities of d_{H_0} in Example 6.

Example 7. *Student evaluation data.*

Students' instructor evaluation at the Hebrew University has been done for years on a 20-category scale. In the year 2000, the provost established a professional committee to re-examine the entire evaluation methodology and come up with recommendations for improvement. One major recommendation was to replace the original scale by an 11-category scale. An experiment was carried out by one of us (Gilula) on second year BA students in psychology taking a course in "research methods". The class had about 140 students and 105 of them filled the regular 20-category form. A sample of anonymous 43 of them later filled in the 11-category form. Although the two underlying samples are not fully independent, we used our proposed methodology for illustrative purposes. The data appears in the following tables.

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
3	3	0	0	3	0	3	0	3	0
11	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
3	0	6	6	6	0	12	6	12	39

Table 10: Student evaluation data for scale I.

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
2	1	1	0	1	1	3	6	5	7	16

Table 11: Student evaluation data for scale II.

There are $\binom{19}{10} = 92,378$ possible collapsings. With this number the computation time for the model checking is considerable but the computations for inference about the true collapsing are still very fast. In Figure 4 we have plotted the prior and posterior densities for the squared distance used in model checking. We see that the posterior has concentrated much more closely about 0 than the prior. In fact the P-value here is 0.967 so there is clearly no evidence against the model.

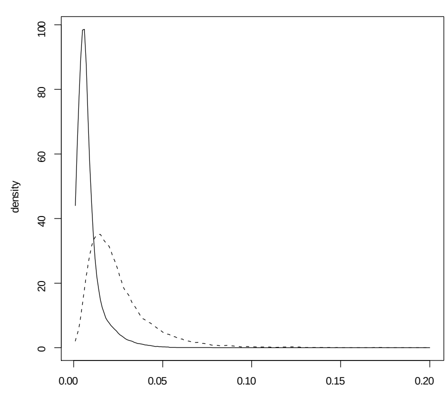

Figure 4. The prior (- -) and the posterior (–) densities of d_{H_0} in Example 7.

We then proceeded to make inference about which of the possible collapsings was appropriate. Experience with these surveys lead to the hypothesis that an appropriate collapsing was given by $(4, 5, 7, 8, 10, 11, 13, 15, 17, 19, 20)$. In Figure 5 we have plotted the relative belief ratios for each of the collapsings where they are labelled sequentially according to the ordering we discussed in Example 1. The hypothesized collapsing corresponded to #86388, with relative belief ratio equal to 45.477, while the LRSE corresponds to collapsing #81260, with relative belief ratio equal to 47.987. So the hypothesized collapsing is well supported and in fact the P-value equaled .991 when we tested the null hypothesis that this is the true collapsing. The LRSE collapsing is given by $(3, 5, 7, 8, 9, 11, 13, 16, 17, 19, 20)$ which we see is similar to the hypothesized collapsing. Note that because we are using uniform priors, the LRSE is also the posterior mode.

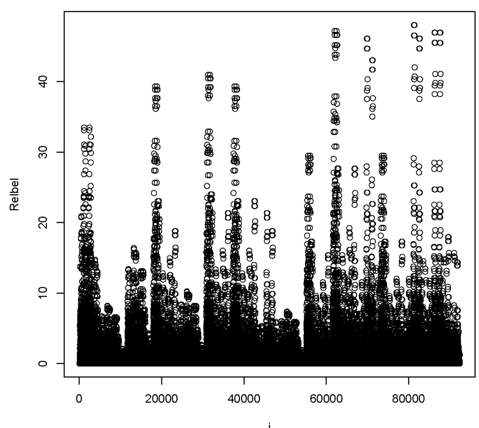

Figure 5: The relative belief ratios of the collapsings in Example 7.

7 Conclusions

We have formalized the problem of collapsing one categorical ordinal scale into another as an inference problem. Inferences then proceed by placing a prior on the parameter for the finer scale and a prior on the set of collapsings and applying Bayesian methodology. We have developed a test to assess whether or not a collapsing makes sense for a given data set. Furthermore, we have discussed how one can convert one scale into another and assess the uncertainty in such a conversion. The authors will be happy to provide the R code that was used to implement the calculations.

References

Box, G.E.P. and Tiao, G.C. (1973) Bayesian Inference in Statistical Analysis. John Wiley and Sons, New York.

Evans, M., Gilula, Z., and Guttman, I. (1993) Computational issues in the Bayesian analysis of categorical data : loglinear and Goodman's RC model. Statistica Sinica, 3, 391-406.

Evans, M., Gilula, Z., Guttman, I., and Swartz, T. (1997) Bayesian analysis of stochastically ordered distributions of categorical variables. Journal of the American Statistical Association, Vol. 92, No. 437, 208-214.

Evans, M., Guttman, I., and Swartz, T. (2006). Optimality and computations for relative surprise inferences. Canadian Journal of Statistics, Vol. 34, No. 1, 113-129.

Evans, M., and Shakhatreh, M. (2008). Optimal properties of some Bayesian inferences. Electronic Journal of Statistics, Vol. 2, 1268-1280.

Green, P.E., and Rao, V.R. (1970). Rating scales and information recovery - How many scales and response categories to use? Journal of Marketing, 34, 33-39.

Miller G.A. (1956) The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 63, 81-97.

Rossi, P.E, Gilula, Z. and Allenby, G.M. (2001). Overcoming scale usage heterogeneity: a Bayesian hierarchical approach. Journal of the American Statistical Association, 96, 453, 20-31.