Special Issue Paper

Environmetrics

Received: 31 October 2012, Revised: 3 January 2013, Accepted: 10 January 2013, Published online in Wiley Online Library: 25 February 2013

(wileyonlinelibrary.com) DOI: 10.1002/env.2202

Bayesian inference about odds ratio structure in ordinal contingency tables†

M. Kateri[a]* and A. Agresti[b]

When the goal of a study is to compare two groups on an ordinal categorical scale, a large number of inferential methods are available. Most methods are designed to detect a location effect, such as by focusing a single-degree-of-freedom test on an effect parameter. Often, rather than merely summarizing by a P-value to describe the evidence against a null hypothesis, it is of interest to consider whether a stronger conclusion can be made. For example, can we conclude that the population distributions are stochastically ordered? For parameter space regions described by order restrictions, frequentist methods are not well designed for significance testing. For example, a frequentist P-value for testing identical distributions against an alternative of stochastically ordered distributions can be very small even when the sample distributions give clear evidence that the distributions do not have the ordering property. The Bayesian approach seems better equipped to handle such questions. We discuss this in the context of stochastic ordering and other types of ordinal odds ratio structure, for the two-group comparison and for more general contexts. For Dirichlet priors, simple simulations provide posterior probabilities of particular ordinal odds ratio structures. Copyright © 2013 John Wiley & Sons, Ltd.

Keywords: $2 \times J$ contingency tables; generalized odds ratios; monotone association; ordered categorical data; ordinal logistic models; proportional odds; stochastic orderings

1. PROLOGUE

George Casella's career was incredibly broad and deep. His research interests ranged from theoretical aspects of decision theory to methodological aspects of developing inferential statistical tools such as confidence intervals (CIs) to applied aspects focusing on substantive issues in several disciplines, most notably the environment and genetics. His book publications were likewise amazingly diverse. How many statisticians could publish state-of-the-art books on as wide a variety of topics as statistical theory, variance components, experimental design, Monte Carlo methods, and statistical genomics?

One admirable trait of George's was his non-dogmatic approach to an appropriate framework for statistical inference. He used all of frequentist, Bayes, and empirical Bayes approaches throughout his career, according to what made the most sense for the particular problem he was considering. With his tutorial articles and textbook on Markov chain Monte Carlo (MCMC) methods, he helped to make many statisticians feel more comfortable using Bayesian methods. Both of us (Agresti and Kateri) have primarily used the frequentist framework in our own research, but we both have also occasionally found Bayesian methods more relevant. It is in the spirit of so much of George Casella's work that we contribute this research article, presenting a solution to a problem that can be addressed with traditional frequentist methods, but which seems better suited to be addressed with a Bayesian approach.

The methods we will discuss pertain to determining whether an association between an ordinal categorical response variable and a categorical explanatory variable satisfy any of a set of possible ordinal structures. To explain the basic question addressed in a non-technical manner, we first consider Table 1, based on data from the 2010 General Social Survey in the USA. This probability sample is administered every two years and asks respondents their opinions about a wide variety of issues. In 2010, one module asked questions with an environmental focus, such as "How concerned are you about environmental issues?" "Would you be willing to pay higher taxes to help the environment?" and "Do you think that a rise in the world's temperature is dangerous for the environment?" You can see a long list of the questions asked at http://sda.berkeley.edu/D3/GSS10/Doc/gs10.htm. At that site, it is easy to construct tables of the form of Table 1 to investigate associations and to track changes in opinions over time.

As part of this module, respondents were asked whether they agreed or disagreed with the statement, "Many of the claims about environmental threats are exaggerated." This was measured with the scale (strongly agree, agree, neither agree nor disagree, disagree, strongly disagree). Table 1 shows results, but here, for simplicity, we combine the two agree categories and combine the two disagree categories. Table 1 cross classifies responses by the respondent's political party affiliation. Inspection of sample percentages on the response, also shown in this table, suggests that Republicans tend to be most in agreement with this statement, and Democrats the least in agreement.

* Correspondence to: M. Kateri, Institute of Statistics, RWTH Aachen University, D-52056 Aachen, Germany. E-mail: maria.kateri@rwth-aachen.de

a Institute of Statistics, RWTH Aachen University, D-52056 Aachen, Germany

b Department of Statistics, University of Florida, Gainesville, FL 32611-8545, U.S.A

†This article is published in Environmetrics as part of a Special Issue in memory of George Casella and his contributions to environmental statistics; guest-edited by Linda J. Young, Department of Statistics and Institute of Food and Agricultural Sciences, University of Florida, Gainesville, Florida, USA.

Table 1. Results of 2010 General Social Survey about political party identification (variable PARTYID in the GSS) and opinion about whether environmental threats are exaggerated (variable GRNEXAGG)

	Environmental threats are exaggerated		
Political party	Agree	Neutral	Disagree
Republican	172	57	82
	(55.3%)	(18.3%)	(26.4%)
Independent	178	115	227
	(34.2%)	(22.1%)	(43.7%)
Democrat	111	78	283
	(23.5%)	(16.5%)	(60.0%)

Can we infer that in the corresponding population, this is also true, with, for example, the conditional distribution for Democrats being stochastically lower than the one for Republicans? We will analyze these data in Section 5. First, however, we will consider another data set from the existing literature on order-restricted methods that we will re-analyze, to help us connect our proposed method with that substantial literature.

2. INTRODUCTION

Table 2, analyzed by Agresti and Coull (1998, 2002), is a 4×5 contingency table resulting from a clinical trial regarding the outcome for patients who experienced trauma due to subarachnoid hemorrhage. The response categories are ordered, ranging from "death" to "good recovery." The table has four ordered treatment groups, corresponding to three dose levels of a medication and a vehicle infusion serving as a control group. A study objective was to determine whether a more favorable outcome tends to occur as the dose increases. We first consider the 2×5 table resulting from combining the three dose levels, and let Y denote the response and x an indicator for dose ($x = 1$) versus placebo ($x = 0$). The two rows have sample conditional distributions, in percentages, (28, 12, 22, 23, 15) and (23, 6, 25, 29, 17). These are stochastically ordered in favor of the dose. Standard methods provide strong evidence against the null hypothesis of identical response distributions. For example, for the adjacent-categories logit model with proportional odds structure (Agresti, 2010, Sec. 4.1),

$$\text{logit}[P(Y = j + 1)/P(Y = j)] = \alpha_j + \beta x,$$

the likelihood-ratio test of H_0: $\beta = 0$ has P-value 0.012 for the alternative H_a: $\beta > 0$ by which the response tends to be better under the dose than under the placebo.

A characteristic of this logit model is that the local log odds ratios (defined in equation (1)) are uniformly equal to β. But, of course, a small P-value for H_a: $\beta > 0$ does not imply that the local log odds ratios are uniformly nonnegative. Suppose we actually were interested in making an inference about whether the true local log odds ratios were uniformly nonnegative, regardless of whether a model such as this one is valid. We could test H_0 against H_a that this ordinal log odds ratio structure holds. This sort of order-restricted test was considered in a survey article by Agresti and Coull (2002). For this 2×5 table, they reported a likelihood-ratio test statistic of 7.9. They used an exact conditional distribution for the test statistic (i.e., conditioning on the column marginal totals, which are sufficient for the unknown common multinomial probabilities under H_0), and reported that the P-value is 0.012. This test suggests that the data are more consistent with uniformly nonnegative local log odds ratios than they are with identical conditional distributions. However, as discussed in Agresti and Coull (2002), this small P-value does not enable us to conclude that the true distributions have uniformly nonnegative log odds ratios. In fact, in the sample 2×5 table, two of the four local log odds ratios are negative. So, how can one summarize the evidence in favor of this condition?

Table 2. Cross classification of treatments and extent of trauma due to subarachnoid hemorrhage

	Outcome				
Treatment group	Death	Vegetative state	Major disability	Minor disability	Good recovery
Placebo	59	25	46	48	32
Low dose	48	21	44	47	30
Medium dose	44	14	54	64	31
High dose	43	4	49	58	41

The main point of our article is that the Bayesian approach is natural for such a conclusion. For any particular prior distributions for the two sets of multinomial probabilities, we can find the posterior probability that the true distributions have a particular structure such as uniformly nonnegative log odds ratios. For this 2×5 table, with uniform Dirichlet priors, this posterior probability is only 0.014. We will see that there is much stronger evidence in favor of another ordinal odds ratio structure.

In Section 3, we summarize ordinal odds ratio structures that are of interest for ordinal contingency tables. In Section 4, we present the Bayesian approach, illustrating in Section 5 for various structures for this 2×5 table and the complete 4×5 table, as well as for Table 1. Section 6 discusses frequentist methods for summarizing the evidence about the ordinal structure. Section 7 concludes by mentioning extensions of such Bayesian inferences for alternative structures in contingency tables.

There is quite a large literature on order-restricted methods, as surveyed for contingency tables by Agresti and Coull (2002) and as considered in more general contexts in Silvapulle and Sen (2005). More recent literature includes Bartolucci and Scaccia (2004), who considered exact conditional tests for positive association alternatives, Forcina and Bartolucci (2004), who modeled ordinal quality of life data using item response type models for which hypotheses of interest are expressed partly by inequality constraints, Iliopoulos *et al*. (2007, 2009), who considered Bayesian order-restricted inference for Goodman's association models, and Laudy and Hoijtink (2007), who used a Bayesian approach for estimation and testing when cell probabilities satisfy inequality constraints. Most recently, Bartolucci *et al.* (2012) developed a Bayesian model selection approach within a class of marginal models for categorical variables that are formulated through equality or inequality constraints on various types of log odds ratios. The focus of such articles differs from ours in that they assume order-restricted models in null or alternative hypotheses, whereas we consider unrestricted models and focus on posterior evidence supporting order restrictions of various types. Of particular relevance is a classic article by Altham (1969) that showed how to find a posterior probability of a stochastic ordering for two groups. Also, Bhattacharya and Nandram (1996) and Evans *et al.* (1997) considered such probabilities and proposed Bayes factors for comparing stochastic ordering to a lack of order restrictions.

3. ORDINAL ODDS RATIO STRUCTURE IN TWO-WAY CONTINGENCY TABLES

For an $I \times J$ contingency table of ordinal classification variables X (rows) and Y (columns), often, whatever association exists is expected to be monotone. Monotone associations can be expressed in terms of various types of odds ratios that reflect the ordering of the rows and columns.

The distribution underlying an $I \times J$ probability table can be equivalently expressed by a set of $(I-1)\times(J-1)$ odds ratios, expressed in terms of the joint distribution or conditional distributions of one variable given the other. Here, we will focus on conditional distributions, as most applications such as Tables 1 and 2 have a clear distinction between response and explanatory variables. For $i = 1, \ldots, I$, let

$$\pi_{j|i} = P(Y = j|X = i),\ j = 1, \ldots, J.$$

The most popular such sets of ordinal odds ratios are the *local odds ratios*,

$$\theta_{ij}^{L} = \frac{\pi_{j|i}/\pi_{j+1|i}}{\pi_{j|i+1}/\pi_{j+1|i+1}} \tag{1}$$

the *cumulative odds ratios*

$$\theta_{ij}^{C} = \frac{\left(\sum_{k\leqslant j}\pi_{k|i}\right)/\left(\sum_{k>j}\pi_{k|i}\right)}{\left(\sum_{k\leqslant j}\pi_{k|i+1}\right)/\left(\sum_{k>j}\pi_{k|i+1}\right)} \tag{2}$$

and the *global odds ratios*

$$\theta_{ij}^{G} = \frac{P(Y\leqslant j \mid X\leqslant i)/P(Y>j \mid X\leqslant i)}{P(Y\leqslant j \mid X>i)/P(Y>j \mid X>i)} \tag{3}$$

for $i = 1, \ldots, I-1,\ \ j = 1, \ldots, J-1$. Less common are *continuation odds ratios*, such as

$$\theta_{ij}^{CO} = \frac{P(Y=j \mid X=i)/P(Y>j \mid X=i)}{P(Y=j \mid X=i+1)/P(Y>j \mid X=i+1)} \tag{4}$$

Separate continuation odds ratios result from reversing the order of Y categories. Another type of continuation odds ratio refers to 2×2 tables that provide decompositions of independent chi-squared statistics,

$$\theta_{ij}^{CO2} = \frac{P(Y=j \mid X=i)/P(Y>j \mid X=i)}{P(Y=j \mid X>i)/P(Y>j \mid X>i)} \tag{5}$$

The local, cumulative, and first type of continuation odds ratio are also relevant when the explanatory variable is nominal, for example, for comparing any pair of levels of X (not only the adjacent ones).

Agresti and Coull (1998, 2002) and Douglas *et al.* (1990) presented these odds ratios and their properties and provided many references of their use. The most popular model for ordinal responses uses logits of cumulative probabilities (so-called *cumulative logits*) and uses cumulative odds ratios to interpret effects. Goodman (1981) noted that the local odds ratios describe association models that fit well when there is an underlying bivariate normal distribution. The continuation-ratio logit model is useful when a sequential mechanism, such as survival through various age periods, determines the response outcome (e.g., Tutz, 1991), so that the ordering of response categories is relevant in one direction but not in the other. These and most other models for ordinal responses focus on location effects and imply stochastic orderings of the response distributions at various levels of explanatory variables. They can be derived by latent variable models for underlying continuous response variables. For example, cumulative logit and cumulative probit models result from underlying standard regression models that assume logistic and normal response distributions with constant variance.

The analyses we consider in this article can be regarded as referring to a more general latent variable model that merely implies nonnegative log odds ratios for certain ordinal structures. For Table 2, for example, the model of nonnegative global log odds ratios corresponds to a latent structure, which says that no matter how we collapse the explanatory variable into dose above versus below some level, and no matter how we collapse the response variable into outcome better than versus worse than some particular outcome, the resulting 2×2 table has a nonnegative log odds ratio. The model of nonnegative cumulative log odds ratios corresponds to a latent structure that says that no matter how we take a pair of dosage levels (one of which may be placebo) and no matter how we collapse the response variable into outcome better than versus worse than some particular outcome, the resulting 2×2 table has a nonnegative log odds ratio.

Lehmann (1966) discussed several types of bivariate structure for continuous variables. When the variables are collapsed in certain ways to discrete ones, these structures are equivalent to uniformly nonnegative or uniformly nonpositive values of the log odds ratios. The best known type of stochastic ordering is the *simple stochastic ordering*, whereby the distribution of Y when $X = x$ is stochastically larger than the distribution of Y when $X = x'$ if

$$P(Y \leqslant y \mid X = x) \leqslant P(Y \leqslant y \mid X = x'), \text{ all } y.$$

In the discrete case with $x = i + 1$ and $x' = i$, this is equivalent to uniformly nonnegative cumulative log odds ratios. A stronger condition is *likelihood-ratio ordering*, which in the discrete case corresponds to uniformly nonnegative local log odds ratios. A weaker condition is *positive quadrant dependence*, which in the discrete case corresponds to uniformly nonnegative global log odds ratios. Yet another condition is the *hazard rate ordering*, which in the discrete case corresponds to the first type of continuation log odds ratio being uniformly nonnegative. The various possible stochastic orders are related. For example, in the discrete case,

$$\text{all } \log\left(\theta_{ij}^{L}\right) \geqslant 0 \Rightarrow \text{all } \log\left(\theta_{ij}^{CO}\right) \geqslant 0 \Rightarrow \text{all } \log\left(\theta_{ij}^{C}\right) \geqslant 0 \Rightarrow \text{all } \log\left(\theta_{ij}^{G}\right) \geqslant 0$$

(Douglas *et al.*, 1990). We denote the conditions of uniformly nonnegative log odds ratios by L for local odds ratios, CO for continuation odds ratios, C for cumulative odds ratios, and G for global odds ratios.

Assuming independent multinomial sampling in the I rows, Agresti and Coull (1998, 2002) surveyed the literature for finding ML estimates of cell probabilities and for testing independence against the alternative corresponding to any particular ordinal odds ratio structure. The asymptotic chi-bar-squared null distributions of such order-restricted test statistics are awkward, involving complex mixtures of chi-squared variates with unknown weights for the various degrees-of-freedom (df) values. The lack of software for such analyses may be a reason such methods seem to be rarely used, and they proposed simulating exact conditional tests for the likelihood-ratio statistic. For example, for the full version of Table 2, they noted that the likelihood-ratio statistics are 16.1 for alternative L, 27.7 for alternative C, and 27.8 for alternative G. Each test has exact conditional P-value $\leqslant 0.002$.

Agresti and Coull pointed out that a small P-value does not suggest that the order restriction truly holds, but only that strong evidence exists against the null hypothesis of independence, based on that test criterion. In fact, it is possible to obtain small P-values even if many of the sample odds ratios violate the inequality constraints. For Table 2, for example, two of the 12 sample cumulative log odds ratios are negative, and five of the sample local log odds ratios are negative, so it is not sensible to conclude that the population satisfies condition C or L. So, a small P-value in an order-restricted test does not suggest that the order restriction truly holds, but merely that the criterion provides strong evidence against the null. An analogy occurs in regression analysis, in which a test based on the slope of a straight line may provide strong evidence against independence, even though we would not want to conclude that the true relationship is exactly linear.

So, given that sample data sometimes fall quite far from either the null hypothesis or an order-restricted alternative, how do we summarize the strength of evidence in favor of a particular ordinal structure? As noted in the literature (e.g., Laudy and Hoijtink, 2007), asymptotic distribution theory is complex for testing a constrained model against an unconstrained or constrained alternative. Although frequentist inference is not well suited for this, in Section 6 we will discuss such methods for summarizing the evidence. This seems to be a context, however, in which Bayesian solutions are more appealing. One simple such solution is described next.

4. BAYESIAN POSTERIOR PROBABILITIES OF NONNEGATIVE ORDINAL ASSOCIATION

We assume that each row of the table follows an independent multinomial distribution, that is,

$$\mathbf{n}_i = (n_{i1}, n_{i2}, \ldots, n_{iJ}) \sim \mathcal{M}(n_{i\cdot}, \pi_i), \quad i = 1, \ldots, I,$$

where π_i is the vector of conditional probabilities in row i. It is convenient to assume conjugate Dirichlet priors. Thus, we set

$$\pi_i \sim \mathcal{D}(\alpha_i), \quad i = 1, \ldots, I,$$

with hyperparameter vector $\alpha_i = (\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{iJ})$. The posterior then is

$$\pi_i \mid \mathbf{n}_i \sim \mathcal{D}(\mathbf{n}_i + \alpha_i), \quad i = 1, \ldots, I.$$

Priors with $\alpha_{ij} = c$, for all i, j, are non-informative in terms of association directions, satisfying $\theta_{ij} = 1, i = 1, \ldots, I-1, j = 1, \ldots, J-1$, for all types of ordinal odds ratios. Commonly used special cases are the uniform distribution ($c = 1$) and the Jeffreys prior ($c = 0.5$).

One could, as in Laudy and Hoijtink (2007), limit the prior to the part of the parameter space satisfying the order restriction of interest, but we shall not do so, as we do not wish to assume such structure, but merely find the posterior probability supporting it. That is, on the basis of the posterior Dirichlet distribution, we are interested in the posterior probability of a condition such as L or C. This merely involves integrating the posterior density over the relevant part of the parameter space.

In the case $I = 2$ of comparing two groups, the conditions C and G are equivalent and were considered by Altham (1969, Sec. 5 and 6). For symmetric Dirichlet priors with $\alpha_{ij} = c$, for all i, j, the prior probability that all cumulative log odds ratios are nonnegative is $P(C) = 1/J$ (Altham, 1969, Equation (8)). By the independence of the $J-1$ 2×2 tables formed for the continuation odds ratios, $P(CO) = \left(\frac{1}{2}\right)^{J-1}$, and this prior probability bounds above $P(L)$. For the posterior Dirichlet for a 2×J table, using the equivalence of a Dirichlet random variable with differences between certain order statistics from a uniform distribution over [0, 1], Altham found an expression for the posterior probability of condition C. Let $\{\mu_j = n_{1j} + \alpha_{1j}\}$ and $\{\nu_j = n_{2j} + \alpha_{2j}\}$. Let $\mu = (\mu_1 + \cdots + \mu_J)$ and $\nu = (\nu_1 + \cdots + \nu_J)$ be the effective posterior sample sizes. The posterior probability that group 2 is stochastically larger than group 1 equals

$$P(C \mid \mathbf{n}) = \sum_{s_1} \cdots \sum_{s_c} \frac{\binom{\mu_1+\nu_1-1}{s_1}\binom{\mu_2+\nu_2}{s_2}\cdots\binom{\mu_{c-1}+\nu_{c-1}}{s_{c-1}}\binom{\mu_c+\nu_c-1}{s_c}}{\binom{\mu+\nu-2}{\nu-1}},$$

where each s_j index varies between 0 and the upper limit in the corresponding binomial coefficient, but such that $(s_1 + \cdots + s_j) \leqslant (\mu_1 + \cdots + \mu_j - 1)$ for $1 \leqslant j \leqslant J-1$. For fixed (μ, ν) and fixed $(\mu_j + \nu_j)$ for each j, this posterior probability is monotone increasing as $(\mu_1 + \cdots + \mu_j)$ increases for any $j < J$, that is, as relatively more posterior probability falls at the low end of the scale for group 1. Weisberg (1972) proposed an algorithm for computing $P(C \mid \mathbf{n})$.

Altham's result has the appealing aspect of providing a posterior probability of a simple stochastic ordering. So, if this probability is close to 1, you can conclude that there truly is a stochastic ordering in the population of interest. Her derivation does not seem to extend to alternative types of ordering (such as L) for 2×J tables or to $I \times J$ tables. However, with Dirichlet priors, the posterior probabilities of the conditions can be simulated simply and very accurately, as follows. We simulate a large number R (say 1,000,000) of Dirichlet random vectors $\mathcal{D}(\mathbf{n}_i + \alpha_i)$ for each group i and form the R corresponding $I \times J$ contingency tables with row marginals $n_{i\cdot} + \alpha_{i\cdot}$ $(i = 1, \ldots, I)$. Let A_L, A_{CO}, A_{CO2}, A_C and A_G be the subsets of the R simulated tables for which the distributions have uniformly nonnegative log odds ratios of the associated type, so that $A_L \subseteq A_{CO} \subseteq A_C \subseteq A_G$ and $A_L \subseteq A_{CO} \subseteq A_{CO2} \subseteq A_G$. The sample proportions of cases satisfying the various conditions converge as $R \to \infty$ to the true posterior probabilities that satisfy

$$P(L \mid \mathbf{n}) \leqslant P(CO \mid \mathbf{n}) \leqslant P(C \mid \mathbf{n}) \leqslant P(G \mid \mathbf{n}),$$
$$P(L \mid \mathbf{n}) \leqslant P(CO \mid \mathbf{n}) \leqslant P(CO2 \mid \mathbf{n}) \leqslant P(G \mid \mathbf{n}).$$

In practice, the prior and posterior probabilities of non-negativity for the various types of log odds ratios are often relatively small and tend to be smaller as I and J increase. (Recall that even for a 2×J table, an uninformative prior has $P(C) = 1/J$, and $P(L)$ is considerably smaller.) For larger tables with data more sparsely distributed over the cells, it tends to become more difficult to establish a condition such as L, even when all the sample odds ratios satisfy it. The condition G, using all data for each log odds ratio and having smaller standard errors for them, would tend to be more robust in this sense. In particular, by the relative sizes of the various posterior probabilities, in practice it can be quite difficult for this Bayesian analysis to establish property L.

When there is very strong evidence that a particular ordinal log odds ratio is uniformly nonnegative, it is still relevant to judge whether the association that seems to exist is practically significant. One way to assess this is to find the posterior probability that the log odds ratios are uniformly larger than some threshold value. For example, if treatment is considered practically better than placebo when the odds of responding above some point instead of below it are 20% higher for treatment than placebo, then one could report the posterior probability that all the cumulative odds ratios exceed 1.20. Again, for the Dirichlet prior structure, it is easy to simulate from the posterior density to estimate precisely such a probability.

Another analysis that can also be useful considers the effect of the data on the prior belief of a particular ordinal structure. For example, a Bayes factor evaluating the change in the odds in favor of a particular ordering type OT is

$$BF = \frac{P(OT \mid \mathbf{n})/[1 - P(OT \mid \mathbf{n})]}{P(OT)/[1 - P(OT)]},$$

where $P(OT)$ is the corresponding prior probability. Likewise, the change in the odds in favor of an ordering type OT_1 relative to another type OT_2 is summarized by the Bayes Factor (BF)

$$BF_{12} = \frac{P(OT_1 \mid \mathbf{n})/P(OT_2 \mid \mathbf{n})}{P(OT_1)/P(OT_2)}.$$

These can also be easily simulated.

5. EXAMPLES

We now return to the 2×5 collapsing of Table 2 referred to in the introductory section. The sample ($\log(\hat{\theta}_{11})$, $\log(\hat{\theta}_{12})$, $\log(\hat{\theta}_{13})$, $\log(\hat{\theta}_{14})$) values are (0.28, 0.47, 0.32, 0.15) for the cumulative and global log odds ratios, (0.28, 0.75, 0.06, −0.10) for the continuation log odds ratios, and (−0.38, 0.72, 0.10, −0.10) for the local log odds ratios. Using Altham's expression, $P(C \mid \mathbf{n}) = 0.705$ with the uniform Dirichlet prior and 0.709 with the Jeffreys prior. For the narrower types of stochastic ordering, the corresponding probabilities are estimated with 1,000,000 simulations as $P(CO \mid \mathbf{n}) = 0.192$ and 0.196 and $P(L \mid \mathbf{n}) = 0.014$ and 0.015. The Monte Carlo standard errors are bounded above by 0.001 for this many simulations. It seems plausible that the two distributions are truly stochastically ordered, but unlikely that the narrower L condition holds. For a more detailed summary, one can use the simulations to plot the full posterior distribution or to construct a credible interval for each of the odds ratios that constitute a particular ordinal structure.

The standard models characterized by cumulative odds ratios and local odds ratios are the cumulative logit models and the adjacent-categories logit models. With proportional odds structure for the binary group predictor, they imply uniform values for such odds ratios. Both such models fit these data adequately, with deviance 4.77 for the cumulative logit model and 4.56 for the adjacent-categories logit model (df = 3). So, given either model, the observed data would not be unusual. However, given the data, the structure that is more general than the cumulative logit model by which the cumulative log odds ratios are uniformly nonnegative is not suspect but the structure that is more general than the adjacent-categories logit model by which the local log odds ratios are uniformly nonnegative *is* suspect.

For the uniform Dirichlet prior ($c = 1$), the prior probabilities of the various ordinal structures are $P(C) = 0.200$, $P(CO) = 0.0625$, and $P(L) = 0.008$. Hence the Bayes factors describing the change in evidence supporting each of these are estimated by $BF(C) = 9.6$, $BF(CO) = 3.6$, and $BF(L) = 1.7$. The Bayes factors describing the change in odds for conditions CO (2) and L (3) relative to the most likely condition C (1), are $BF_{21} = 0.9$ and $BF_{31} = 0.5$.

We next analyze the full 4×5 version of Table 2. For uniform priors and 1,000,000 simulations, the estimated posterior probabilities are $P(L \mid \mathbf{n}) = 0.000$, $P(CO \mid \mathbf{n}) = 0.0002$, $P(C \mid \mathbf{n}) = 0.018$, and $P(G \mid \mathbf{n}) = 0.519$. The small value for $P(L \mid \mathbf{n})$ is not surprising, since five of the 12 sample odds ratios violate condition L. All the sample global log odds ratios are positive, and it seems plausible that this is also true in the population.

Although the posterior probabilities of a positive association have been small for most conditions in this example, they need not be. When the sample data show a strong association according to such criteria, and when the sample size is large so that such association is unlikely to be due to sampling error, such posterior probabilities can be large.

For example, consider Table 1 on responses to the statement about whether many claims about environmental threats are exaggerated. The General Social Survey sample is more complex than a simple random sample, but we use the data for illustrative purposes under such an assumption. For these data, for each ordinal structure, each sample log odds ratio is positive and is large relative to sampling error except for one of the local log odds ratios. With uniform Dirichlet priors and using 1,000,000 Monte Carlo samples, $P(L \mid \mathbf{n}) = 0.640$, $P(CO \mid \mathbf{n}) = 0.937$, $P(C \mid \mathbf{n}) = 0.9999$, and $P(G \mid \mathbf{n}) = 1.0000$. Thus, there is very strong evidence in favor of simple stochastic ordering and positive quadrant dependence, but the evidence for the narrowest condition of likelihood-ratio ordering is much less strong. For conditions C and G, the association also seems to be substantively strong. For example, in a comparison of Republicans and Democrats, the sample cumulative log odds ratios are 1.39 and 1.43, and the posterior probability that both cumulative log odds ratios exceed 1.0 is estimated to be 0.991 (based on 1,000,000 Monte Carlo samples).

The uniform positivity of the local odds ratios is not strongly supported, but one can also consider this probability for each individual probability. Table 3 reports these. It shows that there is quite strong evidence that $\log\left(\theta_{11}^L\right) > 0$ and $\log\left(\theta_{22}^L\right) > 0$. The most questionable

Table 3. Ninety-five per cent equal-tail credible intervals for the log local odds ratios $\log\left(\theta_{ij}^L\right)$, $i, j = 1, 2$, of the example in Table 1 (based on 1,000,000 generated tables and for prior parameters $\alpha_j = 1$ ($j = 1, 2, 3$) for all rows of the table), along with their simulated coverage probabilities (based on 1,000,000 simulations) and the posterior probabilities of negative log local odds ratios

	95% equal-tail CI		Coverage Probability	$P\left[\log\left(\theta_{ij}^L\right) > 0 \mid \mathbf{n}\right]$
$\log\left(\theta_{11}^L\right)$	0.2869	1.0459	0.9473	0.9998
$\log\left(\theta_{12}^L\right)$	−0.0884	0.7204	0.9474	0.9374
$\log\left(\theta_{21}^L\right)$	−0.2868	0.4553	0.9470	0.6726
$\log\left(\theta_{22}^L\right)$	0.2731	0.9427	0.9474	0.9998

CI, credible intervals.

positivity case is $\log\left(\theta_{21}^L\right)$, for which the posterior $P\left(\log\left(\theta_{21}^L\right) < 0\right)$ is estimated to be 0.327. Table 3 also reports 95% credible intervals for these log odds ratios. As a check on the adequacy of such intervals, we also conducted simulations that treated the sample distributions as if they were the true population values and conducted simulations to estimate the probability that the interval captured the true parameter; these were found to perform well, as illustrated in the table.

Our approach is computationally easily feasible, and the corresponding computation times are not of worrying length. For example, on an Intel Core 2 Duo P9500 processor (2.53 GHz, 3.48 GB of RAM), the procedure, based on 1,000,000 replications, for deriving for our first data set the posterior probabilities of the various types of stochastic ordering required user, system and elapsed CPU times 629.9, 0.14 and 633.8, respectively (in seconds). The procedure that additionally saved all simulated odds ratios (of all types) in order to produce posterior density plots and all inferential results for all types of odds ratios (not just for local odds ratios presented earlier), required user, system and elapsed CPU times 698.0, 0.3 and 702.3, respectively (in seconds).

In some examples, the variable defining the groups might be nominal rather than ordinal (e.g., religious affiliation, region of country of residence), and the interest might be in determining whether an ordering of those nominal categories exists for which the groups are stochastically ordered or satisfy one of the other ordinal structures we have discussed. We could then simulate the posterior probabilities for various permutations of the groups. Evans *et al.* (1997) considered such analyses for the stochastic ordering structure.

For example, in Table 1, one might prefer to treat the political party affiliations as nominal rather than ordinal, then evaluate posterior probabilities of the various ordinal structures for each of the six possible orderings of the three groups. In this case, however, none of the other five orderings are at all plausible. For example, the posterior probability of a stochastic ordering is zero to several decimal places for orderings other than the one given in the table.

When there is strong evidence that a particular ordinal structure does not hold, one can use models designed for that type of odds ratio to characterize the non-monotonicity. For example, for the L structure, it is natural to consider association models that are characterized by local odds ratios. In particular, with Goodman's (1970) RC model, possibly nonmonotone row scores $\{\mu_i\}$ and column scores $\{\nu_j\}$ are estimated for which the $\log\theta_{ij} = (\mu_{i+1} - \mu_i)(\nu_{j+1} - \nu_j)$ need not have uniformly one sign. This model is also particularly useful for a predictor that is nominal, as when the $\hat{\nu}_j$ are monotone, the order of the $\hat{\mu}_i$ suggests a possible ordering of the rows for which structure L (or other ordinal structures) may hold for the response variable. In the Evans *et al.* (1997) approach with a nominal factor, all possible permutations were considered for such orderings, whereas a modeling approach with the RC model identifies directly the permutation of the nominal factor categories that corresponds to the ordering of highest posterior probability.

6. FREQUENTIST SUMMARIZING OF INFORMATION ABOUT ORDINAL STRUCTURE

We have noted that for order-restricted alternatives, ordinary hypothesis testing that uses a narrow null hypothesis such as independence or identical distributions results in limited information. In particular, the union of H_0 and H_a is not the entire parameter space, so such inference is not well suited for determining whether an ordinal structure such as nonnegative log odds ratios of a certain type truly holds. In future research, it is of interest to develop frequentist tests of a more general null hypothesis corresponding to the complement space of the order-restricted condition against the alternative hypothesis of the order-restricted condition. One way it currently is possible to do this is with an intersection–union test Berger (1982). We are pleased to mention this approach in this article, as its author (Roger Berger) was a graduate school colleague of George Casella and later co-author of their highly regarded and influential *Statistical Inference* textbook.

For a particular type of ordinal odds ratio, one regards the null hypothesis H_0 that the order restriction does not strictly hold as the union of $(I-1)(J-1)$ events, where event k says that the kth of the log odds ratios is $\leqslant 0$. The alternative hypothesis states that the order restriction strictly holds, that is, it is the intersection event that all $(I-1)(J-1)$ of the log odds ratios are positive. To achieve overall size α, for each individual log odds ratio k we conduct an α-level one-sided test that it is $\leqslant 0$ versus the alternative that it is > 0. For example, we could use the signed square root of the Pearson statistic for the relevant 2×2 table, which is the one-sided two-sample z statistic for comparing two proportions using the pooled standard error (Agresti, 2013, p. 78). We then reject the overall H_0 if each of the $(I-1)(J-1)$ individual tests is significant. From Berger (1982), this test has large-sample size of α. Moreover, Berger (1997) noted that this test is also the likelihood-ratio test for this decomposition of the parameter space. One can regard the P-value for the test as equaling the maximum of the P-values for the individual tests.

To illustrate, consider cumulative odds ratios for the 2×5 collapsing of Table 2. Using the two-sample one-sided z test statistic, the minimum individual z score occurs for the test of $\log\left(\theta_{14}^C\right) = 0$, with $z = 0.66$ generating a P-value $= 0.25$. The corresponding Bayesian $P(C \mid \mathbf{n}) = 0.71$ for the priors we used. For Table 1, the minimum individual z score occurs for the test of $\log\left(\theta_{21}^C\right) = 0$, with $z = 3.71$ generating a P-value $= 0.0001$. With uniform prior, the corresponding Bayesian $P(C \mid \mathbf{n}) = 0.9999$. Here, we have conducted the individual tests under the null condition that $\log\left(\theta_{ij}^C\right) = 0$ and acted as if this gives the same P-value as testing under $\log\left(\theta_{ij}^C\right) \leqslant 0$, which seems intuitively reasonable, but requires further justification.

When a table truly satisfies independence, with all log odds ratios equal to 0, the size of such a test is much smaller than α (indeed $\alpha^{(I-1)(J-1)}$ if the separate tests were independent, which they are not). However, over the entire union H_0 space, one can achieve size α by letting one of the log odds ratios equal 0, but letting all the other log odds ratios extremely large. This is why nothing similar to a Bonferroni correction is needed for the individual tests in order to achieve the desired overall size.

Compared with the Bayesian approach, an unsatisfying aspect of this frequentist test is its severe discreteness. For example, if all except one of the sample log odds ratios take large positive values, the P-value is governed essentially completely by that one exception; that is, one could change the counts in the table in any way whatever keeps that one corresponding 2×2 table (for the exception) the same but for which all the other individual P-values remain smaller than the P-value for the exception, and the overall P-value of the intersection–union test will not change. Also, compared with the Bayesian approach, this approach is relying on simultaneous large-sample tests, which may not be suitable for all the 2×2 tables considered for a particular ordinal odds ratio.

It is also possible, of course, to learn something about the non-model-based association structure using frequentist estimation methods. A rather ad hoc estimation approach constructs simultaneous CIs for the entire set of $(I-1)(J-1)$ log odds ratios of interest in the $I \times J$ table. When we use the score CI for each case, each such $100(1-\alpha)\%$ CI containing only nonnegative values is equivalent to rejecting H_0 in favor of an order-restricted alternative at the α-level for the intersection–union test. When I and J are large, a simple but somewhat conservative approach that is sensible even if the order restriction does not seem to hold uses the Bonferroni method.

7. EXTENSIONS

The basic idea presented in this article extends to many other settings. For example, for the comparison of two ordinal responses, Kateri (2011) proposed a method for testing for stochastic ordering as well as for umbrella ordering and the detection of the crossing points of the two distribution functions. The approach introduced here could be extended also to find the posterior probability of a particular umbrella ordering. As another example, for multivariate ordinal responses, one could find the posterior probability that a pair of marginal distributions has a particular ordinal structure such as stochastic ordering, as an alternative to marginal homogeneity.

We have focused on structure for ordinal response variables, but similar ideas can be useful for ordinal explanatory variables with binary or nominal-scale response variables. For example, a common problem (such as in dose–response studies) deals with analyzing whether the probability of a "success" on a binary response increases as the level of an ordinal predictor increases. As a supplement to such analyses as a linear logit model and its Cochran–Armitage score test, we could find the non-model-based posterior probability that the success probability is monotone increasing.

Extensions are also possible for stratified contingency tables. However, as the number of cells increases, the non-model-based posterior probability of any particular ordinal structure may tend to be quite small.

Finally, we used simple Dirichlet priors, which provide quite a bit of flexibility yet result in a simple posterior and easy simulation of posterior probabilities. In some applications it might be of interest to use an alternative type of prior, such as a hierarchical prior or a prior that mixes a continuous prior such as the Dirichlet with one that places discrete probability mass on the set of cell probabilities for which the ordinal log odds ratios are uniformly equal to 0. This would be a reasonable approach in applications in which "no effect" is plausible, such as in many genetics investigations.

Acknowledgements

The authors are grateful to two referees who made suggestions that improved this paper. In particular, we are grateful to one referee for suggesting that an intersection–union test would be one way to approach this from a frequentist perspective.

REFERENCES

Agresti A. 2010. *Analysis of Ordinal Categorical Data*, 2nd ed. Wiley: Hoboken, NJ.
Agresti A. 2013. *Categorical Data Analysis*, 3rd ed. Wiley: Hoboken, NJ.
Agresti A, Coull B. 1998. Order-restricted inference for monotone trend alternatives in contingency tables. *Computational Statistics & Data Analysis* **28**: 139–155.
Agresti A, Coull B. 2002. The analysis of contingency tables under inequality constraints. *Journal of Statistical Planning and Inference* **107**: 45–73.
Altham PME. 1969. Exact Bayesian analysis of a 2 × 2 contingency table and Fisher's "exact" significance test. *Journal of the Royal Statistical Society, Series B* **31**: 261–269.
Bartolucci F, Scaccia L. 2004. Testing for positive association in contingency tables with fixed margins. *Computational Statistics & Data Analysis* **47**: 195–210.
Bartolucci F, Scaccia L, Farcomeni A. 2012. Bayesian inference through encompassing priors and importance sampling for a class of marginal models for categorical data. *Computational Statistics & Data Analysis* **56**: 4067–4080.
Berger RL. 1982. Multiparameter hypothesis testing and acceptance sampling. *Technometrics* **24**: 295–300.
Berger RL. 1997. Likelihood Ratio Tests and Intersection-Union Tests. In *Advances in Statistical Decision Theory and Application*, Panchapakesan S, Balakrishnan N (eds). Birkauser: Boston; 225-237.
Bhattacharya B, Nandram B. 1996. Bayesian inference for multinomial populations under stochastic ordering. *Journal of Statistical Computation and Simulation* **54**: 45–163.
Douglas R, Fienberg SE, Lee MT, Sampson AR, Whitaker LR. 1990. Positive dependence concepts for ordinal contingency tables. In *Topics in Statistical Dependence*, Insitute of Mathematical Statistics Lecture Notes. Institute of Mathematical Statistics: Hayward, CA; 189–202.
Evans M, Gilula Z, Guttman I, Swartz T. 1997. Bayesian analysis of stochastically ordered distributions of categorical variables. *Journal of the American Statistical Association* **92**: 208–214.
Forcina A, Bartolucci F. 2004. Modelling quality of life variables with non-parametric mixtures. *Environmetrics* **15**: 519–528.
Goodman LA. 1981. Association models and the bivariate normal for contingency tables with ordered categories. *Biometrika* **68**: 347–355.
Iliopoulos G, Kateri M, Ntzoufras I. 2007. *Computational Statistics & Data Analysis* **51**: 4643–4655.
Iliopoulos G, Kateri M, Ntzoufras I. 2009. Bayesian model comparison for the order restricted RC association model. *Psychometrika* **74**: 561–587.
Kateri M. 2011. On the comparison of two ordinal responses. *Communications in Statistics: Theory & Methods* **40**: 3748–3763.
Lehmann E. 1966. Some concepts of dependence. *Annals of Mathematical Statistics* **37**: 1137–1153.
Laudy O, Hoijtink H. 2007. Bayesian methods for the analysis of inequality constrained contingency tables. *Statistical Methods in Medical Research* **16**: 139–153.
Silvapulle MJ, Sen PK. 2005. *Constrained Statistical Inference*. Wiley: New-York.
Tutz G. 1991. Sequential models in categorical regression. *Computational Statistics and Data Analysis* **11**: 275–295.
Weisberg HI. 1972. Bayesian comparison of two ordered multinomial populations. *Biometrics* **28**: 859–867.

