A Bayesian credible set for ranking effect sizes

Abstract

The relative magnitude, or ranking, of multiple parameters is often more interesting or trustworthy than their absolute magnitude, or than strawman tests of their equality. Yet there is currently nothing akin to a simple $\beta\%$ confidence or credible region for this ranking. We first show how such a credible region can be encoded as a finite collection of full rankings. For example, collecting together the two disjoint parts of parameter space $\mathbb{R}^3$ satisfying full rankings $w_1 < w_2 < w_3$ and $w_2 < w_1 < w_3$ respectively, we attain a coherent (partial) ranking which we can denote $1, 2|3$. We provide a general construction for the best $\beta\%$ Bayesian posterior credible set of such rankings, henceforth denoted $\mathcal{C}_\beta$. The approach can be applied in both exact and approximate - MCMC and variational - settings and offers interpretational advantages over conventional analyses. By sidestepping the computation of marginal likelihood and Bayes factors our procedure requires neither informative priors nor informative hypotheses.

Keywords: Regression, multilevel, order theory, informative hypotheses, region of practical equivalence, encompassing priors, compositional data analysis (coda), simplex.

Word count: 3600

A Bayesian credible set for ranking effect sizes

Introduction

Scientific research often scrutinizes theoretically-inspired hypotheses via classical tests or Bayesian model comparison. But the impact of theory on data analysis is sometimes premature and questionable and there remains an important place for more open-ended or exploratory analyses. We apply this maxim to ranking parameters, see Figure 1, a task currently dominated by confirmatory methods (Gu, Mulder, Deković, & Hoijtink, 2014; Klugkist, Kato, & Hoijtink, 2005; Mulder & Olsson-Collentine, 2019; Mulder, Hoijtink, & Klugkist, 2010). To our knowledge only one exploratory scheme for general multi-parameter inequalities has been advanced (Stern, 2005), but this avenue of research appears to have been stifled by concerns that 1) scientifically impossible hypotheses should first be excluded from the search space or that 2) a greater multiple comparisons problem compromises the false positive rate (Klugkist, Laudy, & Hoijtink, 2005).

This critique is however easily overstated. In situations of genuine scientific uncertainty, it is the confirmatory approach which is more risky, because it assumes with certainty that the true data generating process is somewhere among the limited set of theoretically-inspired hypotheses. When this assumption breaks, the approach predictably incurs false positives: a hypothesis with overwhelming positive evidence over its limited theoretical competitors may nonetheless pale beneath the (omitted) true model. Posterior model probabilities generally depend on the set of models considered. Clearly with all else being equal, a hypothesis that competes well in a broader field meets a higher - not a lower - standard of evidence. Conversely, there is no logical requirement to exclude "scientifically impossible" hypotheses from consideration *a priori*: Bayesian inference automatically downgrades their posterior probability. As we show in this work, exploratory schemes can at once search a larger set of possibilities and readily control the multiple comparisons problem.

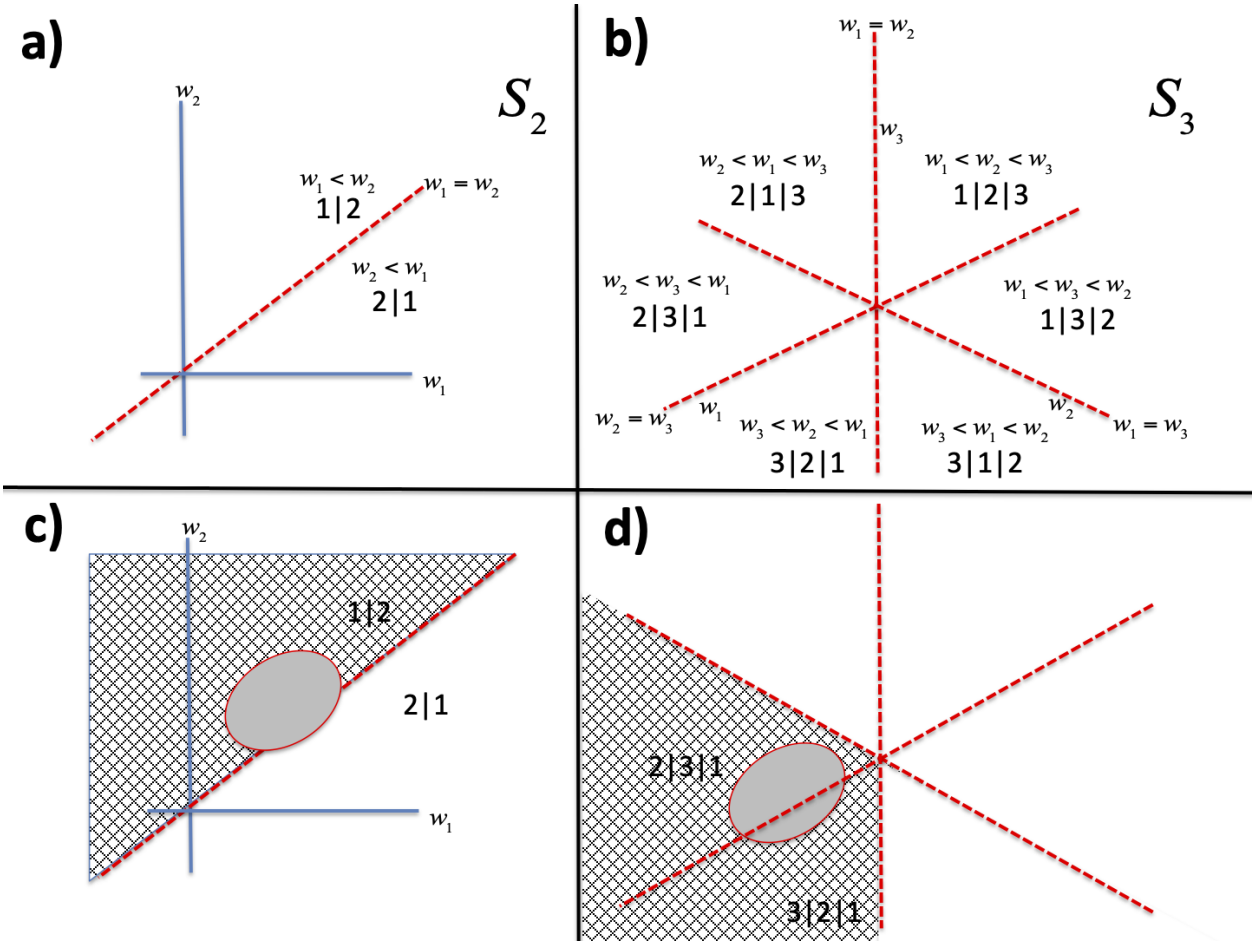

Figure 1. Full rankings in 2 and 3 dimensions. a) The two full rankings in $\mathbb{R}^2$ are reflections about $w_1 = w_2$. b) It is hard to visualize a 3 dimensional parameter space. We therefore simply illustrate a representative plane of type $\mathbb{R}^2$, orthogonal to $w_1 = w_2 = w_3$ which cuts through the full parameter space $\mathbb{R}^3$. In this transect we can see how the $6 = 3!$ full rankings in $\mathbb{R}^3$ reflect across $w_1 = w_2, w_2 = w_3$ and $w_1 = w_3$ and intersect at $w_1 = w_2 = w_3$. These equalities themselves have neglible probability under any uncertain prior, which leaves all the mass to full rankings, as discussed in the text. c) A standard 95% credible set in solid grey alongside our credible ranking 1|2 in hashed grey. d) Similarly for 3 dimensions. Here $2, 3|1$. If our parameter space is $\mathbb{R}^3$ then $2, 3|1$ is the hashed grey region which extends infinitely bottom left and projectively into the axis of gaze. Figure 2 illustrates a procedure for identifying the credible sets just discussed.

We propose simply summarizing the posterior distribution of any well-specified parametric model, e.g. ordinary linear or multilevel regression, which has been estimated either analytically or by MCMC. As with the confirmatory methods mentioned above, we view this parametric model as "encompassing" many submodels, i.e. different putative full rankings like $w_1 < w_2 < w_3$ or $w_2 < w_1 < w_3$ competing to explain the data. However, in analogy to "highest β% posterior density credible sets", we then seek to aggregate together the smallest number of these submodels required to attain a coherent *partial* ranking like $1, 2|3 \equiv 2, 1|3$, for a fixed posterior probability β. This set can also be aptly described as the *finest* partial ranking (FCR) with probability β, which we denote $\mathcal{C}_\beta$. Figure 2 schematically illustrates $\mathcal{C}_\beta$, and our proposed approach for identifying it, see the supplemental material for the mathematical formalization.

Our choice to limit the cardinality or complexity of our posterior inference for fixed β, contrasts with the confirmatory approach which seeks to maximize posterior model probability: in that confirmatory setting, assuming suitable complexity penalties, higher probability models are superior by definition (Gu et al., 2014; Klugkist et al., 2005; Mulder & Olsson-Collentine, 2019; Mulder et al., 2010). Also note that while Bayes factors and therefore informative priors make sense in that confirmatory work (Jeffreys, 1961; Mulder, 2010), our procedure requires neither informative priors nor informative hypotheses (because we can sidestep computation of marginal likelihood). This is fitting in exploratory settings with scientific uncertainty.

Throughout this work $P(\cdot)$ to denote a probability, $p(\cdot)$ for a probability density and $\mathbb{W}$ for a generic parameter space, e.g. $\mathbb{R}^d$, $(0, \infty)^d$ or the regular simplex Δ^d. We will assume that the underlying estimand $\mathbf{w}^* \in \mathbb{W}$ is a single point whose components are on the same scale of measurement, and that we have collected some data to derive posterior density $p(\mathbf{w}|y) = p(y|\mathbf{w})p(\mathbf{w})/p(y)$. Our work is partly inspired by the work of Lebanon & Mao (2008) on discrete preferences, and we have borrowed some other notation and definitions

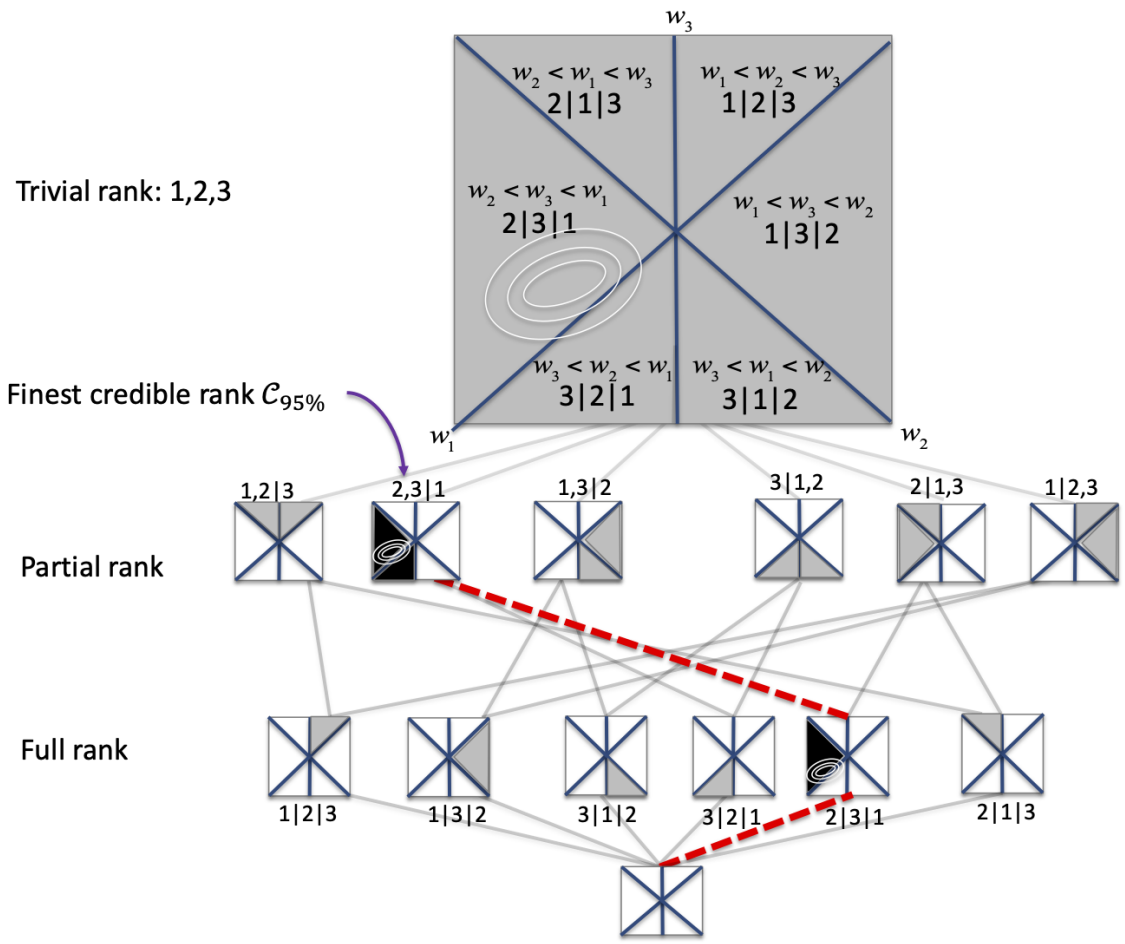

Figure 2. This graphic depicts our proposed method by considering the posterior distribution of a schematic model with three parameters. Each square depicts the same representative plane as Figure 1b, still sitting orthogonally to $w_1 = w_2 = w_3$. The schematic also includes (a slice of elliptical) posterior probability iso-contours. Our algorithm proceeds bottom to top. Starting at the base, we advance up the diagram (choosing the most probable event at the next level up). Note that upward paths correspond to the subset inclusion relation $\subseteq$ so posterior probability monotonically increases accordingly: see Figure 3 for a numerical example. We stop at the first region with probability greater than 95% and call this region the 95% finest credible region. Our procedure is inconclusive if (and only if) the first such region is the vacuous ranking at the top of the diagram. This latter result should not clearly be confused with the assertion that all parameters are equal, i.e. the classical omnibus hypothesis $w_1 = w_2 = w_3$!

from Critchlow (2012), Lebanon & Mao (2008) and Stanley (2011). In what follows we first introduce some simulations which statistically evaluate our approach before applying it a real dataset. A detailed mathematical justification of our approach can be found in the supplementary material.

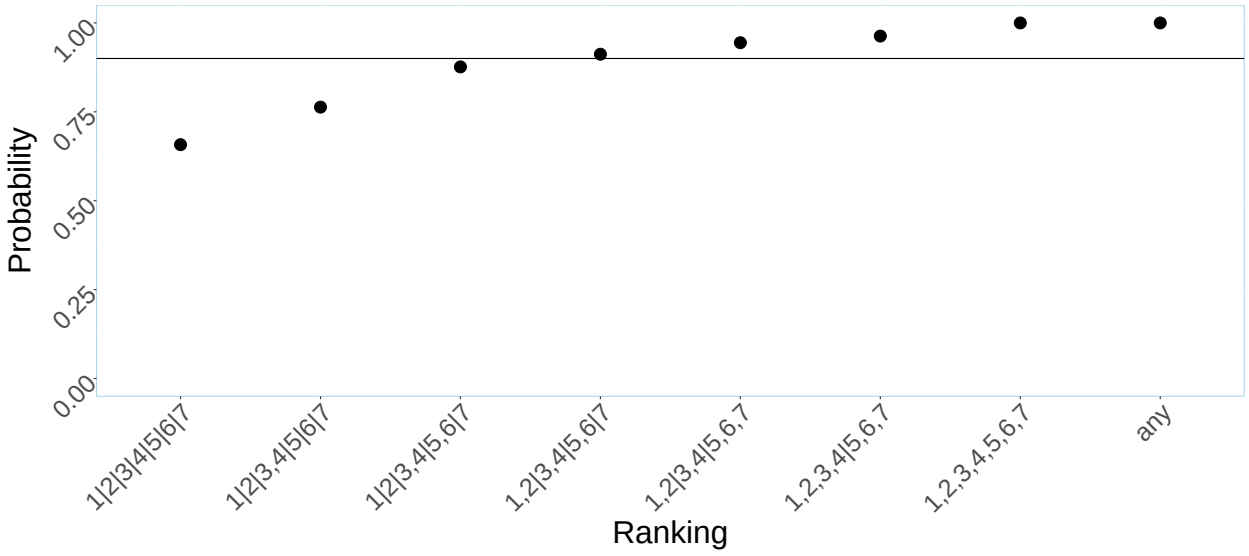

Figure 3. Posterior cumulative density over increasingly coarse partial rankings. The ground truth in this example was $1|2|3|4|5|6|7$. Progressing from left to right across the x-axis, rankings become coarser by the loss of one distinction ("|"). All points above the horizontal black line have at least 90% posterior credibility.

Rationale

Let the function f label each point $\mathbf{w} \in \mathbb{W}$ with its full ranking. For example, f gives the point $\mathbf{w} = (0.2, 0.7, 0.1)$ the label $3|1|2$ because the third component is least important, followed by the first, with component two being the most important. The true ranking of the underlying estimand is $f(\mathbf{w}^*)$, which we expect to be a full ranking with probability 1 under any continuous prior. This expectation is fitting in situations of high scientific uncertainty and reflects that parameters equalities are actually very special. Importantly, the notation extends to coarser, partial rankings, such as $3, 1|2$ which are simply coherent collections of

full rankings. Being only partial, these carry less information about the true full ranking $f(\mathbf{w}^*)$. Because full rankings are disjoint, the probability of a partial rankings is simply the sum of its constituent full ranks. A fruitful strategy is therefore to first calculate the probability of each full ranking before aggregating them as follows.

In perfect analogy to the familiar construction of "highest posterior density credible sets", a simple algorithm can then start at the maximum full ranking *a posteriori*, before recursively seeking the maximum *a posteriori* ranking at the next level of resolution (among all partial rankings with one additional bar "|"" exchanged for a ","). This is depicted in Figure 2 and gives the optimal nested sequence of partial rankings. We define the $\beta\%$ "finest credible rank" (FCR), denoted as $\mathcal{C}_\beta$, as the first (finest, or most informative) such partial ranking with β posterior probability. This is the most informative statement that can confidently be made about the ranking of parameters.

By ranking parameters as above we consider how they relate to one another. In some settings it becomes important to also know their relation to a null value $\star$. For example, in settings when we can not safely assume in advance that the underlying parameters have the same sign, the null value $\star$ is often 0, e.g. Chumbley 2021. One might answer this question by resorting to conventional sign tests or by complicating the preceding FCR problem specification and algorithm. Another pragmatic approach is as follows. First, identify the FCR as above, which will have say n distinct ranks. There are then $n+1$ ways relating it to null value $\star$: just compare the posterior probability of these. For example, if the FCR were found to be $4|1,2,3|5$ in some 5 parameter problem, then $n=3$, and the following $3+1$ situations are all compatible with this FCR: $\star|4|1,2,3|5$, $4|\star|1,2,3|5$, $4|1,2,3|\star|5$ and $4|1,2,3|5|\star$. Mulder provides tools for this.

Simulations

Simulated data

To specify the ground truth in our simulations we first sampled $d = 3, 5, 7$ parameter vectors $\mathbf{v}^*$ uniformly on the d part unit simplex, i.e. $(w_1, .., w_d) \in \mathbb{R}^d$ satisfying $w_t > 0, \sum_t w_t = 1$. We then independently scaled these samples to give $\mathbf{w}^* = \delta \mathbf{v}^*$ for δ between $0 - 5$. Here smaller δ makes parameter ranking harder because any two components of $\mathbf{w}^*$ are closer on average. These simulations include the degenerate case $\delta = 0$ where $\mathbf{w}^* = (0, ..., 0)$. For all other cases $\delta > 0$, $\mathbf{w}^*$ could be fully ordered, i.e. $\mathbf{w}^*$ contained $d - 1$ true distinctions (inequalities or bars). Given $\mathbf{w}^*$, we then generated $y_i = \sum_{j=1}^{d} x_{ij} w_j^* + \epsilon_i$ with independent $\epsilon_i \sim N(0, 1)$, for $i = 1, ..., n$, for four sample sizes $(n = 70, 140, 700, 1400)$.

Inference

Our construction for $\mathcal{C}_\beta$ requires calculating posterior probabilities for a general partial ranking. This may be achieved via closed-form integrals over the relevant subsets of posterior $p(\mathbf{w}|y)$, or by using a sample from that posterior. In the latter case, the posterior probability is (up to Markov error) just the fraction of samples satisfying the ranking constraint. Sampling schemes such as MCMC have the tremendous advantage of generality, but arguably the disadvantage of computational burden and tedius Markov error. See Chumbley et al 2021 for an application of MCMC to a ranking problem. For variety, in this work below we evaluate closed form integrals, borrowing the deterministic scheme due to Mulder (2014) to calculate the probability of each partial ranking in the most credible chain. we do not exploit the full grandeur of his method, but simply evaluate all probability of all full rankings once. These are the atoms which our algorithm aggregates into the final partial rank, our β FCR. Crucially, the probability axioms directly permit us to simply add the probability of these parts together to attain the probability of the whole FCR.

In agreement with the data generating process above, the likelihood $p(y|\mathbf{w})$ was a standard linear regression model.

Goals

Knowing the simulated ground truth $\mathbf{w}^*$ and its corresponding *true* full ranking $f(\mathbf{w}^*)$, we can examine its relation to our inferred partial ranking, the finest $\beta\%$ credible ranking denoted $\mathcal{C}_\beta$. First, is the FCR $\mathcal{C}_\beta$ consistent with the true full ranking? We say it is *in*consistent if there is any $i \neq j$ for which the FCR asserts $w_i < w_j$ while in fact $w_j < w_i$ in the simulated ground truth. Otherwise, we say it is consistent. Second, how much "information" does $\mathcal{C}_\beta$ retain? Here,we $q = r/r^*$ to measure the quality of information in $\mathcal{C}_\beta$, where r is the number of distinctions (inequalities or bars "|") in $\mathcal{C}_\beta$ and r^* is the true number in $f(\mathbf{w}^*)$. Therefore, a larger q means a more informative inference. The first question expresses the minimal requirement that $\mathcal{C}_\beta$ does not contradict the truth. The second question is motivated by the desire for $\mathcal{C}_\beta$ to be as informative as possible. Ideally, it should faithfully retain all distinctions made in the true ranking $f(\mathbf{w}^*)$. If $f(\mathbf{w}^*) = \mathcal{C}_\beta$ then $q = 1$ reflecting that our credible set contains just one full ranking. Smaller q implies higher uncertainty about the full rank.

Our simulations included the degenerate case of $\delta = 0$ where $w_1^* = w_2^* = ... = w_d^*$ is akin to the conventional omnibus null hypothesis. This is depicted as the central intersection in Figure 1b. We can identify "false positives" in this setting with $\mathcal{C}_\beta \neq \mathbb{W}$. Conversely, the remaining simulations assumed that $\{w_1^*, w_2^*, ..., w_d^*\}$ can be totally ordered. In this setting we can identify "false negatives" with $\mathcal{C}_\beta = \mathbb{W}$ and can therefore ask a third question: What is the proportion of false negatives and false positives so defined?

Results

We can answer the three questions posed above as follows. First, we found that our inferred partial rank $\mathcal{C}_\beta$ violated the ground truth in only 0.02 percent of the simulations. Second, on average over all simulations 0.50 % of the distinctions were preserved. Table 1 shows that q, the proportion of distinctions preserved in $\mathcal{C}_\beta$, increased with the simulated sample size. We naturally expected q to improve with δ, because the components $\mathbf{w}^*$ will vary more relative to one another and to the observation noise. Figure 4 captures this dependence by plotting jittered q against the per-simulation standard deviation over components of $\mathbf{w}^*$ (which reflects different settings of δ). In particular the x-axis of Figure 4 is $\frac{1}{d}\sum_{i=1}^{d}(w_d^* - \bar{\mathbf{w}}^*)^2$, with $\bar{\mathbf{w}}^*$ denoting the mean over components of the ground truth $\mathbf{w}^*$.

Table 1

The proportion q of distinctions preserved by the posterior credible ranking increased with the simulated sample size.

n	q
70	0.35
140	0.42
700	0.59
1400	0.64

Our third and final question is answered in Table 2, which summarizes the false positive and negative rates in our simulations. Column TRUE indicates that the underlying parameter $\mathbf{w}^*$ can be fully ordered. Column FALSE indicates that there $\mathbf{w}^*$ cannot be ordered at all, because $\mathbf{w}^* = (0, ..., 0)$. Row TRUE indicates that $\mathcal{C}_\beta$ detected at least one distinction or inequality of this order. The diagonals therefore offer one measure of success. Bottom left indicates "false positives" where a false distinction was inferred, so $\mathcal{C}_\beta \neq \mathbb{W}$ despite none actually existing because $\mathbf{w}^* = (0, ...0)$. Top right indicates "false negatives",

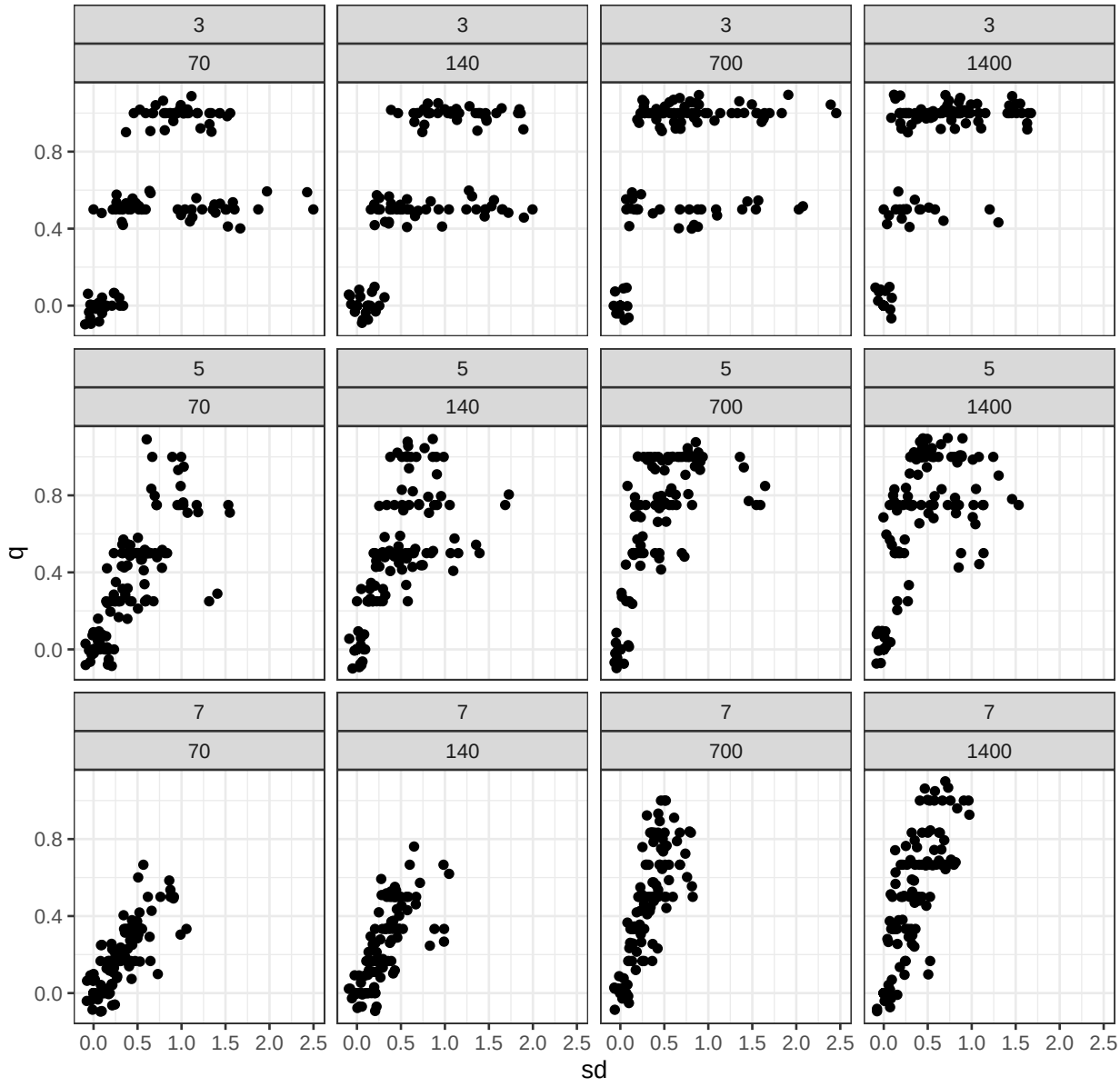

Figure 4. Quality q (jittered) against the per-simulation standard deviation over components. This illustrates the intuitive fact that, in practice, the quality of the relative measure $\mathcal{C}_\beta$ does depend on the scale of underlying differences between component parameters relative to the scale of the noise. Rows vary over the dimension d of simulated $\mathbf{w}^*$ and columns vary over sample size n.

where no distinction was inferred so $\mathcal{C}_\beta = \mathbb{W}$, while in fact $\mathbf{w}^* \neq (0, ...0)$ was fully ordered. Thus false positives were very rare over all conditions. While false negatives were more prevalent overall, their rate vanishes to practically zero for stronger effects, indexed for example by δ below (or equally the per-simulation standard deviation over components discussed above).

Table 2

False positives and negatives of the posterior credible ranking, over all simulations. The column variable indicates the ground truth: whether the underlying parameter can be ordered (TRUE) or not. The row variable indicates whether any order was detected by our method.

	FALSE	TRUE
FALSE	117	41
TRUE	3	559

Application of our method to real behavioral data

To better understand percieved discrimination against female applicants in the hiring process, Carlsson & Sinclair (2018) regressed perceptions of discrimination towards female victims on "belief in discrimination against women", "stigma consciousness", and "feminist identification", while controlling for "gender" and "belief in discrimination against men". As a regression equation, this can be expressed as

$$\begin{aligned} y_{discriminationW,i} &= \beta_0 + \beta_{beliefW} X_{beliefW,i} \\ &\quad + \beta_{stigma} X_{stigma,i} \\ &\quad + \beta_{feminist} X_{feminist} \\ &\quad + \beta_{gender} X_{gender,i} \\ &\quad + \beta_{beliefM} X_{beliefM,i} + \text{error}_i. \end{aligned}$$

where the β's are standardized regression effects of the variables on perceived

Table 3

False positives and negatives of the posterior credible ranking, delta = 0.

	FALSE	TRUE
FALSE	117	0
TRUE	3	0

discrimination, see Mulder & Olsson-Collentine (2019) for more details. In this setting, the finest credible ranking above 90% was "stigma , feminist | beliefW", with posterior probability of 94%. This means that belief in discrimination about women is more important than either stigma consciousness or feminist identification, the latter two being indiscriminable. This result suggests that, in attempting to explain the percieved discrimination against female applicants in the hiring process, the celebrated "prototype" explanation of social psychology trumps the "same-gender bias" explanation, see Carlsson & Sinclair (2018). Importantly, this conclusion was reached automatically, without the requirement to explicitly specify any confirmatory hypotheses in advance. This insures the method against local minima, i.e. selecting a bad ranking that is nonetheless relatively plausible amoung the overly-restricted set of hypotheses considered. Note that while a typical coefficient-based analyses might only consider the marginal distribution on each parameter $w_i|y$ at a time, this analysis summarizes the full posterior distribution.

Conclusions

Classical hypothesis tests are increasingly discouraged and demand careful attention to the multiple comparisons problem. Confidence or credible intervals are a valuable alternative but become prohibitively complicated in higher dimensions. Bayesian model comparison is ideally matched to confirmatory research, where prior data and theory clearly outline the competing explanations. We consider another way to summarize multi-parameter hypotheses

Table 4

False positives and negatives of the posterior credible ranking, delta = 1.

	FALSE	TRUE
FALSE	0	34
TRUE	0	86

in situations of high scientific uncertainty, when the relative magnitude of parameters is of interest. In analogy to the familiar highest posterior density credible sets, this entails evaluating increasingly coarse tilings of the parameter space for the smallest with partial ranking with β probability. We call this the finest β% credible ranking (FCR) supported by the data. Just as a highest density credible set contains the posterior maximum, our finest credible set of full rankings $\mathcal{C}_\beta$ contains the posterior maximum full rank. It can be shown that the FCR is generally a connected set of regions, meaning that any full order in the FCR is at most Kendall distance 1 from some other full order in the FCR, see pg 89-91 of Stanley & others (2004). In other words, the simple topology of FCR (Euler characteristic = 1) mirrors the unimodal prior and likelihood assumptions common in many applications. We have shown that FCR has low error and that the information it preserves about true ranking unsurprisingly depends on the signal to noise in the data: it increases with the sample size and the average distance between components of the underlying estimand $\mathbf{w}^*$.

A couple of points on interpretation. First, the trivial ranking $\mathcal{C}_\beta = \mathbb{W}$ is inconclusive about the partial ranking of parameters: it simply reflects high posterior uncertainty is not to be taken as evidence for the popular omnibus equality hypothesis $w_i = w_j$ for all i, j. (This can also be pictured by contracting or expanding the isocontours in our schematic posterior density in Figure 2.) It does not imply $p(w_i = w_j, \forall i, j|y) \geq \beta$. In fact we have not studied equality hypotheses at all in this work, in part avoid complicated distractions such as the Lindley paradox, see Chumbley et al 2021, Kruskal & Majors (1989), Johnson & LeBreton (2004), Mulder & Olsson-Collentine (2019). Note however that any non-trivial conclusion

Table 5

False positives and negatives of the posterior credible ranking, delta = 2.

	FALSE	TRUE
FALSE	0	5
TRUE	0	115

$\mathcal{C}_\beta \neq \mathbb{W}$ does indeed contradict some equality hypothesis: if $w_i < w_j$ in the $\mathcal{C}_\beta$, then $w_i \neq w_j$. For example, any $\mathcal{C}_\beta$ with 2 or more ranks like $w_1 < \{w_2, w_3\}$ contradicts the omnibus hypothesis above. Similarly, if the $\mathcal{C}_\beta$ is a composition of 3 or more distinct ranks, we can reject any set of equality hypothesis constraining parameters to be one of only two values. This may be useful, for example, in developmental epidemiology these two cases amount to rejecting the so-called accumulation hypothesis and the critical period hypothesis in favor of any so-called sensitive period hypothesis consistent with the $\mathcal{C}_\beta$ Chumbley et al 2021.

The ideas presented here are agnostic to the particular method used for posterior inference: they apply to either a posterior sample or to some deterministic or exact scheme. In this work we have chosen to adapt to our purposes a deterministic scheme originally proposed for confirmatory analyses and available in the R package "lmhyp" (Mulder, 2014). This scheme is fast and has no Markov error, albeit currently having limited generality. We have also validated our proposed approach for use with MCMC samples from the popular "brms" package, see Chumbley et al 2021.

While we have focused here on the mean components $(w_1, ..., w_d)$ of a regression model, our approach readily applies to rank all or some components of a covariance or precision matrix. It is therefore useful when studying say the variance components $(\sigma_1, ..., \sigma_d)$ in a multi-level model. In that setting, the finest credible ranking of variance parameters in a simple multilevel dataset with 4 random factors, each with say 20 levels, might transpire to be $\mathcal{C}_\beta = \sigma_3, \sigma_1 | \sigma_4 | \sigma_2$. This indicates that random factor 2 explains most variation in the

Table 6

False positives and negatives of the posterior credible ranking, delta = 3.

	FALSE	TRUE
FALSE	0	1
TRUE	0	119

outcome, followed by factor 4, but the other factors are indiscriminable. Standard existing methods obviously cannot directly support such relativistic conclusions, as they simply isolate the absolute contributions of each factor.

Acknowledgements

We would like to thank Margaret Bellamy, Helene Schernberg, Edward Moyles, Wenjia Xu, Cecilia Potente.

Software

An R package is available on request to the corresponding author.

Table 7

False positives and negatives of the posterior credible ranking, delta = 4.

	FALSE	TRUE
FALSE	0	0
TRUE	0	120

Supplementary material

Preliminaries

Our procedure should find the maximum posterior *full ranking*, which may be say $w_1 < w_2 < w_3$, and tell us whether it is $\beta\%$ plausible. If not, we naturally want to know whether some coarser, *partial ranking*, say $w_1 < w_2, w_3$, is $\beta\%$ plausible. Note that this partial ranking is less informative or "coarser" because it subsumes the first. It simply asserts that w_1 is the smallest component *irrespective of the relative ranking of* w_2, w_3. A nice notation for these rankings, which we define more formally in section 3 below, is

$$h_1 = 1|2|3 = \{\mathbf{w} : w_1 < w_2 < w_3\}$$
$$h_1' = 1|2,3 = \{\mathbf{w} : w_1 < w_2 < w_3 \; or \; w_1 < w_3 < w_2\}.$$

In this notation a recursively constructed coarsening of any partial rank can be represented simply by replacing some of the "|" with ",".

Note that h_1 treats an entire class of points $\mathbf{w} \in \mathbb{W}$ as equivalent, because individual points are not of practical interest. It is therefore akin to the popular "regions of practical equivalence" (Kruschke, 2011) but with two points deemed equivalent $\mathbf{w} \equiv \mathbf{w}'$ not because they are close to being equal but because they have the same full ranking. More abstractly, any *partial ranking* reflects that some set of full rankings are in turn deemed equivalent, as

Table 8

False positives and negatives of the posterior credible ranking, delta = 5.

	FALSE	TRUE
FALSE	0	1
TRUE	0	119

with h_1' above which equates $h_1 \equiv h_2$ where $h_2 = \{\mathbf{w} : w_1 < w_3 < w_2\}$. Our posterior credible set $\mathcal{C}_\beta$ will be constructed in this way, using equivalences to glue together just as many full rankings as is necessary to attain the desired $\beta\%$ credibility, see Figures 1,2 and section 3. Briefly, we identify this smallest credible set $\mathcal{C}_\beta$ such that $P(\mathcal{C}_\beta|y) \geq \beta$ among a well-defined space of partial rankings by simple application of the additivity and monotonicity of probability (calculating the posterior probability of each generic ranking h as $P(h|y) = \int_h p(\mathbf{w}|y)d\mathbf{w}$).

Definition: An equivalence relation on a set $(Q, \equiv)$ is a set Q with a relation $\equiv \subseteq Q \times Q$ such that for all $x, y, z \in Q$ the relation $\equiv$ is 1) reflexive $x \equiv x$ 2) symmetric $x \equiv y \implies y \equiv x$ and 3) transitive $x \equiv y$ *and* $y \equiv z \implies x \equiv z$.

Given a surjective function $f : A \to B$ onto a finite set B, any partition of $B = B_1 \uplus ... \uplus B_n$ into n blocks correspondingly partitions $A = A_1 \uplus ... \uplus A_n$ into n blocks as follows. Simply let each block of A be the set of all points which map to a given block of B, meaning $f(x_1), f(x_2) \in B_i$ for some i. In other words, $f : A \to B$ together with a partition of B, imply an equivalence relation on A where two points $x_1 \equiv x_2$ are equivalent if they map to the same block of B. We will use equivalence relations constructed in this way to abstract over all real points in $\mathbb{W}$ that share some partial ranking, and use subscripts such as s in $\equiv_s$ to distinguish equivalence relations - or partitions - of different coarseness.

Definition: A binary operation on a set $(Q, \circ)$ is a function $\circ : Q \times Q \to Q$. We will use the special binary operation $\circ$ of an algebraic group - namely a closed, associative binary

operation with inverses and an identity element - to construct the equivalence relations introduced above. This group structure also provides a natural way for us to calculate the size or coarseness of each class.

Definition: A partially ordered set or *poset* $(Q, \preceq)$ is a set Q such that for all $x, y, z \in Q$ the relation $\preceq$ is 1) reflexive $x \preceq x$ 2) anti-symmetric $x \preceq y \text{ and } y \preceq x \implies x = y$ and 3) transitive $x \preceq y \text{ and } y \preceq z \implies x \preceq z$. We will use this to compare the coarseness of the equivalence relations introduced above, which in turn guides our search for $\mathcal{C}_\beta$.

Definition: Any partially ordered set has a corresponding *covering relation* $\vdash$. Given the poset $(Q, \preceq)$, we write $x \prec y$ if $x \preceq y$ and $x \neq y$. We say that y *covers* x and write $x \vdash y$ when $x \prec y$ and there is no $z \in Q$ such that $x \prec z \prec y$. In other words y is as close as possible to x. This covering relation $\vdash$ on the poset Q can be conveniently represented by a Hasse diagram, as in Figure 2.

Full rankings

We construct our space of partial rankings via the rank transformation $f : \mathbb{W} \to \mathcal{S}_d$ which maps each point $\mathbf{w}$ to its full ranking π where here $\mathcal{S}_d$ is the set of all full rankings, that is all permutations of parameter indices $[d] := \{1, ..., d\}$. More precisely, $f : \mathbb{W} \setminus \mathcal{B}_d \to \mathcal{S}_d$ where the so-called braid arrangement $\mathcal{B}_d = \{\mathbf{w} \in \mathbb{W} : w_i = w_j, \textit{some } i \neq j\}$ contains points which cannot be fully ranked because 2 or more components are exactly equal. We can suppress this technical detail because $\mathcal{B}_d$ is a "tiny" set with probability zero under any jointly continuous prior. In other words, any $\mathbf{w}$ can be fully ranked with probability 1 in the current setting.

For example, f maps $\mathbf{w} = (4.6, 8.9, -11.4)$ to $\pi = (2, 3, 1)$, where $\pi(l)$ gives the rank of component w_l. This value π of f can also be represented verbosely as a matrix whose first

row records the component indices $1, 2, 3$ of $\mathbf{w} = (w_1, w_2, w_3)$ and whose second row records the rank of each component parameter, here yielding $\pi = \left(\begin{smallmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{smallmatrix}\right)$. For clarity in our setting we will usually follow a different and more intuitive convention, instead writing this same full ranking or permutation as $3|1|2$, with a "|" notation meaning $\pi^{-1}(1)|\pi^{-1}(2)|...|\pi^{-1}(d)$. While this bar notation is commonly used in this way to express permutation π on d discrete points (say integers or objects), we extend its interpretation to the continuous domain $\mathbb{W}$ as follows. First note that the preimages of f, denoted $\mathcal{R} := \{f^{-1}(\pi) \in \mathbb{W} : \pi \in \mathcal{S}_d\}$, implicitly define an equivalence relation $\equiv_d$, the so-called equivalence kernel of f. This is because f partitions the domain $\mathbb{W}$ of f into equivalence classes, one per full ranking. Thus we say $\mathbf{w} \equiv_d \mathbf{w}'$ if and only if both $\mathbf{w} \in f^{-1}(\pi)$ and $\mathbf{w}' \in f^{-1}(\pi)$. The subscript d here indicates that each class corresponds to a total ordering of all d component parameters of $\mathbf{w}$. Geometrically, elements of $\mathcal{R}$ correspond to regions of a tesselated $\mathbb{W}$, dissected by a "braid arrangement" of hyperplanes, denoted $\mathcal{R} = \mathcal{R}(\mathcal{B}_d)$ by (Stanley & others, 2004). Figure 1b attempts to depict $\mathcal{R}$ in the case of our running example, where h_1, h_2 but not $h_1' := h_1 \cup h_2$ are members of $\mathcal{R}$. The collection of $d!$ full rankings (regions) in $\mathcal{R}$ forms our finest discretization of $\mathbb{W}$, whose elements we emphasize are defined by

$$\begin{aligned}\mathcal{R} &:= \{f^{-1}(\pi) \in \mathbb{W} : \pi \in \mathcal{S}_d\} \\ &= \{\{\mathbf{w} \in \mathbb{W} : w_{\pi^{-1}(1)} < w_{\pi^{-1}(2)} < ... < w_{\pi^{-1}(d)}\} : \pi \in \mathcal{S}_d\}.\end{aligned}$$

Note the one-to-one correspondence between the regions of $\mathcal{R}$ and $\mathcal{S}_d$

$$\begin{aligned}\mathcal{S}_d &\leftrightarrow \mathcal{R} \\ \pi &\leftrightarrow \mathcal{R}_\pi := \{\mathbf{w} \in \mathbb{W} : w_{\pi^{-1}(1)} < w_{\pi^{-1}(2)} < ... < w_{\pi^{-1}(d)}\}\end{aligned}$$

which justifies our reinterpretation below of the symbol π or equivalently $\pi^{-1}(1)|\pi^{-1}(2)|, ..., |\pi^{-1}(d)$ as $\mathcal{R}_\pi \subseteq \mathbb{W}$ as above[1], e.g. $h_1 = 1|2|3, h_2 = 1|3|2$.

[1]Similarly, notation such as $P(\pi) = P(\pi^{-1}(1)|...|\pi^{-1}(d)|)$ below unambiguously refers to $P(\mathcal{R}_\pi)$, the

We have not yet used any special algebraic structure to construct the equivalence relation above. Recall however that the set of $[d]$ integer permutations $\mathcal{S}_d$ is an algebraic group when viewed as a set of bijective functions $[d] \to [d]$, with the group operation $\circ$ being function composition. This group structure applies equally to $\mathcal{R}$: any region (full ranking) $\mathcal{R}_\pi \in \mathcal{R}$ can be bijectively mapped onto another by a series of reflections, the rigid body isometries encoded formally by the Coxeter reflection group of $\mathcal{R}$, see Figure 1b and Stanley & others (2004).

Because $\mathcal{R}$ partitions or quotients $\mathbb{W}$ up to a set with measure 0, it discretizes $P(\mathbf{w}|y)$ satisfying $1 = \int_{\mathbb{W}} p(\mathbf{w}|y)d\mathbf{w}$ into $P(\mathcal{R}_\pi|y)$ satisfying
$1 = \sum_{\pi \in \mathcal{S}_d} \int_{\mathcal{R}_\pi} p(\mathbf{w}|y)d\mathbf{w} = \sum_{\pi \in \mathcal{S}_d} P(\pi|y)$ where the final step is just for notational simplicity, equating $\pi \in \mathcal{S}_d$ with its corresponding $\mathcal{R}_\pi$. If one $\mathcal{R}_\pi$ happens to satisfy our criterion $P(\pi|y) > \beta\%$, then $\pi = \mathcal{C}_\beta$ and no smaller $\mathcal{C}_\beta$ is possible: our search for the finest credible rank $\mathcal{C}_\beta$ is complete. Otherwise, the fact that $\mathcal{R}$ has a group structure will make it easy for us to define the coarser, partial rankings which will necessarily contain $\mathcal{C}_\beta$.

To summarize, any element of $\mathcal{R}$ corresponds to a full ranking or total ordering of component parameters $1, ..., d$ (each component of $\mathbf{w}$ has a single, unambiguous rank and there are no ties). We have noted that $\mathcal{R}$ shares structure with $\mathcal{S}_d$. Viewed discretely, $\mathcal{R}$ exchanges the familiar continuous linear algebra and metrics of $\mathbb{W}$, on which standard confidence intervals are based, for the simpler group algebra and metrics of discrete *regions* within that vector space. We next define partial ranks more generally because these constitute our finite set of candidates for $\mathcal{C}_\beta$.

---

probability of the corresponding subset of $\mathbb{W}$, see Figure 1. The context should distinguish whether π or $\mathcal{S}_d$ refer to permutations of integers $[d]$, or their corresponding $\mathcal{R}_\pi \subset \mathbb{W}$.

Partial rankings

Figure 2 attempts to depict our space of partial rankings for $d = 3$. As discussed in the previous section, for brevity we continue to casually use the notation conventionally used for subsets $\mathcal{S}_d$, such as $3|1|2$ and $1|2,3$, but to refer to the corresponding subsets of continuous $\mathbb{W}$. Thus row 3 of Figure 2 shows all elements of $\mathcal{S}_3$, the complete set of full ranking equivalence classes of $\mathbb{W}$. Row 2 shows two coarser equivalence relations denoted $\mathcal{S}_3/\mathcal{S}_{1,2}$ and $\mathcal{S}_3/\mathcal{S}_{2,1}$. Here the subscripts in the denominator indicate the integer "composition" of $d = 3$ parameters over $r = 2$ ranks.

Less casually, an integer *composition* denoted $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_r)$, is defined to be a sequence of r positive integers which sum to d. Associate with each integer composition the unique partition of integers $1, ..., d$, namely
$N_1 = \{1, 2, ..., \gamma_1\}, N_2 = \{1 + \gamma_1, ..., \gamma_1 + \gamma_2\}, N_r = \{1 + \sum_{l=1}^{r-1} \gamma_l, ..., d\}$. Then the set $\mathcal{S}_{\boldsymbol{\gamma}}$ is defined to be the set of all permutations $\sigma \in \mathcal{S}_d$ for which the following set equalities hold

$$\sigma(N_i) = N_i, \qquad i = 1, ..., r.$$

Thus $\mathcal{S}_{\boldsymbol{\gamma}}$ contains permutations of integers $1, ..., d$ that permute within but not across the N_i. It can be shown that this set is a subgroup of $\mathcal{S}_d$, and can thereby induce an equivalence relation on $\mathcal{S}_d$, defined as the set of right cosets of $\mathcal{S}_{\boldsymbol{\gamma}}$. We can write this new and coarser equivalence relation on integer permutations using conventional divisor notation as in $\mathcal{S}_3/\mathcal{S}_{1,2}$. Each equivalence class in this relation is a partial ranking of integers $\mathcal{S}_{\boldsymbol{\gamma}}\pi := \{\sigma\pi : \sigma \in \mathcal{S}_{\boldsymbol{\gamma}}\}$ with composition $\boldsymbol{\gamma}$, represented here by the fixed $\pi \in \mathcal{S}_d$. For example, $\pi = \left(\begin{smallmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{smallmatrix}\right)$, otherwise known as $3|2|1$, can alternatively represent the equivalence class $2, 3|1 = \{3|2|1, 2|3|1\}$ in $\mathcal{S}_3/\mathcal{S}_{2,1}$ as

$$\mathcal{S}_{2,1} \circ \left(\begin{smallmatrix}1&2&3\\3&2&1\end{smallmatrix}\right) = \{\left(\begin{smallmatrix}1&2&3\\1&2&3\end{smallmatrix}\right), \left(\begin{smallmatrix}1&2&3\\2&1&3\end{smallmatrix}\right)\} \circ \left(\begin{smallmatrix}1&2&3\\3&2&1\end{smallmatrix}\right) := 2,3|1.$$

Here the group operation of function composition is explicitly denoted $\circ$ and the second equality just substitutes in our preferred notation. This second equality more formally defines our bar notation for any partial ranking as the set of all full orders of integers which shuffles components within but not across "|" bars[2]. The set of such partial rankings comprise an equivalence relation with the three classes $2,3|1$, $1,2|3$ and $1,3|2$. Recalling the correspondence $\mathcal{R} \leftrightarrow \mathcal{S}_d$ we have just constructed an equivalence relation $\equiv_{\boldsymbol{\gamma}}$ on our native parameter space $\mathbb{W}$. According to this relation we can now unambiguously say $\mathbf{w} \equiv_{\boldsymbol{\gamma}} \mathbf{w}'$ if and only if both $\mathbf{w} \in f^{-1}(\mathcal{S}_{\boldsymbol{\gamma}}\pi)$ and $\mathbf{w}' \in f^{-1}(\mathcal{S}_{\boldsymbol{\gamma}}\pi)$ for some $\pi \in \mathcal{S}_d$. But this notation is cumbersome so we continue to refer to these coarser equivalence classes using discrete bar notation, identifying symbols like $2,3|1$ with their corresponding subsets of $\mathbb{W}$ as throughout row 2 of Figure 2.

The lattice of partial rankings

Following Lebanon & Mao (2008) we consider the set $\mathcal{T}_d := \{\mathcal{S}_{\boldsymbol{\gamma}}\pi : \pi \in \mathcal{S}_d, \forall\gamma\} \cup \{\emptyset\}$ of all partial rankings augmented by a null element $\emptyset$. Unlike that work, we have emphasized that all these symbols are to be interpreted as subsets of $\mathbb{W}$. Crucially these subsets form a partially ordered set $(\mathcal{T}_d, \preceq)$ (see definition above and Figure 2) if we define $x \preceq y \leftrightarrow x \subseteq y$. Similarly, we will write $x \vdash y$ if there is no superset of x smaller than y. Then one rank covers another exactly when it is a single step up some chain in the Hasse diagram, see Figure 2. The *covering set* of x, denoted $C_x = \{z \in \mathcal{T}_d : x \vdash z\}$, is naturally the set of all partial ranks which cover x.

[2]As a further example, a partial ranking of composition say $\boldsymbol{\gamma} = (1,4,2)$ in $\mathcal{S}_7/\mathcal{S}_{1,4,2}$ refers to something with $r = 3$ ranks over $d = 7$ parameters like $7|1,2,3,6|4,5$.

Prior probability on partial rankings

In linear modeling scenarios, the prior density $p(\mathbf{w})$ on parameters is commonly chosen to be continuous and exchangeable with mean zero. For example, $p(\mathbf{w})$ may be an independent and identically distributed, mean zero Gaussian. Continuity means we believe the unobserved parameters are distinct and therefore fully ranked without any ties (the event that any two components are exactly equal has probability zero). Exchangeablity means we know nothing about the relative rank *a priori*: the density $p(\mathbf{w})$ has the symmetry $p(w_1, w_2..., w_d) = p(w_{\pi(1)}, w_{\pi(2)}, ..., w_{\pi(d)})$ for any permutation π of $[d]$. This implies a uniform prior over full rankings $P(\mathcal{R}_\pi) = 1/d!$. More generally, the uniformity $P(h) = P(h')$ holds for any two partial rankings $h, h' \in \mathcal{S}_d/\mathcal{S}_{\boldsymbol{\gamma}} \subseteq \mathcal{T}_d$ with the same composition $\boldsymbol{\gamma}$. This follows because Lagrange's theorem guarantees that cosets h, h' contain the same number of permutations or full rankings, and exchangeability implies that each of these constituent permutations is equiprobable. We therefore need only the size of $\mathcal{S}_d/\mathcal{S}_{\boldsymbol{\gamma}}$ to attain $P(h) = 1/\#(\mathcal{S}_d/\mathcal{S}_{\boldsymbol{\gamma}})$ for all $h \in \mathcal{S}_d/\mathcal{S}_{\boldsymbol{\gamma}}$. We know this size to be $\binom{d}{\boldsymbol{\gamma}}$, implying that exchangeability effectively discretizes $p(\mathbf{w})$ into $P(h) = 1/\binom{d}{\boldsymbol{\gamma}}$ for $h \in \mathcal{S}_d/\mathcal{S}_{\boldsymbol{\gamma}}$ for *any* composition $\boldsymbol{\gamma}$.

Posterior probability of partial rankings and $\mathcal{C}_\beta$

Definition: We define the *MAP full ranking* to be the full ranking with maximum posterior probability is $\pi_{MAP} = \underset{\pi \in \mathcal{S}_n \subseteq \mathcal{T}_d}{\operatorname{argmax}} P(\pi|y)$ with π interpreted as $\mathcal{R}_\pi \subseteq \mathbb{W}$.

Definition: A *chain* is a sequence $(R_0, R_1, ..., R_d)$ of d progressively coarser partial rankings in $\mathcal{T}_d$ so that $R_i \vdash R_{i+1}$ where $\vdash$ is the covering relation defined above.

Starting from a full ranking R_1, say $3|4|2|1$, the elements of this sequence are attained by progressively removing one bar "|" at a time. This will result in the trivial ranking "$1, 2, 3, 4$" (no distinct bars or ranks) after 3 steps. Any chain is an upward path in a Hasse

diagram, see Figure 2. Note that the first element in this chain $R_0 = \phi$, R_1 must be a single full ranking, and $R_d = \mathbb{W}$. The number of distinct bars "|" at each step R_t in the upward sequence is $r = d - t$ for $t > 0$, which is the number of ways for the chain to proceed.

Definition: The *most credible chain* is the special chain $(\bar{R}_0, \bar{R}_1, ..., \bar{R}_d)$ such that for all t, $\bar{R}_t$ has higher probability than any competing partial ranking $R'_t \in C_{R_{t-1}}$. Thus $\bar{R}_0 = \emptyset$, $\bar{R}_1 = \pi_{MAP}$ and generally $\bar{R}_t$ satisfies the recursion $\bar{R}_t = \underset{R_t \in C_{\bar{R}_{t-1}}}{\operatorname{argmax}} P(R_t|y)$.

Definition: The $\beta\%$ *credible ranking* is defined to be $\mathcal{C}_\beta = \bar{R}_t$ where t is the smallest index such that $P(\bar{R}_t|y) \geq 0.95$. This motivates our algorithm below.

By definition therefore $P(\mathcal{C}_\beta|y) \geq \beta$. In practice, equality is very unusual so this inequality is strict. This is because the posterior probability increases discontinuously with each progressive coarsening in our construction of $\mathcal{C}_\beta$. This makes overshoot likely, in that the first partial ranking with the desired credibility of β may actually be much more credible than β. To avoid understating the credibility of a partial ranking in this situation, it therefore makes sense to additionally report $P(\mathcal{C}_\beta|y)$, for example $\mathcal{C}_{90\%} = 3|2|4|1|5$ and $P(\mathcal{C}_{90\%}) = 0.99$.

Algorithm

Our initial condition for $\mathcal{C}_\beta$ is the π_{MAP}, the maximum posterior full ranking. Finding this initialization is the most intensive step of our procedure, involving a discrete maximization over $d!$ possibilities. In higher dimensional problems, a fast approximation is $\hat{\pi}_{MAP} = f(\mathbf{w}_{MAP})$, where $\mathbf{w}_{MAP} := \underset{\mathbf{w}}{\operatorname{argmax}} P(\mathbf{w}|y)$ and f is still the rank transformation introduced above. In other words, start at the unique full rank on all d parameters implied by $\mathbf{w}_{MAP}$. If, for example $\mathbf{w}_{MAP} = (3.2,\ 5.9,\ 4.0,\ 8.7)$, then at iteration 1 this corresponds to setting $R_1 = 1|3|2|4$. Our simulation study below shows that such a heuristic can yield surprisingly good results. Note that this approximation is always well-defined: continuity of

our prior/posterior implies that the components of $\mathbf{w}_{MAP}$ can indeed be fully ranked with probability one.

Starting from $R_1 := \pi_{MAP}$ (or $\hat{\pi}_{MAP}$), we immediately stop if $p(R_1) \geq \beta$ and declare $\mathcal{C}_\beta = R_1$. Otherwise, enumerate all hypotheses *covered by* R_1 in the ranking poset - the smallest ranks which strictly include R_1. Each of these candidates has an associated probability. Choose *the most credible* of these candidates. If two or more rankings have equal (and maximal) credibility, then pick one arbitrarily. Repeat until there is a ranking with credibility ≥ 0.95. See Figure 3, which depicts the trace of this algorithm for one simulated dataset.

Proof of convergence: The initial full rank will have probability ≤ 1. Probability must increase at each step progressing up any chain in the poset, by the monotonicity of probability. We can therefore see that the algorithm will trace out a "cumulative density function" as it progresses along the *most credible chain* in the poset. If the algorithm proceeds uninterrupted to the end, it will remove all $d - 1$ inequalities from the initial full rank. This yields the trivial ranking $\mathcal{C}_\beta = \mathbb{W}$ with $P(\mathbb{W}|y) = 1$. Because probability increases *monotonically* up to this point, it must at some point pass the desired $\beta\%$ of credibility. If the algorithm terminates before reaching $\mathbb{W}$, it will have identified the non-trivial $\mathcal{C}_\beta$ which has credibility $\geq \beta\%$.

REFERENCES

Carlsson, R., & Sinclair, S. (2018). Prototypes and same-gender bias in perceptions of hiring discrimination. *The Journal of Social Psychology*, *158*(3), 285–297.

Critchlow, D. E. (2012). *Metric methods for analyzing partially ranked data* (Vol. 34). Springer Science & Business Media.

Gu, X., Mulder, J., Deković, M., & Hoijtink, H. (2014). Bayesian evaluation of inequality constrained hypotheses. *Psychological Methods*, *19*(4), 511.

Jeffreys, H. (1961). Theory of probability, clarendon. Oxford.

Johnson, J. W., & LeBreton, J. M. (2004). History and use of relative importance indices in organizational research. *Organizational Research Methods*, *7*(3), 238–257.

Klugkist, I., Kato, B., & Hoijtink, H. (2005). Bayesian model selection using encompassing priors. *Statistica Neerlandica*, *59*(1), 57–69.

Klugkist, I., Laudy, O., & Hoijtink, H. (2005). Bayesian eggs and bayesian omelettes: Reply to stern (2005).

Kruschke, J. K. (2011). Bayesian assessment of null values via parameter estimation and model comparison. *Perspectives on Psychological Science*, *6*(3), 299–312.

Kruskal, W., & Majors, R. (1989). Concepts of relative importance in recent scientific literature. *The American Statistician*, *43*(1), 2–6.

Lebanon, G., & Mao, Y. (2008). Non-parametric modeling of partially ranked data. *Journal of Machine Learning Research*, *9*(Oct), 2401–2429.

Mulder, J. (2010). *Bayesian model selection for constrained multivariate normal linear*

models (PhD thesis). Utrecht University.

Mulder, J. (2014). Prior adjusted default bayes factors for testing (in) equality constrained hypotheses. *Computational Statistics & Data Analysis*, *71*, 448–463.

Mulder, J., & Olsson-Collentine, A. (2019). Simple bayesian testing of scientific expectations in linear regression models. *Behavior Research Methods*, 1–14.

Mulder, J., Hoijtink, H., & Klugkist, I. (2010). Equality and inequality constrained multivariate linear models: Objective model selection using constrained posterior priors. *Journal of Statistical Planning and Inference*, *140*(4), 887–906.

Stanley, R. P. (2011). Enumerative combinatorics, vol. 49. Cambridge university press.

Stanley, R. P., & others. (2004). An introduction to hyperplane arrangements. *Geometric Combinatorics*, *13*, 389–496.

Stern, H. S. (2005). Model inference or model selection: Discussion of klugkist, laudy, and hoijtink (2005).

Carlsson, R., & Sinclair, S. (2018). Prototypes and same-gender bias in perceptions of hiring discrimination. *The Journal of Social Psychology*, *158*(3), 285–297.

Critchlow, D. E. (2012). *Metric methods for analyzing partially ranked data* (Vol. 34). Springer Science & Business Media.

Gu, X., Mulder, J., Deković, M., & Hoijtink, H. (2014). Bayesian evaluation of inequality constrained hypotheses. *Psychological Methods*, *19*(4), 511.

Jeffreys, H. (1961). Theory of probability, clarendon. Oxford.

Johnson, J. W., & LeBreton, J. M. (2004). History and use of relative importance

indices in organizational research. *Organizational Research Methods*, *7*(3), 238–257.

Klugkist, I., Kato, B., & Hoijtink, H. (2005). Bayesian model selection using encompassing priors. *Statistica Neerlandica*, *59*(1), 57–69.

Klugkist, I., Laudy, O., & Hoijtink, H. (2005). Bayesian eggs and bayesian omelettes: Reply to stern (2005).

Kruschke, J. K. (2011). Bayesian assessment of null values via parameter estimation and model comparison. *Perspectives on Psychological Science*, *6*(3), 299–312.

Kruskal, W., & Majors, R. (1989). Concepts of relative importance in recent scientific literature. *The American Statistician*, *43*(1), 2–6.

Lebanon, G., & Mao, Y. (2008). Non-parametric modeling of partially ranked data. *Journal of Machine Learning Research*, *9*(Oct), 2401–2429.

Mulder, J. (2010). *Bayesian model selection for constrained multivariate normal linear models* (PhD thesis). Utrecht University.

Mulder, J. (2014). Prior adjusted default bayes factors for testing (in) equality constrained hypotheses. *Computational Statistics & Data Analysis*, *71*, 448–463.

Mulder, J., & Olsson-Collentine, A. (2019). Simple bayesian testing of scientific expectations in linear regression models. *Behavior Research Methods*, 1–14.

Mulder, J., Hoijtink, H., & Klugkist, I. (2010). Equality and inequality constrained multivariate linear models: Objective model selection using constrained posterior priors. *Journal of Statistical Planning and Inference*, *140*(4), 887–906.

Stanley, R. P. (2011). Enumerative combinatorics, vol. 49. Cambridge university press.

Stanley, R. P., & others. (2004). An introduction to hyperplane arrangements.

Geometric Combinatorics, *13*, 389–496.

Stern, H. S. (2005). Model inference or model selection: Discussion of klugkist, laudy, and hoijtink (2005).