A Bayesian credible set for ranking parameters by relative magnitude

Justin Chumbley[1]

[1] Jacobs Center for Productive Youth Development, University of Zurich

Author Note

Correspondence concerning this article should be addressed to Justin Chumbley, Jacobs Center for Productive Youth Development, Andreasstrasse 15, CH-8050 Zurich. E-mail: justin.chumbley@pm.me

Abstract

The relative magnitude or ranking of statistical parameters - say treatment effects or variance components - is often more interesting or trustworthy than their absolute magnitude, or than strawman tests of their equality. Yet despite some existing tools for model selection, there is currently nothing akin to a simple $\beta\%$ confidence or credible interval for ranking parameters of interest. We therefore aim here to automatically identify the probable ranking of (an interesting subset of) components of a parameter vector or matrix: to find the most informative or *finest* assertion about this ranking at the desired $\beta\%$ credibility. For example, we may learn from linear regression on a given dataset that "$w_2 < w_1, w_3 < w_4$ holds with 95% probability", so our 4 coefficients of interest contain at least $r = 3$ distinct ranks. Or we may discover from the covariance matrix of a time-series that observation 1 has a U-shaped dependence with the subsequent 5 observations: "$\sigma_{13}, \sigma_{14}, \sigma_{15} < \sigma_{12}, \sigma_{16}$ with 95% probability". To do this we construct a new Bayesian posterior credible set of full rankings $\mathcal{C}_\beta$ using simple group-theoretic properties of the tesselated parameter space. The approach can be applied in both exact and approximate - MCMC and variational - settings and offers interpretational advantages over typical analysis. 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 parameters by relative magnitude

Introduction

Most behavioral, social and biological research scrutinizes theoretically-inspired hypotheses via classical tests or Bayesian model comparison. But premature or excessive impact of theory on data analysis also carries some risk of missing the truth, so there remains an important place for more open-ended or exploratory analyses. We apply this maxim to the case of infering inequalities between parameters, a task currently dominated by confirmatory methods for comparing so-called "informative hypotheses" (Gu, Mulder, Deković, & Hoijtink, 2014; Klugkist, Kato, & Hoijtink, 2005; Mulder & Olsson-Collentine, 2019; Mulder, Hoijtink, & Klugkist, 2010).

While there are now confirmatory methods for parameter ranking - via Bayes factors or the generalized order restricted information criterion (Gu et al., 2014) - there is currently nothing to match the much-beloved credible intervals. 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).

Such a critique perhaps overstates the respective dis/advantages of confirmatory and exploratory approaches. In many situations a confirmatory approach 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 readily incurs false positives: a hypothesis with overwhelming positive evidence over it's 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 false-positives.

We propose a novel way to summarize the posterior distribution of a parametric model, say standard linear or multilevel regression, which has been estimated either analytically or by MCMC without irksome inequality constraints. We assume this so-called "encompassing" model is well-specified and supported by theory or an Occam's razor to balance fit and complexity, e.g. the Bayes Factor, AIC, BIC or DIC. We show that our subsequent procedure requires no further complexity penalty. In analogy to the familiar "highest β% posterior density credible intervals", we seek to minimize the size or complexity of our posterior credible set - the number of distinct ranks - for a given posterior probability β. Just as standard confidence/credible intervals can become too wide to be useful at higher probabilities, arbitrarily increasing probability β generally results in inferior, coarser inferences about ranking. Contrast this with a confirmatory approach where one seeks to maximize model probability: 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). Here the confirmatory tool of choice is the Bayes factor with informative priors, because improper prior distributions do not furnish sensible Bayes factors (Jeffreys, 1961; Mulder, 2010). By sidestepping the computation of marginal likelihood and Bayes factors our procedure requires neither informative priors nor informative hypotheses.

Our specific aim in this paper is to rank multiple unobserved parameters without strong prior information or hypotheses. For example, parameters $\mathbf{w} = (w_1, w_2, w_3) \in \mathbb{W} := \mathbb{R}^3$ may originate in some regression quantifying the relationship between pupil's test scores Y and

their IQ in three exchangeable schools $i = 1, 2, 3$ which we would like to rank. We 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 $p(\mathbf{w}|y) = p(y|\mathbf{w})p(\mathbf{w})/p(y)$. Throughout this work, $P(\cdot)$ denotes a probability and $p(\cdot)$ a probability density. 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

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

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 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 $P(h|y) = \int_h p(\mathbf{w}|y)d\mathbf{w}$).

The remaining work is structured as follows. Section 3 contains technical details which readers may postpone or skip according to taste. It first rehearses some general definitions, before specifically defining full ranks, partial ranks and their algebraic structure. From these building blocks, we then define our Bayesian posterior $\beta\%$ credible set $\mathcal{C}_\beta \subseteq \mathbb{R}^d$ and offer a convenient construction. Section 4 introduces some simple simulations. Section 5 offers a very simple empirical example and we conclude in Section 6. Our work is partly inspired by the work of Lebanon and Mao (2008) on discrete preferences, and we have borrowed notation and definitions from Critchlow (2012), Lebanon and Mao (2008) and Stanley (2011).

Preliminary definitions

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 it's 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 probability of the corresponding subset of $\mathbb{W}$, see Figure 1. The context should distinguish whether π or $\mathcal{S}_d$

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 and 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 $\mathcal{C}_\beta$.

---

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 and 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 $\geq \beta\%$.

Simulations

Simulated data

To specify the ground truth in our simulations we first sampled $d = 3, 5, 7$ dimensional parameters $\mathbf{w}^*$ uniformly on $\Delta^{d-1} := \{(w_1, .., w_d) \in \mathbb{R}^d : w_t > 0, \sum_t w_t = \delta^*\}$, the $d-1$ dimensional regular simplex. This was repeated for δ between 0-5, with smaller δ^* posing a harder challenge for inference because it implies smaller distance between any two

components of $\mathbf{w}^*$ 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 proposed construction for $\mathcal{C}_\beta$ requires a way to readily calculate the posterior probability of any general partial ranking. In practice, one might evaluate closed-form integrals over the relevant subsets of posterior $p(\mathbf{w}|y)$ or use a sample from that posterior. In the latter case, the posterior probability of each partial ranking is just the fraction of samples satisfying it's set indicator function, up to Markov error. Our simulations below pursue the former strategy, borrowing the deterministic scheme due to Mulder (2014) to calculate the probability of each partial ranking in the most credible chain. In agreement with the data generating process above, the likelihood $p(y|\mathbf{w})$ was a standard linear regression model.

The goals of our simulation

Knowing the ground truth $\mathbf{w}^*$ and its corresponding full ranking $f(\mathbf{w}^*)$ our principle questions are

1) whether $\mathcal{C}_\beta$ is consistent with the true full ranking, i.e. $f(\mathbf{w}^*) \in \mathcal{C}_\beta$?

2) how much "information" is retained in $\mathcal{C}_\beta$?

The first question expresses the minimal requirement that $\mathcal{C}_\beta$ does not contradict the truth. But we additionally want $\mathcal{C}_\beta$ to be as informative as possible, ideally faithfully retaining *all* distinctions made in the true ranking $f(\mathbf{w}^*)$. Hence our second question above.

In practice, we answer the first question as follows. Let $[r] := \{1, ..., r\}$ be the distinct bins in the observed composition $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_r)$ of $\mathcal{C}_\beta$. Our algorithm for $\mathcal{C}_\beta$ will assign each component w_i of $\mathbf{w}$ to exactly one of these r bins, which we denote $l(w_i) \in [r]$. We therefore ask whether $w_i \leq w_j$ in the true full rank implies $l(w_i) \leq l(w_j)$. If this is true for all w_i, w_j we say that the partial rank contains or satisfies the true full rank, otherwise we say it violates the full rank. This procedure can answer question 1.

Turning to the second question, we use $q = r/(d-1)$ with values between 0 to 1 to measure the quality of $\mathcal{C}_\beta$. Here r is the number of distinctions - inequalities or bars "|" - in $\mathcal{C}_\beta$ and $d-1$ the true number in $f(\mathbf{w}^*)$. We choose this as a simple measure which is comparable over simulations with different settings of d. 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. We report q in each condition of our simulation.

Our simulations also 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 $\mathcal{C}_\beta = \mathbb{W}$.

3) What is the proportion of false negatives and false positives so defined?

Results

1) We found that our inferred partial rank $\mathcal{C}_\beta$ rarely violated the ground truth, meaning $f(\mathbf{w}^*) \notin \mathcal{C}_\beta$. Such a violation only occurred in 0.02 percent of the simulations.

2) 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.

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

We naturally expected q to improve when the components $\mathbf{w}^*$ differed more in absolute terms. Figure 4 therefore plots *jittered* q against the per-simulation standard deviation over components of $\mathbf{w}^*$, namely $\frac{1}{d}\sum_{i=1}^{d}(w_d^* - \bar{\mathbf{w}}^*)^2$, with $\bar{\mathbf{w}}^*$ denoting the mean over components of the ground truth $\mathbf{w}^*$.

3) Table 2 summarizes the false positive and false negative rate 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 very coarse 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", 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

Real data

To better understand percieved discrimination against female applicants in the hiring process, Carlsson and 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 discrimination, see Mulder and 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"

Table 3

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

	FALSE	TRUE
FALSE	117	0
TRUE	3	0

explanation of social psychology trumps the "same-gender bias" explanation, see Carlsson and 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

It is common to test parameters of interest or derive confidence intervals. Yet classical tests are widely discouraged and confidence intervals become increasingly opaque for questions jointly involving multiple parameters: intervals become high-dimensional volumes with arbitrary geometry. We consider another way to summarize multiparameter hypotheses. This approach is useful when we believe component parameters of the estimand $\mathbf{w}^*$ can in principle be ranked, but can only *partially* discern this rank due to posterior uncertainty. Our approach entails 1) constructing a set of increasingly coarse tilings of the parameter space, each interpretable as a partial ranking 2) selecting the smallest of these with at least β probability, namely $\mathcal{C}_\beta$, interpretable as the finest credible ranking supported by the data. We have shown that our credible set $\mathcal{C}_\beta$ has low error rates. It does not guarantee to preserve all rankings in the ground truth parameter $\mathbf{w}^*$, but the amount of detail preserved increases

Table 4

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

	FALSE	TRUE
FALSE	0	34
TRUE	0	86

with both the sample size and the absolute differences between components of the true underlying $\mathbf{w}^*$.

Just as most conventional credible sets typically contain the maximum posterior point $\mathbf{w}_{MAP} = \underset{\mathbf{w}\in\mathbb{W}}{\text{argmax}}\ P(\mathbf{w}|y)$, our credible set of full rankings $\mathcal{C}_\beta$ necessarily contains the maximum posterior rank $\pi_{MAP} = \underset{\pi\in\mathcal{S}_d}{\text{argmax}}\ P(\pi|y)$. It can also be shown that $\mathcal{C}_\beta$ must be a connected set of regions, meaning that any full order in $\pi \in \mathcal{C}_\beta$ is at most Kendall distance 1 from some other full order $\pi' \in \mathcal{C}_\beta$, see pg 89-91 of Stanley and others (2004). This excludes credible sets with complicated topology (e.g. disconnected sets with Euler characteristic $\neq 1$) and naturally mirrors the assumed unimodality of our prior and likelihood. Thus, while conventional highest density sets can have arbitrary topology - multiple disconnected regions, as for mixture models - $\mathcal{C}_\beta$ matches our assumption that the underlying $\mathbf{w}^*$ is a single fixed but unobservable real vector.

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 (not reported here).

Regarding the interpretation of $\mathcal{C}_\beta$, there are a couple of points to emphasize. First,

Table 5

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

	FALSE	TRUE
FALSE	0	5
TRUE	0	115

the trivial ranking $\mathcal{C}_\beta = \mathbb{W}$ amounts to a completely inconclusive partial ranking of parameters. This result does not in itself support the popular omnibus equality hypothesis $w_i = w_j$ for all $i \neq j$, i.e. it does *not* imply $p(w_1 = w_2 = ... = w_d|y) \geq \beta$. Rather $\mathcal{C}_\beta = \mathbb{W}$ just reflects high posterior uncertainty. In fact, any equality hypothesis has probability 0 under any continuous prior $p(\mathbf{w})$. Such a choice implies we believe components of the underlying estimand $\mathbf{w}^*$ can be totally ordered with probability one. If this were doubtful, one might turn to other confirmatory methods (for example, Kruskal & Majors, 1989, Johnson and LeBreton (2004), Mulder and Olsson-Collentine (2019)). Future work should develop exploratory methods which explicitly cater for both equality and inequality statements. We do not pursue this in this initial presentation, partly to side-step any distracting complications such as the Lindley paradox.

Note however that any non-trivial conclusion $\mathcal{C}_\beta \neq \mathbb{W}$ 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 defined in the previous paragraph. 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 $\mathcal{C}_\beta$ (Madathil, Joseph, Hardy, Rousseau, & Nicolau, 2018).

Table 6

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

	FALSE	TRUE
FALSE	0	1
TRUE	0	119

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 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

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

Table 8

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

	FALSE	TRUE
FALSE	0	1
TRUE	0	119

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.

Madathil, S., Joseph, L., Hardy, R., Rousseau, M.-C., & Nicolau, B. (2018). A bayesian approach to investigate life course hypotheses involving continuous exposures. *International Journal of Epidemiology*, *47*(5), 1623–1635.

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).

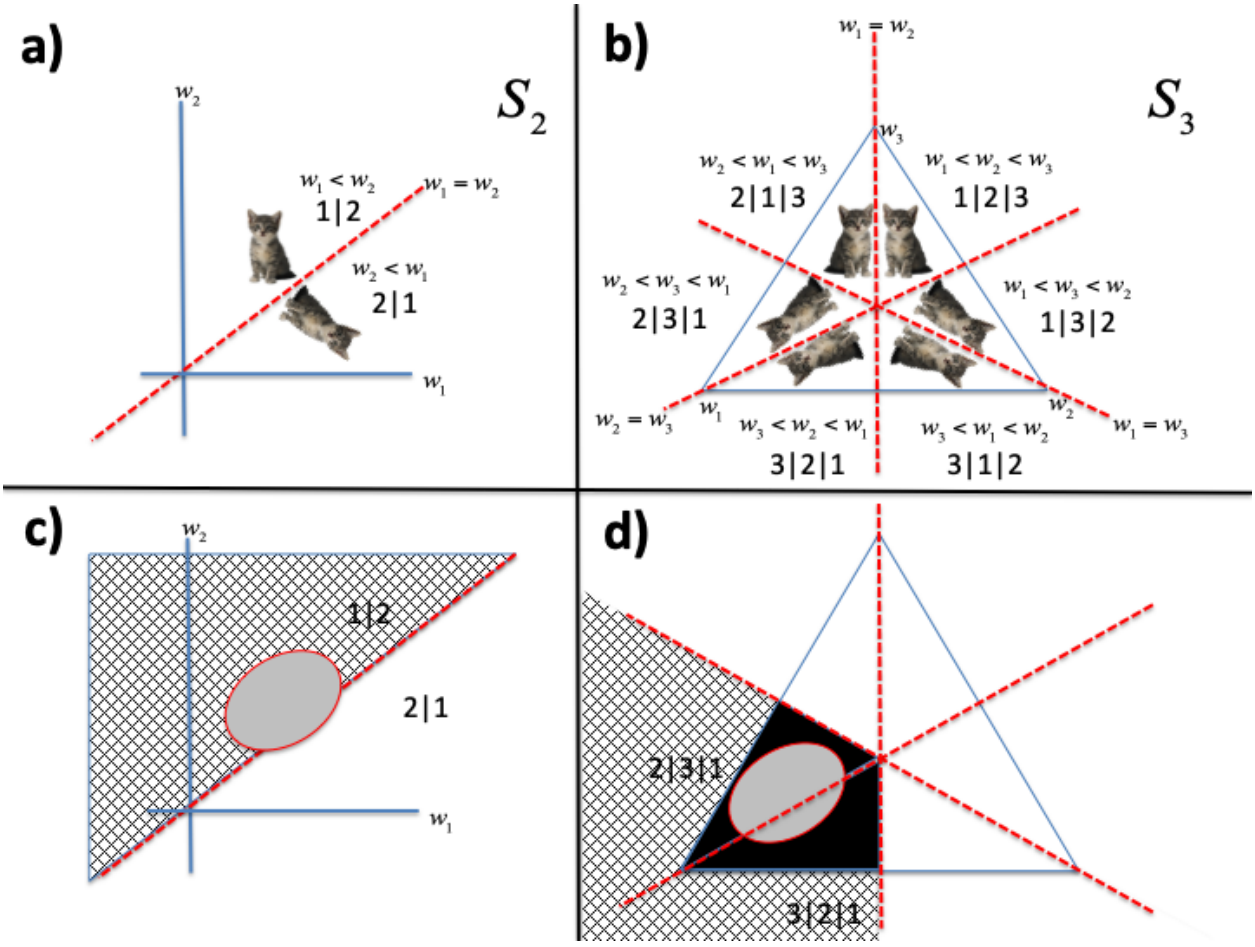

Figure 1. Rankings in 2 and 3 dimensions. a) There are 2 full rankings in 2d, the reflections about $w_1 = w_2$. b) The $6 = 3!$ full rankings in 3d reflect across $w_1 = w_2, w_2 = w_3$ and $w_1 = w_3$, the "braid arrangement" of hyperplanes (dashed lines) intersecting at $w_1 = w_2 = w_3$. We visualize their intersection with a 2d plane orthogonal to $w_1 = w_2 = w_3$. Note that the 2 dimensional unit simplex $\Delta^{3-1} = \{\mathbf{w} : w_1 + w_2 + w_3 = 1 \text{ and } w_i \geq 0\}$ belongs to one such plane, as suggested by the superimposed equilateral triangle. c) A standard 95% credible set in solid grey alongside our credible ranking $\mathcal{C}_{95\%} = 1|2$ in hashed grey. d) Similarly for 3 dimensions (or a 2d simplex). Here $\mathcal{C}_{95\%} = 2, 3|1$. If our parameter space is $\mathbb{W} = \mathbb{R}^3$ then $\mathcal{C}_{95\%} = 2, 3|1$ is the hashed grey region which extends infinitely bottom left and projectively into the axis of gaze. When parameter space is restricted to the simplex $\mathbb{W} := \Delta^{3-1}$, we can still use the $2, 3|1$ notation with the understanding that it denotes the black region, a subset of a 2 dimensional real subspace of $\mathbb{R}^3$. Figure 2 illustrates a procedure for identifying the credible sets just discussed.

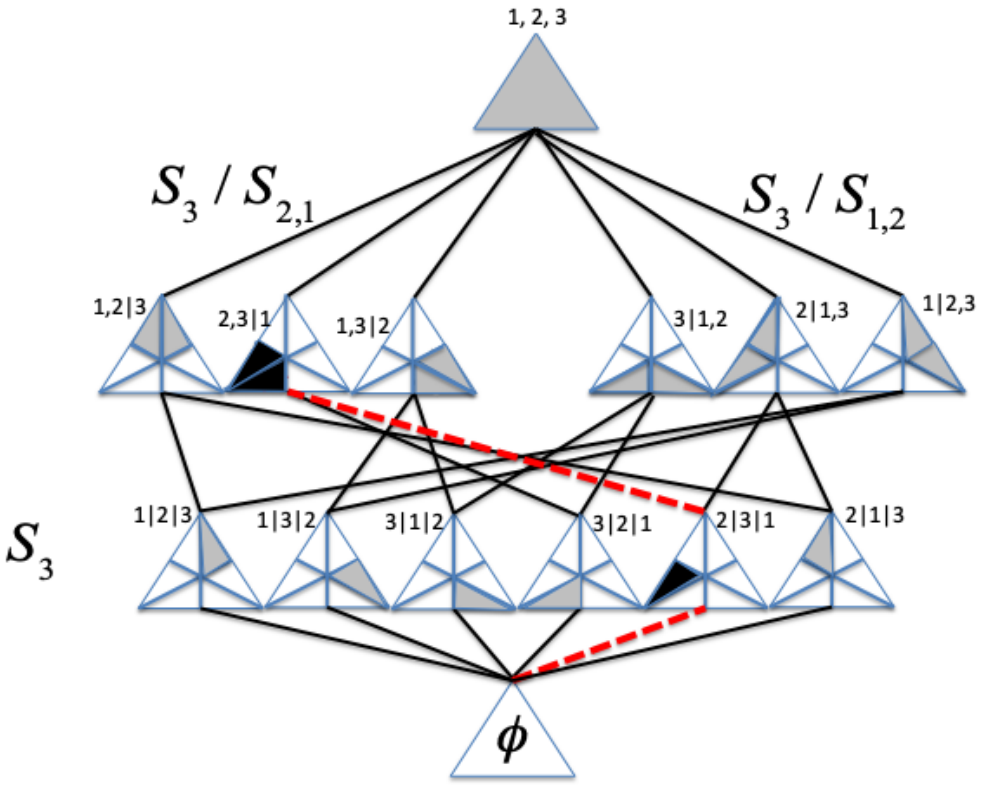

Figure 2. The ranking lattice. Hasse diagram depicting partial rankings in fig 1b. Each rank h denoted in grey or black satifisies some inequalities, say $w_1 < w_2 < w_3$ (leftmost row 3). Row 3 depicts the finest resolution of all $d! = 3! = 6$ full ranks. Ranks h, h' on different rows may or may not be ordered by coarseness (set inclusion), hence a *partial* order. An upward path connecting h, h' in the diagram indicates that $h \subseteq h'$, so $h \preceq h'$. For example, $2|1|3 \preceq 2|1, 3$ and $1|2|3 \preceq 1, 2|3$. If h' is the smallest such ranking, we say h' "covers" h, notationally $h \vdash h'$. We start at the most probable full ranking h_{MAP} at the bottom and ascend in search of the finest 95% credible rank (red dashed lines). Refering to Figure 1d, the maximum posterior full rank would be $h_{MAP} = 2|3|1$ in our row 3 and $\mathcal{C}_{95\%} = 2, 3|1 = 2|3|1 \cup 3|2|1$ in our row 2: the latter being coarser than the former. Associated to each such sequence is a "cumulative density function" of monotone increasing probabilities, as depicted in Figure 3.

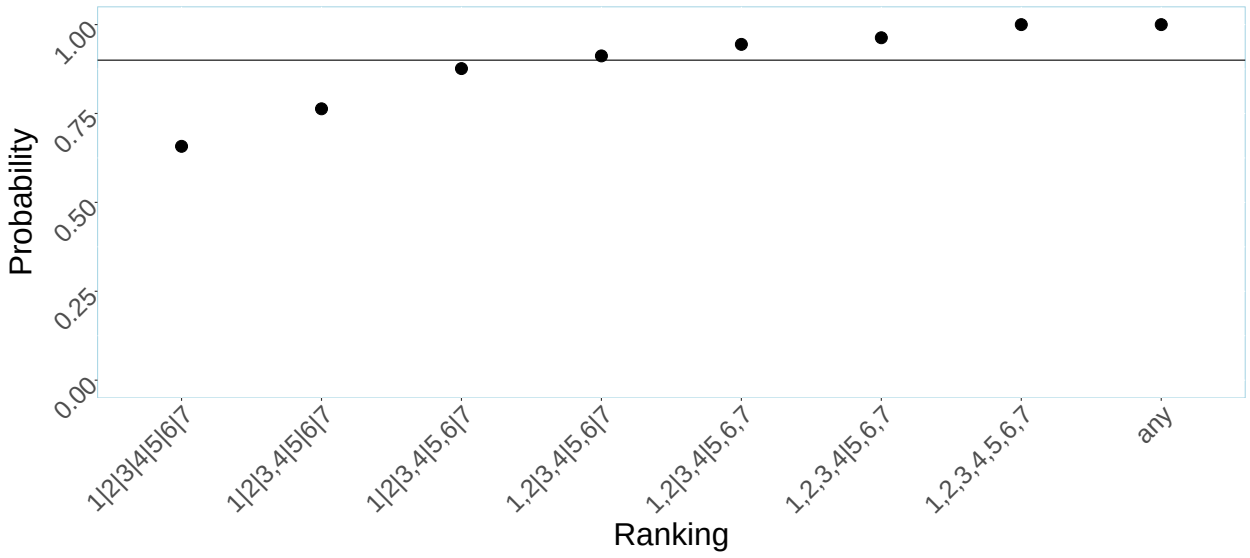

Figure 3. An example cdf over partial rankings, based on the posterior distribution $p(\mathbf{w}|y)$ that obtained in one of our simulations. The true underlying ranking in this simulation was 1|2|3|4|5|6|7. This illustrates how the $\mathcal{C}_{95\%}$ for this 7-dimensional parameter is obtained by our recursive procedure. The x-axis labels express the *most credible chain* up the ranking lattice, see Figure 2 and Section 3. Note that coarser rankings on the right are progressively more probable. For a chosen threshold of β, the $\mathcal{C}_\beta$ is the smallest (leftmost) value of the x-axis with a y-value greater than threshold β, here depicted as a black horizontal line.

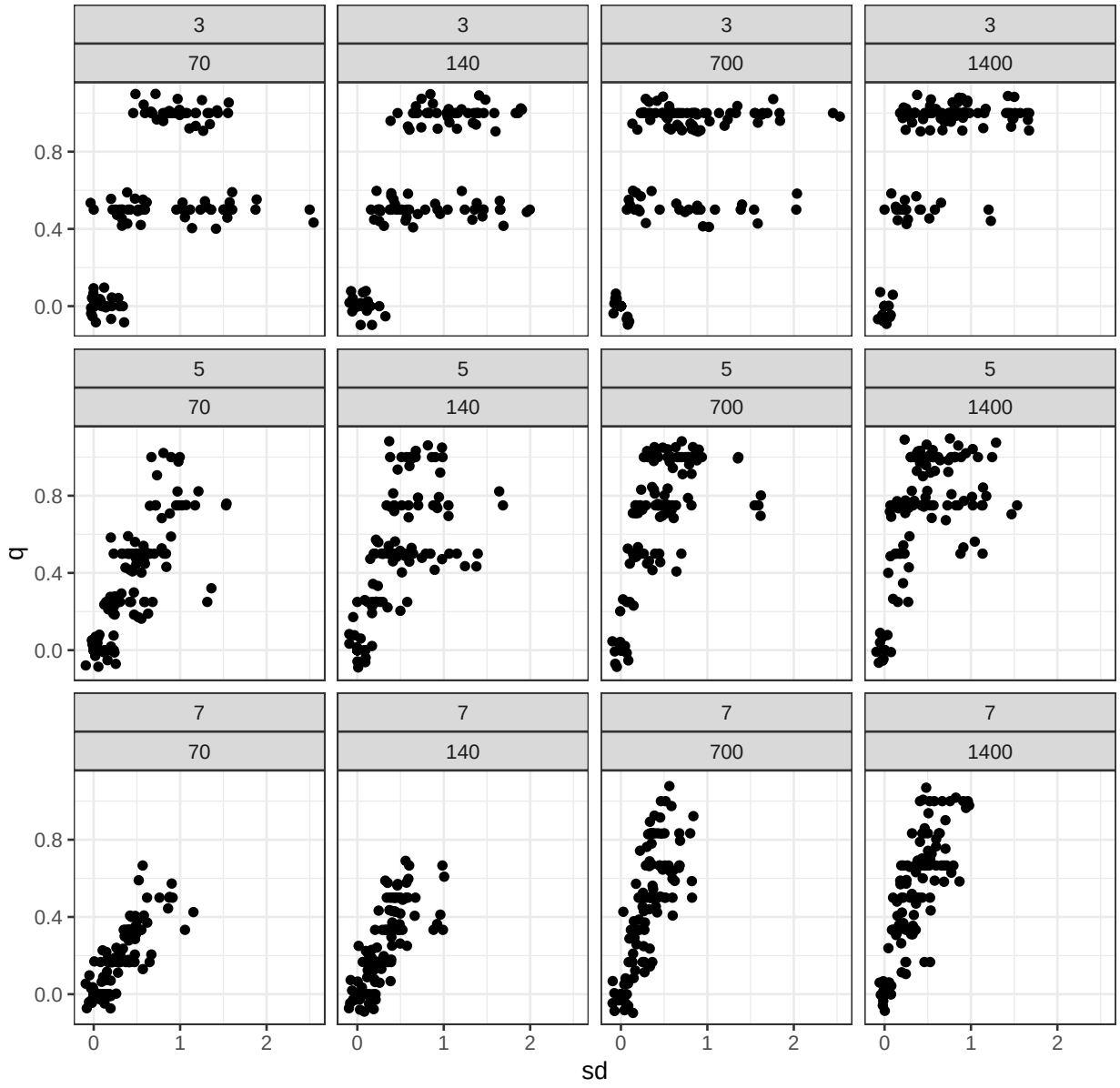

Figure 4. Quality q (jittered) against the per-simulation standard deviation over components. This illustrates the intuitive fact that, in practice, the quality of the relative measure $\mathcal{C}_\beta$ does depend on the absolute underlying differences between component parameters. Rows vary over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