



## anchors: Software for Anchoring Vignette Data

**Jonathan Wand**  
Stanford University

**Gary King**  
Harvard University

**Olivia Lau**  
Harvard University

---

### Abstract

When respondents use the ordinal response categories of standard survey questions in different ways, the validity of analyses based on the resulting data can be biased. Anchoring vignettes is a survey design technique intended to correct for some of these problems. The **anchors** package in R includes methods for evaluating and choosing anchoring vignettes, and for analyzing the resulting data.

*Keywords:* survey research, item response, differential item functioning, R.

---

## 1. Introduction to anchoring vignettes

Consider a survey question along with response categories that is asked as a self-assessment,

How free do you think you are to express yourself without fear of government reprisal? (1) Completely Free, (2) Very Free, (3) Moderately Free, (4) Slightly Free, (5) Not Free at All.

One key difficulty of analyzing the results from such a survey question is the possibility that individuals apply different standards in the selection of a response category. Researchers have tried to ameliorate the problems of interpersonal and cross-cultural incomparability in survey research with careful question wording, translation (and back translation), focus groups, cognitive debriefing, and other techniques, most of which are designed to improve the survey question. In contrast, anchoring vignettes is a technique that seeks to bring additional data to bear on the problem.

For example, vignettes corresponding to the above political freedom question attempt to describe hypothetical individuals who have different levels of freedom from government reprisal. The following six vignettes are intended to correspond to distinct levels of political freedom in order of decreasing freedom,

**vign1** [Kay] does not like many of the government's policies. She frequently publishes her opinion in newspapers, criticizing decisions by officials and calling for change. She sees little reason these actions could lead to government reprisal.

**vign2** [Michael] disagrees with many of the government's policies. Though he knows criticism is frowned upon, he doesn't believe the government would punish someone for expressing critical views. He makes his opinion known on most issues without regard to who is listening.

**vign3** [Bob] has political views at odds with the government. He has heard of people occasionally being arrested for speaking out against the government, and government leaders sometimes make political speeches condemning those who criticize. He sometimes writes letters to newspapers about politics, but he is careful not to use his real name.

**vign4** [Connie] does not like the government's stance on many issues. She has a friend who was arrested for being too openly critical of governmental leaders, and so she avoids voicing her opinions in public places.

**sign5** [Vito] disagrees with many of the government's policies, and is very careful about whom he says this to, reserving his real opinions for family and close friends only. He knows several men who have been taken away by government officials for saying negative things in public.

**vign6** [Sonny] lives in fear of being harassed for his political views. Everyone he knows who has spoken out against the government has been arrested or taken away. He never says a word about anything the government does, not even when he is at home alone with his family.

After each of these vignettes, a corresponding evaluation question is asked with the same response categories as for the self-assessment.

How free do you think [name] is to express [him/her]self without fear of government reprisal? (1) Completely Free, (2) Very Free, (3) Moderately Free, (4) Slightly Free, (5) Not Free at All

In this article we illustrate different approaches for either setting the scale of self-assessments relative to vignettes in a non-parametric setting, or using vignettes to identify a parametric model of differential item functioning (DIF). The focus is on providing annotated examples using techniques proposed in King, Murray, Salomon, and Tandon (2004) and King and Wand (2007). Specifically, we discuss the logic of an empirical analysis of surveys with anchoring vignettes, along with detailed comments on the results of a particular survey, using the **anchors** package (Wand, King, and Lau 2011). A companion document is also available (Wand and King 2007), providing a quick overview of the functions and supporting materials provided by the **anchors** package. The **anchors** package is written in R (R Development Core Team 2011), and is available from the Comprehensive R Archive Network at <http://CRAN.R-project.org/package=anchors>. The latest version of this software and user documentation are also available at the **anchors** website, <http://wand.stanford.edu/anchors/>.

The examples throughout the remainder of this article will rely on survey data measuring political freedom that are made available as part of **anchors** package. The package and the data set can be loaded in the R via,

```
R> library("anchors")
R> data("freedom")
```

## 2. Non-parametric analysis

In this section, we analyze the relative ranks of self-assessment responses compared to the vignette responses. DIF is addressed by using the common set of vignettes given to all respondents to reveal how each respondent uses the survey response scale. This approach depends on assuming a particular “true” ordering of the vignettes and the choice of which vignettes to employ when analyzing self-assessments. As a result, considerable attention is also devoted in this section to the analysis of methods for analyzing the order of vignettes, selecting subsets, and diagnosing problems.

### 2.1. Models of relative ranks

Let  $y_i$  be the self-assessment response and  $z_{i1}, \dots, z_{iJ}$  be the  $J$  vignette responses, for the  $i$ th respondent. For respondents with consistently ordered rankings on all vignettes ( $z_{j-1} < z_j$ , for  $j = 2, \dots, J$ ), we create the DIF-corrected self-assessment  $C_i$

$$C_i = \begin{cases} 1 & \text{if } y_i < z_{i1} \\ 2 & \text{if } y_i = z_{i1} \\ 3 & \text{if } z_{i1} < y_i < z_{i2} \\ \vdots & \vdots \\ 2J + 1 & \text{if } y_i > z_{iJ} \end{cases} \quad (1)$$

Respondents who give tied or inconsistently ordered vignette responses may have an interval values of  $C$ , if the tie/inconsistency results in multiple conditions in equation 1 appearing to be true. A more general definition of  $C$  is defined as the minimum to maximum values among all the conditions that hold true in equation 1. Values of  $C$  that are intervals, rather than scalar, represent the set of inequalities over which the analyst cannot distinguish without further assumption.

Table 1 drawn from King and Wand (2007) gives all thirteen examples that can result from two vignette responses and a self-assessment. Note that beyond the (in)equalities between the responses, the specific numeric values of  $y$ ,  $z_1$ , and  $z_2$  are irrelevant for defining  $C$  and therefore there is no need to make reference to them in the table.

Examples 1–5 have both vignettes correctly ordered and not tied, with the result for  $C$  being a scalar. The vignette responses are tied in Examples 6–8, which produces a censored value for  $C$  only if the self-assessment is equal to them. Examples 9–13 are for survey responses that incorrectly order the vignettes. Within each set, the examples are ordered by moving  $y$  from left to right.

One of the things researchers like to do is characterize the distribution of  $C$  for the survey as a whole or to examine contrasts between subgroups. If values of  $C$  are not exclusively scalar values, a researcher is faced with a challenge of how to accomplish this task. In this section we describe four approaches to answering this question by making different assumptions of how to combine scalar and interval values of  $C$ .

Ex.	Survey responses	1:	2:	3:	4:	5:	$C$
		$y < z_1$	$y = z_1$	$z_1 < y < z_2$	$y = z_2$	$y > z_2$	
1	$y < z_1 < z_2$	1	0	0	0	0	1
2	$y = z_1 < z_2$	0	1	0	0	0	2
3	$z_1 < y < z_2$	0	0	1	0	0	3
4	$z_1 < y = z_2$	0	0	0	1	0	4
5	$z_1 < z_2 < y$	0	0	0	0	1	5
6	$y < z_1 = z_2$	1	0	0	0	0	1
7	$y = z_1 = z_2$	0	1	0	1	0	2,3,4
8	$z_1 = z_2 < y$	0	0	0	0	1	5
9	$y < z_2 < z_1$	1	0	0	0	0	1
10	$y = z_2 < z_1$	1	0	0	1	0	1,2,3,4
11	$z_2 < y < z_1$	1	0	0	0	1	1,2,3,4,5
12	$z_2 < y = z_1$	0	1	0	0	1	2,3,4,5
13	$z_2 < z_1 < y$	0	0	0	0	1	5

Table 1: All examples with two vignettes: this table gives calculations for the nonparametric estimator  $C$  for all possible examples (sans non-response) with two vignette responses,  $z_1$  and  $z_2$  (intended to be ordered as  $z_1 < z_2$ ), and a self-assessment,  $y$ . From [King and Wand \(2007\)](#).

Ex.	$C$	$P(\hat{C} = 1   C)$	$P(\hat{C} = 2   C)$	$P(\hat{C} = 3   C)$	$P(\hat{C} = 4   C)$	$P(\hat{C} = 5   C)$
1,6,9	1	1	0	0	0	0
2	2	0	1	0	0	0
3	3	0	0	1	0	0
4	4	0	0	0	1	0
5,8,13	5	0	0	0	0	1
7	2:4	0	$p_{2,2:4}$	$p_{3,2:4}$	$p_{4,2:4}$	0
10	1:4	$p_{1,1:4}$	$p_{2,1:4}$	$p_{3,1:4}$	$p_{4,1:4}$	0
11	1:5	$p_{1,1:5}$	$p_{2,1:5}$	$p_{3,1:5}$	$p_{4,1:5}$	$p_{5,1:5}$
12	2:5	0	$p_{2,2:5}$	$p_{3,2:5}$	$p_{4,2:5}$	$p_{5,2:5}$

Table 2: Components of a histogram based on survey response orderings enumerated in Table 1: this table specifies the unknown values in  $P(\hat{C} = j)$  that will be filled in by different approaches. The values  $p_{j,m:n} = P(\hat{C} = j | C \in \{m, \dots, n\}) \in [0, 1]$  are described in the text.

In the derivations that follow, a remapped version of  $C$  creating the values for the histogram will be referred to generically as  $\hat{C}$ , which is a function of  $p_{j,m:n} = P(\hat{C} = j | C \in \{m, \dots, n\}) \in [0, 1]$ , and  $p_{m:n} = \sum_{j=1}^J p_{j,m:n} \in \{0, 1\}$ . Assume that  $p_{j,m:n} = 0$  if  $j \notin \{m, \dots, n\}$  and for scalar values of  $C$ ,  $p_{j,j:j} = P(\hat{C} = j | C = j) = 1$ . If the sum  $p_{m:n}$  is equal to zero then that value of  $C$  is omitted from the  $\hat{C}$  histogram. Each bar of the  $\hat{C}$  histogram is calculated by,

$$P(\hat{C} = j) = \sum_{m=1}^{2J+1} \sum_{n=m}^{2J+1} N_{m:n} p_{j,m:n} / D \quad (2)$$

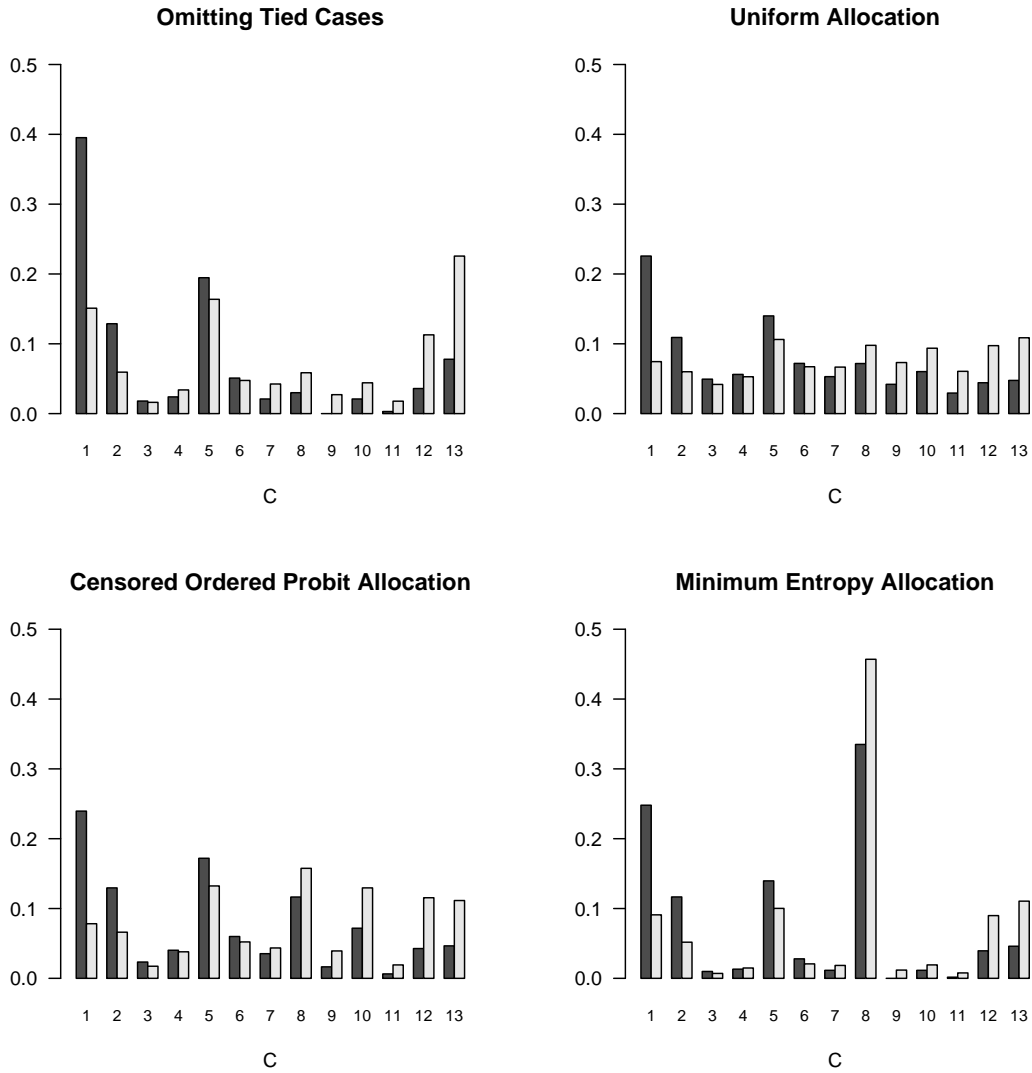


Figure 1: Approaches to summarizing the distribution of  $C$ : all figures are based on same political freedom model specified as, `self vign2 + vign1 + vign3 + vign5 + vign4 + vign6`. Dark histogram is for Oceania subsample, light histogram is for East Asia subsample.

where  $N_{m:n}$  is the number of cases with  $C = \{m, \dots, n\}$  and

$$D = \sum_{m=1}^{2J+1} \sum_{n=m}^{2J+1} N_{m:n} p_{m:n}. \quad (3)$$

Examples for the two vignette case are presented in Table 2. The rest of this section investigates alternative assumptions for calculating the  $p_{j,m:n}$  values.

Using the political freedom data, Figure 1 illustrates the results of the different approach which will each be discussed in turn. The comparison of the distribution of  $C$  is made between

two regions, East Asia and Oceania. Explanations of the code for producing these figures is presented in the following section. Overall, each of the four graphs look quite different and the explanations for why that is will be considered in the following four subsections.

*Approach 1: Omitting interval values*

A simple approach omits cases with interval values of  $C$ , and plots the histogram of the scalar values. Specifically,  $p_{m:n} = 0$  if  $m \neq n$ .

Unless the the observations dropped due to their interval values of  $C$  are treated as missing completely at random (MCAR), this approach obviously loses potentially useful information about the relative propensity of interval values of to occur among different values of  $C$ . Nonetheless, this approach does provide a useful baseline against which to compare the other approaches that use different weighting schemes to allocate interval values of  $C$  to their component scalars.

*Approach 2: Uniform allocation within intervals*

The second approach creates a histogram by allocating equal fractions of a case with an interval  $C$  value to the values it spans. Specifically,  $p_{j,m:n} = 1/(n-m+1)$  for all  $j \in \{m, \dots, n\}$ .

The greater number of cases with interval values and the the more uniform and more variable are the values spanned by the interval, the more uniform the resulting histogram will look. This approach was used in [King \*et al.\* \(2004\)](#).

*Approach 3: Censored ordered probit*

To define this approach developed in [King and Wand \(2007\)](#), it is helpful to begin with the classic ordered probit and then generalize it to take into account of interval responses. First denote  $C_i^*$  (for respondent  $i = 1, \dots, n$ ) as a continuous unobserved dependent variable, and  $x_i$  as a vector of explanatory variables (and for identification, with no constant term). Instead of observing  $C_i^*$ , we instead see  $C_i$  through a specific observation mechanism. Thus, for scalar values, the mechanism is

$$C_i = c \quad \text{if } \tau_{c-1} \leq C_i^* < \tau_c \quad (4)$$

with thresholds  $\tau_c$  (where  $\tau_0 = -\infty$ ,  $\tau_{2J+1} = \infty$ , and  $\tau_{c-1} < \tau_c$ , and for  $c = 1, \dots, 2J + 1$ ). Note that the model here is entirely in terms of  $C$ , without reference to the underlying self-assessment and vignette responses.

The probability of observing an outcome in category  $c$  can be operationalized by specifying a distribution for  $C_i^*$ . One specific operationalization is the ordered probit model which conditions on some observed values  $x_i$ , combined by means of a parameterized linear additive function,

$$\Pr(C_i = c|x_i) = \int_{\tau_{c-1}}^{\tau_c} N(C_i^*|x_i\beta)dy. \quad (5)$$

This model can be generalized by adding notation for interval values of  $C$ , which we do by altering the observation mechanism in Equation 4 to:

$$C_i = \{m, \dots, n\} \quad \text{if } \tau_{m-1} \leq C_i^* < \tau_n \quad (6)$$

such that,

$$\Pr(C_i = \{m, \dots, n\} | x_i) = \int_{\tau_{m-1}}^{\tau_n} N(C_i^* | x_i \beta) dy. \quad (7)$$

Conditional on a given interval value of  $C$ , the probability of a specific value  $\hat{C} = j$  is,

$$p_{ij:m:n} = \frac{\Pr(C_i = j | x_i)}{\Pr(C_i = \{m, \dots, n\} | x_i)} \quad \text{for } j \in \{m, \dots, n\} \quad (8)$$

The new expression in Equation 8 conditions on  $C_i$  by normalizing the probability to sum to one within the set  $C_i$  and zero outside that set. For scalar values of  $C_i$ , this expression simply returns the observed category:  $p_{ij,j:j} = \Pr(\hat{C}_i = j | C_i = j, x_i) = \Pr(\hat{C}_i = j | C_i = j) = 1$  for category  $j$ , and 0 otherwise. For vector valued  $C$ , Equation 8 puts a probability density over the categories within  $C_i = m, \dots, n$ , which in total sum to one.

#### *Approach 4: Minimum entropy*

We suggest also considering the two boundary cases defined by the allocations that provide alternatively the most and least uniform possible histograms. These are described as the maximum entropy  $\hat{C}$  and the minimum entropy  $\hat{C}$ , respectively. Specifically, the minimum remaps interval values of  $C$  in a manner which tries to make as big as possible in the smallest  $\hat{C}$ . A heuristic is to posit that if the interval valued respondents actually had a single true value of  $C$ , we are trying to make as many people as possible have the same single valued  $C$ . From the perspective of a survey instrument designed to differentiate between respondents, this is a worst case scenario.

In this approach  $p_{j,m:n} = 1$  for some  $j$  in each  $m$  and  $n$  interval. The particular  $j$  is selected to minimize the entropy function,

$$H = - \sum_{j=1}^{2J+1} P(\hat{C} = j) \ln P(\hat{C} = j) \quad (9)$$

where for convenience we define  $-0 \ln(0) \equiv 0$  (since  $\lim_{a \rightarrow 0^+} a \ln(1/a) = 0$ ). The measure of entropy is such that  $H = 0$  if and only if  $P(\hat{C} = j) = 1$  for  $j$ . For any given  $J$ ,  $H$  is at a maximum, and equal to  $\ln(2J + 1)$ , when  $p_j = 1/(2J + 1)$  (for all  $j$ ), which is a uniform frequency distribution.

## 2.2. Ordering of vignettes

In the definition of  $C$ , the ordering of the vignettes is fixed *a priori*. The ordering of vignettes is normally chosen by the researchers, but it is also possible to draw upon a consensus ordering by the respondents, so long as only one ordering is used for all respondents for the analysis. Differences between hypothesized ordering of the researchers and the consensus ordering may fruitfully be used for diagnosing problems in the survey instruments, particularly when translating the questions for use in different languages.

Starting with the the order of the political freedom vignettes given above, we consider how responses are ordered empirically with the `anchors.order()` function,

```
R> z1 <- anchors.order(~ vign1 + vign2 + vign3 + vign4 + vign5 + vign6,
+   freedom)
```

where `data("freedom")` provides the `freedom` data set that will also be used in subsequent examples. The list of vignettes is declared using the formula/addition syntax of R, for ease of use and consistency with the `anchors()` function to be introduced shortly. The second argument of `anchors.order()` declares the dataframe used by the function.

The `summary()` method for `anchors.order` class produces the following information.

```
R> summary(z1, top = 10, digits = 3)
```

ANCHORS: SUMMARY OF VIGNETTE ORDERING

Treatment of ties: represent as sets

Number of cases with at least two distinct vignette responses: 3267  
 and with no violations of natural ordering: 773  
 and with no more than 1 violation of natural ordering: 1654  
 and with no more than 2 violation of natural ordering: 2299

Proportion of cases a vignette (row) is less than another (column):

	<1	<2	<3	<4	<5	<6
1	NA	0.235	0.651	0.757	0.715	0.775
2	0.321	NA	0.663	0.732	0.707	0.754
3	0.080	0.121	NA	0.457	0.363	0.575
4	0.056	0.080	0.138	NA	0.183	0.374
5	0.067	0.068	0.198	0.339	NA	0.495
6	0.055	0.070	0.081	0.100	0.103	NA

Upper tri =  $p_{\{ij\}} - p_{\{ji\}}$  (negative values suggest misorderings)

Lower tri =  $1 - p_{\{ij\}} - p_{\{ji\}}$  (big numbers means many ties)

	1	2	3	4	5	6
1	NA	-0.085	0.571	0.702	0.649	0.720
2	0.444	NA	0.542	0.652	0.639	0.684
3	0.269	0.405	NA	0.320	0.165	0.494
4	0.215	0.218	0.477	NA	-0.156	0.275
5	0.187	0.225	0.171	0.345	NA	0.392
6	0.188	0.440	0.176	0.526	0.402	NA

Top 10 orderings (out of 615 unique orderings):

	Frequency	Proportion	Ndistinct	Nviolation
{1,2,3,4,5,6}	233	0.0666	1	0
{1,2},{3,4,5,6}	129	0.0369	2	0
{1,2},3,{4,5,6}	113	0.0323	3	0
{1,2},{3,5},{4,6}	98	0.0280	3	1
{1,2},{3,4,5},6	97	0.0277	3	0
2,1,{3,4,5,6}	93	0.0266	3	1
{1,2},3,{4,5},6	79	0.0226	4	0
{1,2},5,{3,4,6}	63	0.0180	3	2



2,1,3,{4,5},6	58	0.0166	5	1
2,1,{3,5},{4,6}	57	0.0163	4	2

Working from the bottom of the output to the top, the different observed rank orderings of the vignettes are counted and presented in descending order of frequency. For brevity the `top` option in `summary(..., top = )` limits how many of the most frequent observations are shown. In the first column, the numbers in the first column are the index for the vignettes given the order in which they were written (left to right) in the formula passed to `anchors.order()`. It happens in this example that the index values also correspond to the numbers in the labels of the vignettes, but that need not be the case. Vignettes that have the same response value are placed within `{}` brackets.

The most common set of responses is to give all vignettes the same value (Frequency = 233), and the next most common ranking is giving one value for `{vign1,vign2}`, and another greater value for `{vign3,vign4,vign5,vign6}`.

The two columns `Ndistinct` and `Nviolation` are included to facilitate alternative orderings of the summary of vignette rankings, as well as a quick source of information. For example, the fourth row, `{1,2},{3,5},{4,6}`, has `Ndistinct = 3` distinct response levels. Although this is easily calculated by counting the number of distinct sets in the first column, having `Ndistinct` column provides a summary of how many different response values are observed for each constellation of ordering of vignettes. In this example, since there are only five response categories but six vignettes there must be at least two vignettes that have the same response values. The maximum `Ndistinct` value is thus 5.

Also in the fourth row, we also have `Nviolation = 1` because in these cases the 5th vignette has a value less than the 4th vignette. The column `Nviolation` is calculated by the number of times any of the vignette responses are strictly contrary to the natural ordering, as given by the user's formula (ordered left to right).

In this list of vignette response rankings the careful observer might note that ties and order violations occur between two pairs of vignettes, between `vign1` and `vign2` and between `vign4` and `vign5`. The `summary()` function seeks to make it easier than staring at this list to identify troublesome patterns by providing two additional summary statistics.

Immediately above the listing of orderings is a matrix. In the upper triangle is  $p_{ij} - p_{ji}$ , such that negative number indicate a disjunction between the order of the listed vignettes and their responses. Continuing the comparison of `vign1` and `vign2`, we have the negative values for  $p_{12} - p_{21} = -0.09$ , which provides a quick summary that there is an inconsistency of the expert ordering. The proportion of ties between each pair of vignettes is shown in the lower triangle. The proportion of ties in the comparison of `vign1` and `vign2` is 0.44.

There is an issue in the political freedom data with the ordering between `vign1` and `vign2`, and with the ordering between `vign4` and `vign5`. Reasonable people might disagree (and apparently the respondents do) about which scenario indicates less freedom: Bob writes letters to newspapers about politics using a pseudonym, while Michael makes his opinion known on most issues without regard to who is listening. For some respondents the mere existence of a media outlet such as a paper to which one could write a letter discussing political subjects may be the more important indicator of freedom than the ability to talk publicly about politics. The substance of the Sonny and Vito vignettes seem to be correctly ordered, but perhaps the reversal is due to whether or not the vignette ends with the statement about men being

taken away for speaking out against the government. Further indicating that `vign4` describes a more repressive scenario, it is more often tied with the most extreme `vign6` than any other vignette.

Above this matrix is a matrix the number of times that reversals in responses occur as matrix of pairwise comparison between vignettes. Each cell summarizes the proportion of cases that the vignette listed at the beginning of the row  $i$  has a response less than the vignette listed in column  $j$ . Let  $p_{ij}$  be the value in cell  $(i, j)$ , then  $1 - p_{ij} - p_{ji}$  is the number of cases where vignette  $i$  and vignette  $j$  have the same value. For example, the proportion of cases where `vign1`<`vign2` is 0.24, while the proportion of cases where `vign2`<`vign1` is 0.32. Since  $p_{12} < p_{21}$  there appears to be an inconsistency of the expert ordering.

The summary finally provides some basic frequencies for types of responses. Most of the respondents (3267/3500) use at least two different response categories in evaluating the vignettes. The disjunction between the order of vignettes and response values are also evident from the small number of case (773/3500) that have vignette responses that take on more than one level and are (weakly) in the same natural order as the vignettes.

As an alternative, consider the reordering of vignettes such that

```
R> z2 <- anchors.order(~ vign2 + vign1 + vign3 + vign5 + vign4 + vign6,
+   freedom)
R> summary(z2, top = 10, digits = 3)
```

#### ANCHORS: SUMMARY OF VIGNETTE ORDERING

Treatment of ties: represent as sets

```
Number of cases with at least two distinct vignette responses: 3267
and with no violations of natural ordering: 1246
and with no more than 1 violation of natural ordering: 1958
and with no more than 2 violation of natural ordering: 2435
```

Proportion of cases a vignette (row) is less than another (column):

	<1	<2	<3	<4	<5	<6
1	NA	0.321	0.663	0.707	0.732	0.754
2	0.235	NA	0.651	0.715	0.757	0.775
3	0.121	0.080	NA	0.363	0.457	0.575
4	0.068	0.067	0.198	NA	0.339	0.495
5	0.080	0.056	0.138	0.183	NA	0.374
6	0.070	0.055	0.081	0.103	0.100	NA

Upper tri =  $p_{\{ij\}} - p_{\{ji\}}$  (negative values suggest misorderings)

Lower tri =  $1 - p_{\{ij\}} - p_{\{ji\}}$  (big numbers means many ties)

	1	2	3	4	5	6
1	NA	0.085	0.542	0.639	0.652	0.684
2	0.444	NA	0.571	0.649	0.702	0.720
3	0.215	0.440	NA	0.165	0.320	0.494
4	0.269	0.188	0.477	NA	0.156	0.392

```
5 0.225 0.187 0.176 0.345    NA 0.275
6 0.218 0.405 0.171 0.402 0.526    NA
```

Top 10 orderings (out of 615 unique orderings):

	Frequency	Proportion	Ndistinct	Nviolation
{1,2,3,4,5,6}	233	0.0666	1	0
{1,2},{3,4,5,6}	129	0.0369	2	0
{1,2},3,{4,5,6}	113	0.0323	3	0
{1,2},{3,4},{5,6}	98	0.0280	3	0
{1,2},{3,4,5},6	97	0.0277	3	0
1,2,{3,4,5,6}	93	0.0266	3	0
{1,2},3,{4,5},6	79	0.0226	4	0
{1,2},4,{3,5,6}	63	0.0180	3	1
1,2,3,{4,5},6	58	0.0166	5	0
1,2,{3,4},{5,6}	57	0.0163	4	0

Obviously the frequencies are the same since only the indexes of the vignettes were changed. However, the number of violations for these top sets are reduced to zero, except for {1,2},4,{3,5,6} which now has a different (but one fewer) violation as a result of flipping the order of vign4 with vign5.

The analysis of vignettes is useful both at the stage of evaluating a pilot study of survey instruments, as well at the stage of choosing how (and whether) to use particular vignettes. The results of non-parametric anchoring vignettes analysis using  $C$  are entirely dependent on which vignettes are included and the order in which they are specified.

### 2.3. Relative ranks: Calculations, vignette selection, and plotting

After the researcher has deduced or confirmed the order of vignettes, the calculation of  $C$  using the function `anchors()` uses a similar syntax but with the additional designation of the self-assessment variable on the right hand side. In this example, the self-assessment variable is aptly named `self`, and  $C$  is calculated by,

```
R> a2 <- anchors(self ~ vign2 + vign1 + vign3 + vign6, freedom,
+   method = "C", combn = TRUE)
R> summary(a2)
```

ANCHORS: SUMMARY OF RELATIVE RANK ANALYSIS:

Overview of C-ranks

Number of cases: 1013 with interval value, 2487 with scalar value

Maximum possible C-rank value: 9

Interval on C-scale: Frequency and proportions Cs to Ce

	N	Prop	MinEnt
1 to 1	339	0.097	1
2 to 2	129	0.037	2
3 to 3	32	0.009	3
4 to 4	67	0.019	4
5 to 5	393	0.112	5
6 to 6	211	0.060	6
7 to 7	433	0.124	7
8 to 8	448	0.128	8
9 to 9	435	0.124	9
1 to 4	50	0.014	1
1 to 5	33	0.009	5
1 to 6	31	0.009	5
1 to 7	28	0.008	7
1 to 8	32	0.009	7
1 to 9	19	0.005	7
2 to 4	112	0.032	2
2 to 5	74	0.021	5
2 to 6	39	0.011	5
2 to 7	47	0.013	7
2 to 8	177	0.051	7
2 to 9	91	0.026	7
3 to 6	5	0.001	5
3 to 7	1	0.000	7
3 to 8	6	0.002	7
3 to 9	2	0.001	7
4 to 6	16	0.005	5
4 to 7	3	0.001	7
4 to 8	27	0.008	7
4 to 9	12	0.003	7
5 to 8	8	0.002	7
5 to 9	6	0.002	7
6 to 8	162	0.046	7
6 to 9	32	0.009	7

Note: MinEnt is the rank for the interval that minimizes entropy

Summary of C-ranks with ties/intervals broken:

Distribution of ranks omiting interval cases

1	2	3	4	5	6	7	8	9
0.136	0.052	0.013	0.027	0.158	0.085	0.174	0.18	0.175

Distribution of ranks allocating interval cases uniformly

1	2	3	4	5	6	7	8	9
0.107	0.078	0.051	0.064	0.144	0.103	0.161	0.161	0.131

Distribution of ranks allocating interval cases via `cpolr` and conditioning on observed ranks

	1	2	3	4	5	6	7	8	9
	0.110	0.082	0.021	0.047	0.167	0.094	0.175	0.173	0.132

Allocating cases to their `MinEnt` values produces

	1	2	3	4	5	6	7	8	9
	0.111	0.069	0.009	0.019	0.169	0.060	0.310	0.128	0.124

Summary of entropy and intervals by subsets of vignettes:

	Vignettes	Estimated entropy	Minimum entropy
1	1234	2.068	1.902
5	123	1.652	1.517
4	124	1.658	1.522
3	234	1.904	1.845
2	134	1.923	1.834
12	12	1.021	0.920
10	23	1.450	1.429
9	13	1.487	1.436
8	24	1.492	1.467
7	14	1.515	1.472
6	34	1.528	1.489
15	2	0.794	0.794
14	1	0.879	0.879
13	4	0.961	0.961
11	3	1.044	1.044

	Interval Cases	Span avg.	Max. rank
1	1013	2.216	9
5	774	1.836	7
4	763	1.766	7
3	532	1.470	7
2	705	1.691	7
12	684	1.543	5
10	224	1.152	5
9	426	1.329	5
8	196	1.136	5
7	319	1.228	5
6	425	1.285	5
15	0	1.000	3
14	0	1.000	3
13	0	1.000	3
11	0	1.000	3

Note that `anchors()` function again assumes that the vignettes are entered into formula in ascending order from left to right.

The summary of the frequencies of the scalar and interval values of  $C$  can be rather extensive.

The row names **Cs** to **Ce** indicate the interval of  $C$ ; if **Cs** = **Ce** then  $C$  is a scalar. The columns **N** and **Prop** are the frequency and proportion of the cases, respectively.

In all approaches the predictions of scalar valued observations are fixed at their observed values, independent of the assumptions of each model. Predictions of vector valued observations are restricted to within their observed range, also with certainty and independent of modeling choices. Thus the more scalar values, the less the assumptions of a particular approach will affect the histogram. Similarly, the narrower the intervals of non-scalar values, the less assumptions matter as well.

The choice of methods will depend on whether one wants to assume a particular allocation method to make a particular argument (e.g., a worst case scenario) , or whether one would like to believe in a parametric model.

### *Vignette selection and diagnostics*

The function with the addition of the option `anchors(..., combn = TRUE)` looks at all combinations of subsets of vignettes and calculates the minimum entropy and (optionally) the entropy based on estimated values of a censored ordered probit model. The function also summarizes how many interval valued  $C$  are present for a particular subset of vignettes.

Most researchers will want to choose vignettes that split respondents into evenly sized groups of  $C$ . Interval values pose the same problem here as they do in defining a histogram, and the minimum entropy and censored ordered probit approaches are the focus here.

The measure of minimum entropy provides a way of comparing subsets of vignettes that takes into account the distribution of scalar and interval values of  $C$ . Ties in this case are penalized heavily, both in terms of their frequency and their width. The measure should be thought of primarily in terms of providing a ranking. Most usefully it provides a way of comparing subsets conditional on a particular number of vignettes. In general, if a researcher seeks to have vignettes that allocate respondents to equally sized bins, there is no reason not to choose the subset of vignettes (conditional on a fixed number of vignettes) with the greatest minimum entropy.

The `anchors(..., combn = TRUE)` function with also calculates the “estimated entropy”, which relies on the average conditional  $\hat{C}$  probabilities estimated by the `cpolr()` function. The estimated entropy is calculated using the model specified by `cpolr =` in the list of formulas describing the explanatory variables. If no `cpolr =` value is included in the list of formulas, a intercept-only model is assumed.

Using minimum or estimated entropy to select how many vignettes to use is an area of ongoing research, although there are some rules of thumb. Having fewer vignettes means theoretically putting more people into fewer bins of  $\hat{C}$ , but that is traded off against the increased risk of having wider interval values in  $C$ . One would not want to use a larger set of vignettes if it is strictly dominated by a smaller set in terms of minimum entropy. Estimated entropy can (and does in this example) provide estimates of  $\hat{C}$  that are more uniform as the number of ties goes up, and therefore yield higher estimated entropy; rather than leading to a penalty for ties and out of order vignettes, the greater number of intervals are estimated by `cpolr()` to be allocated in a fairly uniform manner.

In this particular data, a researcher would have to have a specific justification for using all six vignettes, instead of four or five (specifically `vign2`, `vign1`, `vign3`, `vign6` or `vign2`, `vign1`, `vign3`, `vign5`, `vign6`). Using all six vignettes creates almost twice as many ties as using only

four vignettes (`vign2`, `vign1`, `vign3`, `vign6`). The difference between Figures 1 and 2, and in particular why the latter is essentially insensitive to the method of creating the histograms, are explained primarily by the reduction in ties by the use of fewer vignettes.

Note that the best minimum entropy selected subsets may not simply drop vignettes when reducing the number of vignettes, but may actually switch vignettes as well. Consider the sequence for the freedom data of the highest minimum entropy for different  $J$ . `vign5` is dropped from  $J = 5$  to  $J = 4$ , reappears at  $J = 2$ , and is the best vignette if one were to choose just one. `vign2` is dropped from  $J = 4$  to  $J = 3$ , but reappears at  $J = 2$ .

Note also that an important feature of including `cpolr = variables` is that cases with any missing value in the covariates will be listwise deleted for both the estimated and minimum entropy calculations to ensure a common basis for comparisons. As such, the minimum entropy values may change as a function of what variables (if any) are included in `cpolr =`. The column `Interval Cases` also shows how many interval valued  $C$  are present for a particular subset of vignettes.

### *Subsetting and plotting*

Since  $C$  is defined for each case only relative to an individual's own responses, it is sometimes useful to do the analysis separately for each region.

```
R> fo <- list(self = self ~ 1, vign = cbind(vign1, vign3, vign6) ~ 1,
+   cpolr = ~ sex + age + educ)
R> a1e <- anchors(fo, freedom, method = "C", subset = country == "Eastasia")
R> a1o <- anchors(fo, freedom, method = "C", subset = country == "Oceania")
```

The `barplot()` method for `anchors.rank` objects plots the (average) fitted proportions for each value of  $C$ . As many `anchors.rank` objects can be passed to `barplot()`, and the fitted proportions from each object will be placed beside each other, in the order given. To produce Figure 2, we invoke:

```
R> par(mfrow = c(2, 2))
R> ylim <- c(0, 0.5)
R> barplot(a1e, a1o, ties = "omit", ylim = ylim, main = "Omit Tied Cases")
R> barplot(a1e, a1o, ties = "uniform", ylim = ylim,
+   main = "Uniform Allocation")
R> barplot(a1e, a1o, ties = "cpolr", ylim = ylim,
+   main = "Censored Ordered Probit Allocation")
R> barplot(a1e, a1o, ties = "minentropy", ylim = ylim,
+   main = "Minimum Entropy Allocation")
```

In practice, there may be a trade-off in selecting a subset of the available vignettes between reducing the frequency of interval valued  $C$  and the number of distinctions that are made by  $C$ . By reducing the number of cases with interval values, the assumptions of the different approach are also reduced. Figure 2 plots the same four types of  $\hat{C}$  histograms, but this time using only `vign1`, `vign3`, and `vign6` to construct  $\hat{C}$ , instead of all six vignettes. The histograms are essentially the same except for the minimum entropy figure which has a larger  $P(\hat{C} = 5)$  for Oceania respondents.

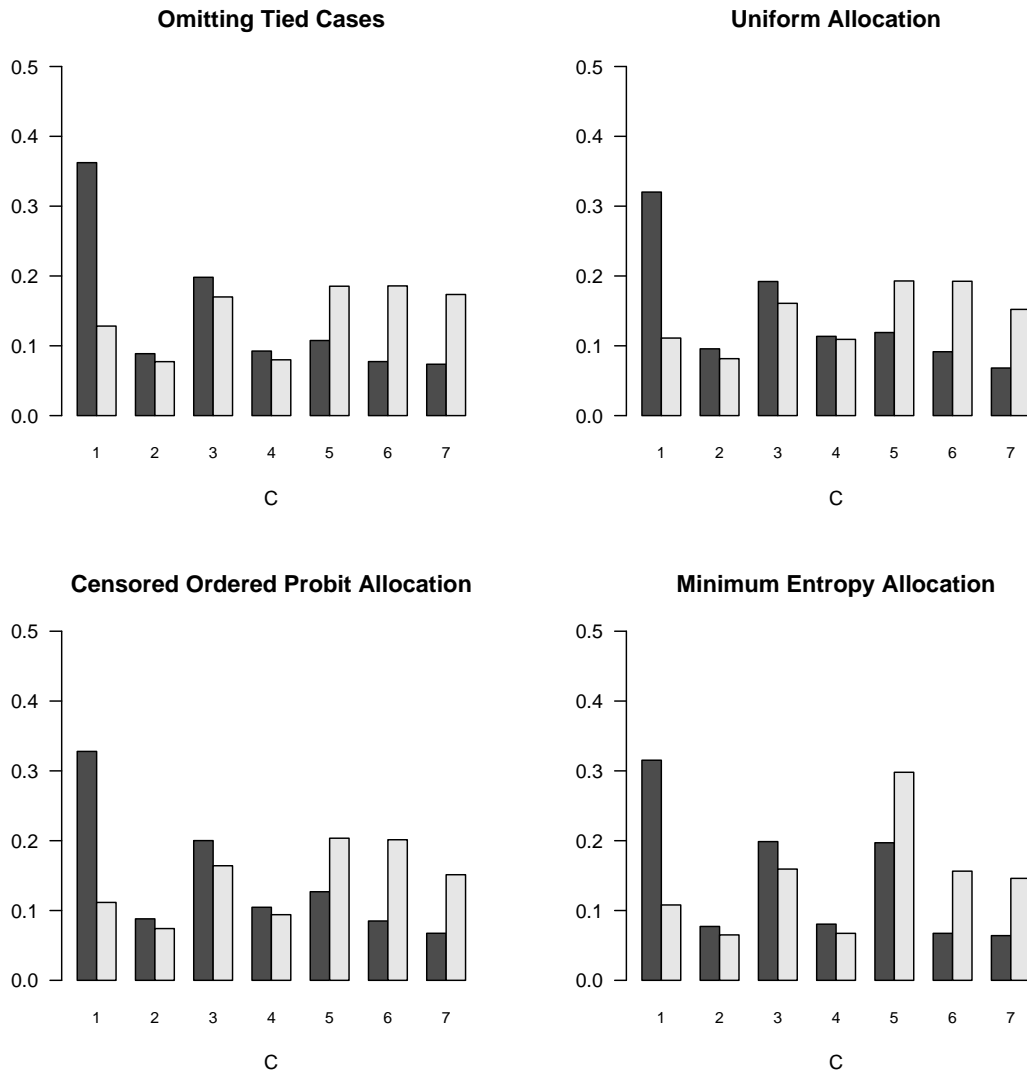


Figure 2: Approaches to summarizing the distribution of *C*: all figures are based on same political freedom model specified as,  $\text{self} \sim \text{vign1} + \text{vign3} + \text{vign6}$ . Dark histogram is for Oceania subsample, light histogram is for East Asia subsample.

## 2.4. Related functions

The functions `insert()` and `cpolr()`, along with the method `fitted()` can be used directly in combination to produce the censored ordered probit histogram values,

```
R> a1 <- anchors(fo, method = "C", data = freedom)
R> freedom2 <- insert(freedom, a1)
R> ca1 <- cpolr(cbind(Cs, Ce) ~ sex + age + educ + as.factor(country),
+ data = freedom2)
R> ca1e <- fitted(ca1, a1e, Cvec = TRUE)
R> ca1o <- fitted(ca1, a1o, Cvec = TRUE)
```



```
R> barplot(rbind(ca1e, ca1o), main = "Minimum Entropy", beside = TRUE,
+         xlab = "C")
```

The helper function `insert()` correctly combines the value of  $C$  found in `a1` with the original data set. Specifically the two columns defining the range of  $C$ , `Cs` and `Ce`, are added to the data set while taking into account that cases may have been omitted when creating  $C$ .

The function `cpolr()` fits a parametric censored ordered probit model, and a few detailed comments are warranted about this function. First, this is simply a slightly modified version of `polr()` from [Venables and Ripley \(2002\)](#), and returns objects of class `c("cpolr", "polr")`. Generic methods for `polr` class work for `cpolr` class objects with the exception of two that are preempted by two new `cpolr` methods for the generics `fitted()` and `vcov()`.

Note that  $C$  is defined for all consecutive integers from 1 to  $2J + 1$ , where  $J$  is the number of anchoring vignettes. If for a given data set, there are no observations that have a  $C$  value of any integer  $j \in \{1, \dots, 2J + 1\}$ , then  $P(\hat{C} = j) = 0$  even if there are interval values of  $C$  that include  $j$ . Thus, `cpolr()` omits the cutpoint associated with missing scalar response categories, just as `polr()` would omit cutpoints for any missing response categories in the standard ordered probit.

The method `fitted` for `cpolr` objects uses a combination of an object with class `cpolr` and an object of class `anchors`. The method `fitted()` for `cpolr` objects finds the intersection of common cases on the basis of matching row names of the two objects. In the current case, the censored ordered probit was fit using the entire sample, but fitted values are extracted for the subsets of each region by passing the `anchors` class object for the separate regions. Cases that are dropped by `cpolr()` due to missing covariates are also handled correctly when finding the intersection of the two objects.

In the function `fitted(..., Cvec = TRUE)` for `cpolr` objects this option requests the return of a  $2J + 1$  vector of the average probabilities for each of the  $\hat{C}$  categories (it simply does `apply(..., 1, mean)` on the fitted matrix of probabilities). This summary vector is common enough a interest (as it is in this example) that it is simply incorporated into the function as an option.

### 3. Parametric analysis

The following section describes the implementation of general class of parametric models for analyzing survey responses with anchoring vignettes. Considerable flexibility is provided in how to specify and identify the models so that users can easily make comparisons to different standard models of analyzing ordinal responses.

#### 3.1. Indexing notation

Our notation is a generalization of that in [King \*et al.\* \(2004\)](#), designed to accommodate our enhancements to the various models. We index survey questions, response categories, and respondents as follows:

- We index *vignette survey questions* by  $j$ . One or more vignette questions are indicated by  $j = 1, \dots, J_s$ . In contrast to the non-parametric model, it is possible to have only a self-assessment for some respondents (indicated by  $j = 0$ ).

- We index *response categories* by  $k$  ( $k = 1, \dots, K_s$ ) Each set of questions (self-assessment and vignettes) must have the same choice categories coded as increasing sequential integers starting with 1. *Missing values* (whether structural, because the question was not asked, or due to non-response) should be coded as  $k = 0$ .
- We index *respondents* by  $i$ . Respondent  $i$  ( $i = 1, \dots, n$ ) may be asked a self-assessment question or vignette questions (or both). Indeed, any combination of questions is possible (e.g., asking  $i$  all vignettes but no self evaluation; asking some vignettes and no self-evaluation, etc.) If an individual is not a question (self-evaluation or vignette) use the missing value code ( $k = 0$ ), and this question will be dropped from the likelihood function.

Thus, every mathematical symbol in the model could be indexed by  $j$ ,  $k$ , and  $i$ .

### 3.2. Chopit: A compound hierarchical ordered probit model

#### *Self-assessment component*

The top panel of Figure 3 summarizes the self-assessment component of the model.

The *actual* level for respondent  $i$  is  $\mu_i$ , a continuous unidimensional variable (with the direction defined by the order of the response categories—values indicating more freedom, better health, etc. are coded as higher or lower by the researcher). Respondent  $i$  perceives  $\mu_i$  with random normal error so that

$$Y_i^* \sim N(\mu_i, \sigma^2) \quad (10)$$

is respondent  $i$ 's unobserved *perceived* level. The actual level is a linear function of observed covariates  $X_i$ , the first column of which can be a constant term (if it is not needed for identification) and an independent normal random effect  $\eta_i$ :

$$\mu_i = X_i\beta + \eta_i \quad (11)$$

with parameter  $\beta$  and

$$\eta_i \sim N(0, \omega^2). \quad (12)$$

The *reported* survey response category is  $y_i$  and is generated by the model via this observation mechanism:

$$y_i = k \quad \text{if } \tau_i^{k-1} \leq Y_i^* < \tau_i^k \quad (13)$$

with a vector of thresholds  $\tau_i$  (where  $\tau_i^0 = -\infty$ ,  $\tau_i^K = \infty$ , and  $\tau_i^{k-1} < \tau_i^k$ , with indices for categories  $k = 1, \dots, K$ ) that vary over the observations as a function of a vector of covariates,  $V_i$  (the first column of which can be a constant term), and unknown parameter vectors  $\gamma$  (with elements the vector  $\gamma^k \geq 0$ ):

$$\begin{aligned} \tau_i^1 &= \gamma^1 V_i \\ \tau_i^k &= \tau_i^{k-1} + \gamma^k V_i \quad (k = 2, \dots, K-1) \end{aligned} \quad (14)$$

#### *Vignette component*

The bottom panel of Figure 3 summarizes the vignette component of the model. The actual level for vignette  $j$  is  $\theta_j$  ( $j = 1, \dots, J$ ), measured on the same scale as  $\mu_i$  and the  $\tau$ 's.

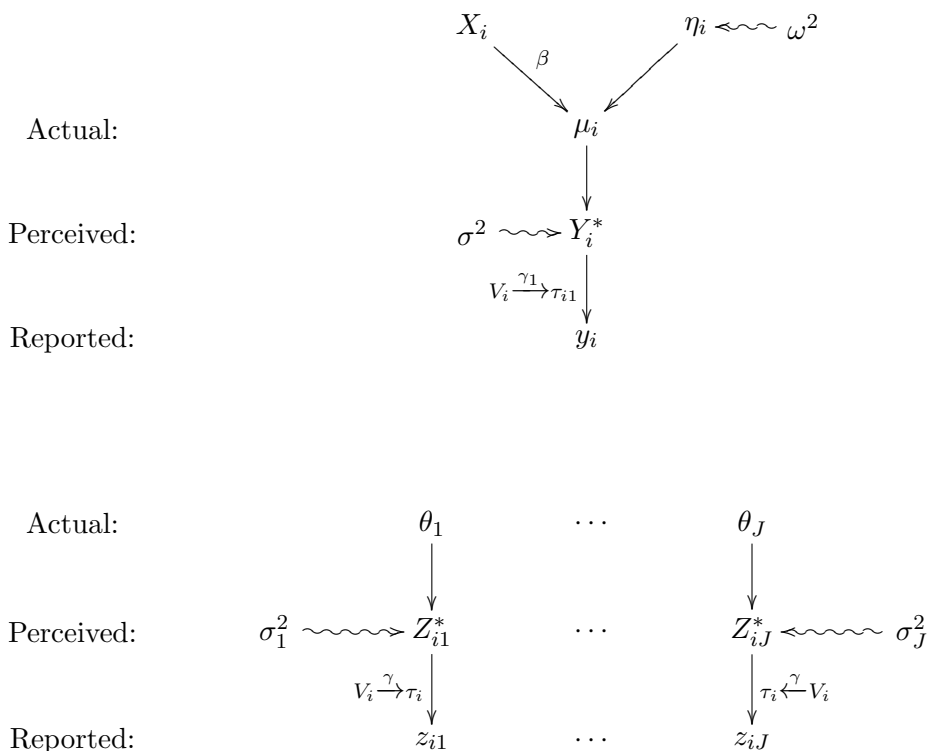


Figure 3: Self-Assessment and Vignette Component. Each solid arrow denotes a deterministic effect; a squiggly arrow denotes the addition of normal random error with variance indicated at the arrow’s source. Self-assessment levels vary over observations  $i$ . Vignette levels vary over observations  $i$ .

Respondent  $i$  perceives  $\theta_j$  with random normal error so that

$$Z_{ij}^* \sim N(\theta_j, \sigma_j^2) \tag{15}$$

represents respondent  $i$ ’s unobserved assessment of the level of vignette  $j$ .

The perception of respondent  $i$  about the level of vignette  $j$  is elicited via a survey question with the same  $K$  ordinal categories as the corresponding self-assessment question. Thus, the respondent turns the continuous  $Z_{ij}^*$  into a categorical answer to the survey question  $z_{ij}$  via this observation mechanism:

$$z_{ij} = k \quad \text{if } \tau_i^{k-1} \leq Z_{ij}^* < \tau_i^k \tag{16}$$

with thresholds determined by the same  $\gamma$  coefficients as in (14) for  $y_{i1}$ , and the same explanatory variables but with values measured for units  $i$ ,  $V_i$ :

$$\begin{aligned} \tau_{i1}^1 &= \gamma^1 V_i \\ \tau_{i1}^k &= \tau_{is}^{k-1} + \gamma^k V_i \quad (k = 2, \dots, K_1 - 1). \end{aligned} \tag{17}$$

### 3.3. Basic example

The `chopit()` function provided by **anchors** at it’s most basic simply requires specifying the

formula's defining  $ys$ ,  $zs$ , and  $\tau_s$ . For example, using variables from the `data("freedom")` data set, we have the *named* list.

```
R> fo <- list(self = self ~ sex + educ + factor(country),
+   vign = cbind(vign1, vign2, vign3, vign4, vign5, vign6) ~ 1,
+   tau = ~ sex + factor(country))
```

Note that the names `self`, `vign`, and `tau` are required.

In this example the self-assessment variable `self` is modeled to have a mean that is a linear additive function of `sex`, `age`, `educ` and `country` dummies. The vignettes are specified as a vector of outcomes `cbind(vign1, vign2, vign3, vign4, vign5, vign6)` as a function of only an intercept, `~ 1`. The  $\tau$  cutpoints shared by the self-assessment and the vignettes are specified as their own formula without a variable to the left of the tilde `~`.

Beyond the formula `fo` and the data `freedom`, we let all other arguments take their default values: i.e.,

```
R> cout <- chopit(fo, data = freedom)
```

which uses the ordinal probit normalization, which identifies/normalizes the model by omitting the intercept in  $\mu$ , and setting  $\sigma_1^2 = 1$  (the variance of the first self-assessment question). If one specified the explanatory variables of `self =` to include an intercept, then that intercept parameter would be constrained to be zero as will be `beta.(Intercept)` in this example. The default invocation also assumes that the  $\tau$  parameters are linearly additive functions of  $\gamma V$  (`use.linear = TRUE`), see equations (14) and (17) for the definitions. One can also estimate the cutpoints as an additive function of non-linear components as in King *et al.* (2004) by setting `use.linear = FALSE`.

The object `cout` contains the results of the analysis, and is of class `chopit`. The method `summary()` for `chopit` objects is useful for printing coefficients, log-likelihoods and related information. The likelihood is listed both overall, as well as for each component self-assessment and vignette.

```
R> summary(cout)
```

ANCHORS: SUMMARY OF RELATIVE CHOPIT ANALYSIS:

Model formula:

\$self

self ~ sex + educ + factor(country)

\$vign

cbind(vign1, vign2, vign3, vign4, vign5, vign6) ~ 1

\$tau

~sex + factor(country)

\$cpolr

~1

Coefficients:

	coeff	se
gamma.cut1.(Intercept)	-1.7165	0.0532
gamma.cut1.sex	0.0546	0.0227
gamma.cut1.factor(country)Eurasia	0.0405	0.0503
gamma.cut1.factor(country)Oceania	-0.1411	0.0307
gamma.cut2.(Intercept)	0.6789	0.0200
gamma.cut2.sex	-0.0422	0.0205
gamma.cut2.factor(country)Eurasia	-0.0211	0.0450
gamma.cut2.factor(country)Oceania	0.0297	0.0273
gamma.cut3.(Intercept)	0.7384	0.0178
gamma.cut3.sex	-0.0265	0.0166
gamma.cut3.factor(country)Eurasia	0.0336	0.0373
gamma.cut3.factor(country)Oceania	-0.0997	0.0216
gamma.cut4.(Intercept)	0.6691	0.0165
gamma.cut4.sex	0.0442	0.0158
gamma.cut4.factor(country)Eurasia	0.0510	0.0361
gamma.cut4.factor(country)Oceania	0.0220	0.0209
sigma.random.effect	1.0000	NaN
sigma.self	1.0000	NaN
sigma.vign1	0.8006	0.0183
sigma.vign2	1.0030	0.0239
sigma.vign3	0.7561	0.0173
sigma.vign4	0.8354	0.0208
sigma.vign5	0.7299	0.0172
sigma.vign6	1.3247	0.0414
theta.vign1	-1.0567	0.0502
theta.vign2	-1.1747	0.0520
theta.vign3	-0.2165	0.0482
theta.vign4	0.1994	0.0497
theta.vign5	-0.0240	0.0483
theta.vign6	0.9754	0.0634
beta.(Intercept)	0.0000	NaN
beta.sex	0.1352	0.0387
beta.educ	-0.0696	0.0109
beta.factor(country)Eurasia	0.4685	0.0895
beta.factor(country)Oceania	-0.7167	0.0513

-Log-likelihood of CHOPIT: 32870.53

Partition of CHOPIT -Log-likelihood by question:

	-LL	N
Self (self)	5185.373	3467
vign1	5116.043	3496
vign2	5288.353	3496
vign3	4843.890	3496
vign4	4397.889	3496

```
vign5      4562.675 3496
vign6      3476.309 3496
```

Number of cases that contribute at least partially to likelihoods:

- a) in self-responses: 3467
- b) in vign-responses: 3496

Note that parameters fixed by identification restrictions are shown but have their standard errors (`chopit.se`) listed as `NaN`. The naming convention of the parameters is as follows. First there are the gamma ( $\gamma$ ) parameters. Among the entire set of `gamma` parameters, the set of parameters associated with each cutpoint are grouped by the second appended label (`cut1`, `cut2`, etc.), and the names of the covariates appended last (e.g., `age`, `educ`). Next, there the standard deviations of the normal distributions in the model: the row `sigma.self` represents  $\sigma$ ; `sigma.vign1` represents  $\sigma_1$ , etc. Similarly, `theta.vign1` represents the mean location of the first vignette, `theta.vign2` represents the mean locating of the second vignette, etc. Finally, the `beta` rows represent the parameters associated with estimating the location of self-evaluations.

### 3.4. Identification

The parameters of the model requires identification restrictions:

1. The location on the actual (latent) scale must be chosen. This can be done by setting the intercept of the self-evaluation equations to a fixed value ( $\beta_0 = 0$  matches the convention of ordered probits) or one of the  $\theta$ 's.
2. The scale of the actual (latent) level must also be chosen. This can be done by setting the self-assessment variance to a fixed value values ( $\sigma^2 = 1$  matches the convention of ordered probits) or by setting another  $\theta$ 's.

Two common parameterizations are as follows:

1. As noted, the ordinal probit parameterization is useful for comparing chopit to this simpler model. Set  $\beta_0 = 0$  and  $\sigma^2 = 1$ . This is the default, and no action is needed by the user.

Note: the variances of the vignettes are estimated under this default normalization. To further constrain the variances of the vignette stochastic terms  $\sigma_j^2$  to also be equal to 1, use the option

```
R> chopit(fo, freedom, options = anchors.options(single.vign.var = TRUE))
```

2. Another option is parameterization defined by the extreme vignettes. Let  $\theta_1 = 0$  and  $\theta_J = 1$ . This lets estimates of  $\mu$  be interpreted on the scale of the vignettes, with 0 being the level of the lowest vignette and 1 the level of the highest. Note that  $\mu$  can still be higher than 1 or lower than 0.

To identify the model by setting  $\theta_1 = 0$  and  $\theta_J = 1$ , use the option

```
R> chopit(fo, freedom, options = anchors.options(normalize = "hilo"))
```

*Caution:* The order of the vignettes does matter for this normalization. If you constrain the  $\theta$  parameters to have an order different from what would be estimated without constraints, odd results such as extremely large standard errors and implausibly large parameter estimates can occur. *Hint:* if in doubt, use the `normalize = "self"` model first to establish the order of the vignettes.

### 3.5. Additional options

There are a variety of options, among which the following are the most often used,

- Instead of the default optimizer `optim()`, use `genoud()` (Mebane, Jr. and Sekhon 2011a,b): e.g.,

```
R> cout2 <- chopit(fo, freedom, options = anchors.options(
+   optimizer = "genoud", start = cout$parm, print.level = 1))
```

As there is as yet no proof of the global concavity of the chopit likelihood, a prudent researcher should investigate whether a choit model fitted using `optim()` is potentially at a local maximum rather than the global maximum of likelihood. Genoud does not rely on global concavity of the likelihood, and is an efficient approach to finding the global maximum.

- The option `use.gr` toggles whether or not to use the analytical gradients that have been derived for the model with a linear parameterization of cutpoints. If `use.gr = TRUE` then analytical gradients are used. The use of numerical gradients via `use.gr = FALSE`, which is currently required if  $\tau$  are specified as non-linear function, is significantly more time consuming to estimate.
- See `help("chopit")` and `demo("chopit")` for additional examples of options.

## 4. Extensions and future directions

Wand (2007a) proposed an alternative non-parametric estimator that relies on a subset of the assumptions used by *C*. Further, Wand (2007b) considered a semi-parametric estimator that uses information in the intrapersonal correlation of anchoring responses in order to identify unparameterized heterogeneity across individuals. Also, instead of assuming a common normal distribution of latent attitudes or attributes is shared by all respondents, the distribution of latent traits are estimated by an ensemble of methods, ranging from bounds of the partially identified distribution to a fully parametric fit. These new estimators will be incorporated into future releases of the **anchors** package.

## Acknowledgments

Authors' note: Our thanks to the Robert Wood Johnson Scholars in Health Policy Research program, to the NIA/NIH (grant P01 AG17625-01), to the NSF (grants SES-0112072 and IIS-9874747), and the Weatherhead Center for International Affairs for generous research

support. Thanks also to Dan Hopkins for research assistance, and to the Center for Political Studies and the Department of Health Management and Policy at the University of Michigan for their hospitality while part of this work was conducted.

## References

- King G, Murray CJL, Salomon JA, Tandon A (2004). “Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research.” *American Political Science Review*, **98**(01), 191–207.
- King G, Wand J (2007). “Comparing Incomparable Survey Responses: New Tools for Anchoring Vignettes.” *Political Analysis*, **15**, 46–66.
- Mebane, Jr WR, Sekhon JS (2011a). “Genetic Optimization Using Derivatives: The **rgenoud** Package for R.” *Journal of Statistical Software*, **42**(11), 1–26. URL <http://www.jstatsoft.org/v42/i11/>.
- Mebane, Jr WR, Sekhon JS (2011b). *rgenoud: R version of GENetic Optimization Using Derivatives*. R package version 5.7-1, URL <http://CRAN.R-project.org/package=rgenoud>.
- R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Venables WN, Ripley BD (2002). *Modern Applied Statistics with S*. 4th edition. Springer-Verlag, New York.
- Wand J (2007a). “Credible Comparisons Using Interpersonally Incomparable Data: Ranking Self-Evaluations Relative to Anchoring Vignettes or Other Common Survey Questions.” Stanford University, URL [http://wand.stanford.edu/anchors/wand\\_anchors.pdf](http://wand.stanford.edu/anchors/wand_anchors.pdf).
- Wand J (2007b). “Stochastic and Latent Class Models of Anchoring Vignettes.” Stanford University.
- Wand J, King G (2007). *Anchoring Vignettes in R: A (Different Kind of) Vignette*. Stanford University, URL <http://wand.stanford.edu/anchors/doc/anchors.pdf>.
- Wand J, King G, Lau O (2011). *anchors: Software for Anchoring Vignette Data*. R package version 3.0-6, URL <http://CRAN.R-project.org/package=anchors>.

### Affiliation:

Jonathan Wand  
 Department of Political Science  
 Stanford University



Encina Hall West  
Stanford, CA, 94305, United States of America  
*and*  
Robert Wood Johnson Health Policy Scholar  
University of Michigan  
E-mail: [wand@stanford.edu](mailto:wand@stanford.edu)  
URL: <http://wand.stanford.edu/>

Gary King  
Harvard University  
Institute for Quantitative Social Science  
1737 Cambridge Street  
Cambridge MA 02138, United States of America  
E-mail: [King@Harvard.Edu](mailto:King@Harvard.Edu)  
URL: <http://GKing.Harvard.Edu>

Olivia Lau  
Institute for Quantitative Social Science  
Harvard University  
1737 Cambridge St  
Cambridge, MA 02138, United States of America  
E-mail: [Olivia.Lau@Post.Harvard.Edu](mailto:Olivia.Lau@Post.Harvard.Edu)  
URL: <http://www.OliviaLau.Org/>