IRT IN R - VIGNETTE

ABSTRACT. umes that the outcome is a probabilistic function of a single latent trait; the distribution of this trait varies across the population and it is informative to condition this distribution on covariates. The responses of those with higher values of the trait first–order stochastically dominate the response of those with lower values of the trait.
Vignette presents a brief introduction to the `irt` package written in R language. Both R and `irt` are open-source software projects. `irt` can be downloaded from `http://code.google.com/p/rsirt/downloads/list`.
(Thanks to Mayssun, Konrad, Anna.)

1. THIS DOCUMENT

While writing this document we follow Sweave formatting of Leisch (2003) which is an implementation of the literate programming style initiated by Knuth (1992). It allows for interplay between code, output and comments...

irt() is not particularly fast or efficient and the problem is highly nonlinear; Because of the highly nonlinear nature of the problem, default values of the convergence parameters are set to very tight tolerances. The warning messages such as:

...

are harmless.

2. INTRODUCTION

There is a response r that is ordered and discrete and a latent trait or factor θ whose value determines te probability distribution of r. The possible outcomes for r will be labelled $1, 2, 3, ...$ where higher values of the outcome are 'more common' for higher values of θ. Specifically, if $\theta_2 > \theta_1$ then the responses of the θ_2 population will stochastically dominate the responses of the θ_1 population. This corresponds to a notion that more of θ leads to more of the response.

In the context at hand, there are $i = 1, ..., n$ individuals who are have characteristics W_i. These characteristics are assumed to have no effect on r except through θ. We embody this assumption by writing

$$p(r|W) = \int p(r|\theta)f(\theta|W)d\theta \tag{1}$$

We wish to estimate both $p(r|\theta)$ and $f(\theta|W)$ with as few assumptions as possible.

We start by observing that if θ were observable, we could rescale it (i.e. apply a strictly monotonic transformation to it) and, provided we adjusted $p(r|\theta)$ appropriately, $p(r|W)$ would remain unchanged, as would all the essential features of the problem. Since here θ is not directly observed, we can choose to 'scale' it for the population or a specific sub-population. So we could for example, say that the entire

Version: August 8, 2008.

population is characterized by θ being standard uniform or normal. An alternative, which we adopt here, is to designate standard characteristic W^s such that $f(\theta|W^s)$ is a 'standard' distribution—either standard normal or standard uniform as taste and convenience dictate.

3. Specification of $p(r|\theta)$

In the example to follow there are 4 responses (answers to an attitudinal question or "item") so we need to specify $F(r=3|\theta)$, $F(r=2|\theta)$, and $F(r=1|\theta)$, which correspond to the probabilities, respectively, of answering "3" or less, "2" or less, and "1". Thus the probability of answering "2" is $F(r=2|\theta) - F(r=1|\theta)$. Stochastic dominance requires e.g. $F(r=2|\theta_2) < F(r=2|\theta_1)$ if $\theta_2 > \theta_1$. Graphing the four estimated $F(r|\theta)$ curves as function of θ with θ measured on $(0,1)$ yields:

When θ is measured on (0,1) (so e.g. $U(0,1)$ for $W = W^s$), we seek a function that is *downward* sloping in θ for each $F(r|\theta)$, and which takes $(0,1)$ into $(0,1)$. There is no particular harm (in the sense of implicitly imposing a substantive restriction) if we take this function to be continuous. So a candidate for $F(r=i|\theta)$ is $1 - G_i(\theta)$, where $G_i()$ is a distribution function on $(0,1)$. (Remember: $F(r|\theta)$ is a discrete distribution function *in* r for each value of θ.) How shall $G_i(\theta)$ be constructed? The stochastic dominance conditions in θ require that the curves show in Figure 1 be downward sloping and not cross.

irt1() constructs $G_i(\theta)$ as the distribution function corresponding to an exponential tilt of the uniform density. Thus

$$G_i(\theta) = \frac{\int_0^\theta e^{1 t_1\gamma_1(u)+t_2\gamma_2(u)+\ldots t_m\gamma_m(u)}du}{\int_0^1 e^{t_1\gamma_1(u)+t_2\gamma_2(u)+\ldots t_m\gamma_m(u)}du},$$

where the functions $\gamma_1(u), \gamma_2(u), ..., \gamma_m(u)$ are m basis functions, here chosen to be the (shifted) Legendre polynomials. (Using these is equivalent to using a polynomial basis but numerically more stable.)

We now turn to the problem of enforcing stochastic dominance. If there are k (ordered) responses, $F(k|\theta) = 1$ by definition. Suppose we are given a (distribution) function $G_{k-1}(\theta)$ and we construct $F(k-1|\theta)$ as $1-G_{k-1}(\theta)$. An arbitrary $G_{k-2}(\theta)$ used in the same way—i.e. to define $F(k-2|\theta) = 1-G_{k-2}(\theta)$— will not in general lie below $F(k-1|\theta)$. But $F(k-2|\theta)$ constructed as $F(k-2|\theta) = [1-G_{k-2}(\theta)]F(k-1|\theta)$ will have the required properties. In addition, it will enforce stochastic dominance in hazard order, which is desirable; this is discussed below. So we will construct $F(k-2|\theta)....F(1|\theta)$ in this fashion.

Once this is done, the probability of the response "k" is given by $p(k|\theta) = 1 - F(k-1|\theta)$; $p(k-1|\theta) = F(k-1|\theta) - F(k-2|\theta)$; etc., with $p(1|\theta) = F(1|\theta)$.

4. Specification of $f(\theta|W)$

Whereas irt1() attempts to be semiparametric about $p(r|\theta)$, it is unabashedly parametric concerning $f(\theta|W)$. In applications, W consists largely of categorical values, so that a standard value of W, denoted W^s, is given by the characteristics which correspond to 0 for all the categorical (or "dummy") variables. When there is a continuous or pseudo–continuous variable (*age* and age^2 in the example to follow), it is natural to rescale the variable so it takes the value zero at e.g. the mean of the sample. Consequently irt1() models θ as $N(W\delta, 1)$, that is, to have a normal distribution with mean $W\delta$ and standard deviation 1. When W^s is defined so $W^s = 0$, (there is no intercept included in W) then $f(\theta|W^s)$ is $N(0,1)$.

It is possible to make this specification more flexible by allowing $f(\theta|W)$ to have a variance that varies with W; $f(\theta|W) \sim N(W\delta, \sigma^2(W))$ with $\log \sigma = W\gamma$ is one possibility we have pursued in other work. It is not implemented here.

Modeling θ on a 'normal' scale ostensibly conflicts with the construction of $p(r|\theta)$ above (embodied in Figure 1), which was done on a $(0,1)$ scale. The curves displayed in Figure 1 are indeed just curves (they are not themselves probability distributions, but a display of the relation between probability distributions), so by applying any inverse distribution function with range $(-\infty, \infty)$ we can "stretch" these curves appropriately. Applying the inverse standard normal distribution to Figure 1's curves gives:

5. The Mechanics of irt1()

`irt` is an R package to estimate the model of equation (...) by maximum likelihood; the integration is carried out by a Gauss–Legendre quadrature.

`irt` is distributed together with an extract of data from the 2002 Pew Survey of American adults (see Spady cemmap papers for some documentation.) The data of 2502 observations (call `data(elections)`) has variables which can be described in 4 categories:

- (CITEMS) 4 category responses to 7 'cultural' items. These are coded so that '4' is the most 'liberal' response.
  These are: *unborn,womentradrole,peacethrustrength searchterrorists, bandangerousbooks, firegayteachers, xratedOK*
- (EITEMS) 4 category responses to 7 'cultural' items. These are coded so that '4' is the most 'liberal' response.
  These are: *govguareatsleep, govtakecarewhocant,govhelpneedy, improveblackpos,equalOPP,poortoodep, richgetricher,govwasteful*
- (VOTE) respondents' report of how they voted in the 2000 Presidential election: Bush, Gore, Other, Did not vote. Other is primarily Nader. Unlike other Pew surveys, the 2002 survey does *not* give vote proportions in accord with the popular vote; Bush is overrepresented.
  These are: *bush,gore,other,novote*
- (Wx) demographic characteristics. 'age' is centered at about 45.8 and agesq.01 is $(yrs - 45.8)^2/100$ These are: ....

6. Getting Started

To install `irt` do ....

Once `irt` installed load it to the R session

```
> library(irt)
```

Help for `irt` can be found by writing in R

```
> help(package = "irt")
> help(irt)
```

An convenient shorthand for the latter command is to type simply `?rq`.

They do following ....

7. Basics of irt

There are four basic steps to conduct IRT estimation with `irt` package

1. upload and manipulate a data set
2. construct an item
3. run irt estimations
4. summarize results

While the first and the last step are standard in any data analysis, construction of item and estimation procedure are typical for `irt` package. We start with very simple example based on `elections` data set (describe data set) from `irt` package. First we load a package, data set and look up at variables.

```
> library(irt)
> data(elections)
> print(names(elections))
 [1] "govguareatsleep"    "govtakecarewhocant"
 [3] "govhelpneedy"       "improveblackpos"
 [5] "equalOPP"           "poortoodep"
 [7] "richgetricher"      "govwasteful"
 [9] "age"                "agesq.01"
[11] "black"              "bornagain"
[13] "blackbornagain"     "rel.catholic"
[15] "rel.nonchr"         "ed.cat1"
[17] "ed.cat3"            "ed.cat4"
[19] "ed.cat5"            "income.1"
[21] "income.3"           "income.4"
[23] "income.dk"          "parent"
[25] "hispanic"           "female"
```

Now we construct an items matrix and design matrix (we do it for consistency with rhs examples).

```
> EITEMS <- elections[, 1:8]
> W <- elections[, 9:26]
```

at this point we are ready to work with items constructors. First we consider single item model for *govguareatsleep*. We use a function `item()` to construct a single item as follows:

```
> singleItem <- item(EITEMS[, 1], W, itemnm = "govguareatsleep")
```

We input optional `itemnm` to force the name of the item. Having an item we can run irt and summarize results.

```
> res <- irt(singleItem)
> print(res)
> summary(res)
```

It is important to note that function `summary` computes jacobian and hessian matrix which may take some time.

We can also summarize the result on a plot.

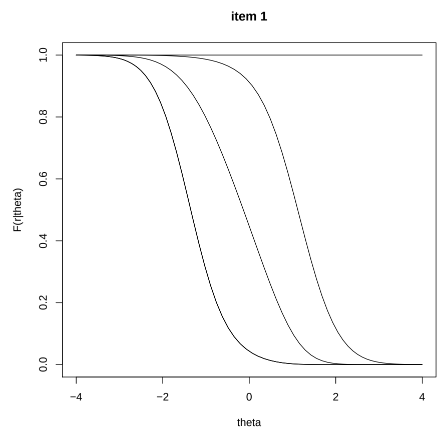


FIGURE 1. Standard irt plot.

8. SINGLE ITEM MODEL

Single item model ...

```
> library(irt)
> data(elections)
> attach(elections)
> W <- cbind(age, agesq.01, black, bornagain, blackbornagain,
+     rel.catholic, rel.nonchr)
> singleItem <- item(govguareatsleep, W)
> res <- irt(singleItem)
> detach(elections)
```

9. SINGLE ITEM MODEL - FORMULA MODE

Single item model has a formula mode for item constructor. Note that we do not need to attach data. Data set enters as an input to item constructor.

```
> data(elections)
> singleItem <- item(govguareatsleep ~ age + agesq.01 +
+     black, ~black, data = elections)
> res <- irt(singleItem)
```

10. MULTIPLE ITEMS MODEL

Multiple items model differs from single items model by items constructor. Construction of an item is done by `multipleitem()` function and the model can be run as follows

```
> library(irt)
> data(elections)
> attach(elections)
> EITEMS <- cbind(govguareatsleep, govtakecarewhocant,
+     govhelpneedy)
> W <- elections[, 9:26]
```

```
> items <- multipleitem(EITEMS, W)
> res <- irt(items, iSet = iSet)
> detach(elections)
```

This is four items model.

11. IRT optimization settings

Package `irt` contains a class `irtSet` which constitutes a basis for optimization. We can look up its default options as follows

```
> iSet <- irtSet()
> print(iSet)
```

```
 Basic settings for irt optimization:
 ========================================
        algorithm : nlm
          Fmethod : ET
            ctype : F
               nb : 2
          gradtol : 1e-11
           reltol : 1e-09
          iterlim : 250
               pl : 2
              tag : 2008-08-08-run25
            bfile : bbest.2008-08-08-run25
       orthobasis : TRUE
```

Output of irtSet() is a list of optimization options. These options determine behavior of the estimator. They can be modified on the initial call.

```
> iSet <- irtSet(iterlim = 10, pl = 0, tag = "mymModel",
+     orthobasis = FALSE)
> print(iSet)
```

```
 Basic settings for irt optimization:
 ========================================
        algorithm : nlm
          Fmethod : ET
            ctype : F
               nb : 2
          gradtol : 1e-11
           reltol : 1e-09
          iterlim : 10
               pl : 0
              tag : mymModel
            bfile : bbest.mymModel
       orthobasis : FALSE
```

Modified settings can be passed to `irt()` function through `iSet` option. Therefore the total run of a single item model would be

```
> data(elections)
> iSet <- irtSet(iterlim = 50, pl = 0, tag = "myModel")
> singleItem <- item(govguareatsleep ~ age + agesq.01 +
```

```
+     black, ~black, data = elections, iSet = iSet)
> res <- irt(singleItem)
```

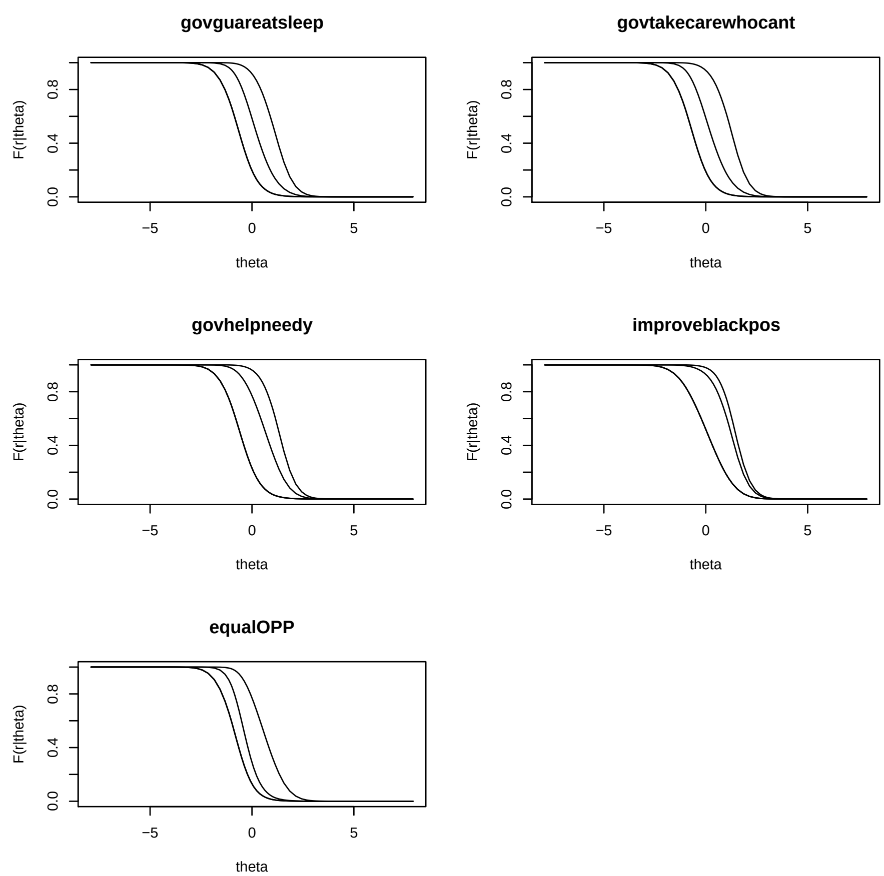
In the same fashion we do it for multiple items model.

12. Plotting Methods

Package has advanced plotting method. They are implemented in a standard R plotting framework as well as `grid` graphics model. Box-Plot model is saved in the `Ftab` in the output from `irt()`. It is of the class `Ftab`. Generic method `plot` operates in this class of objects.

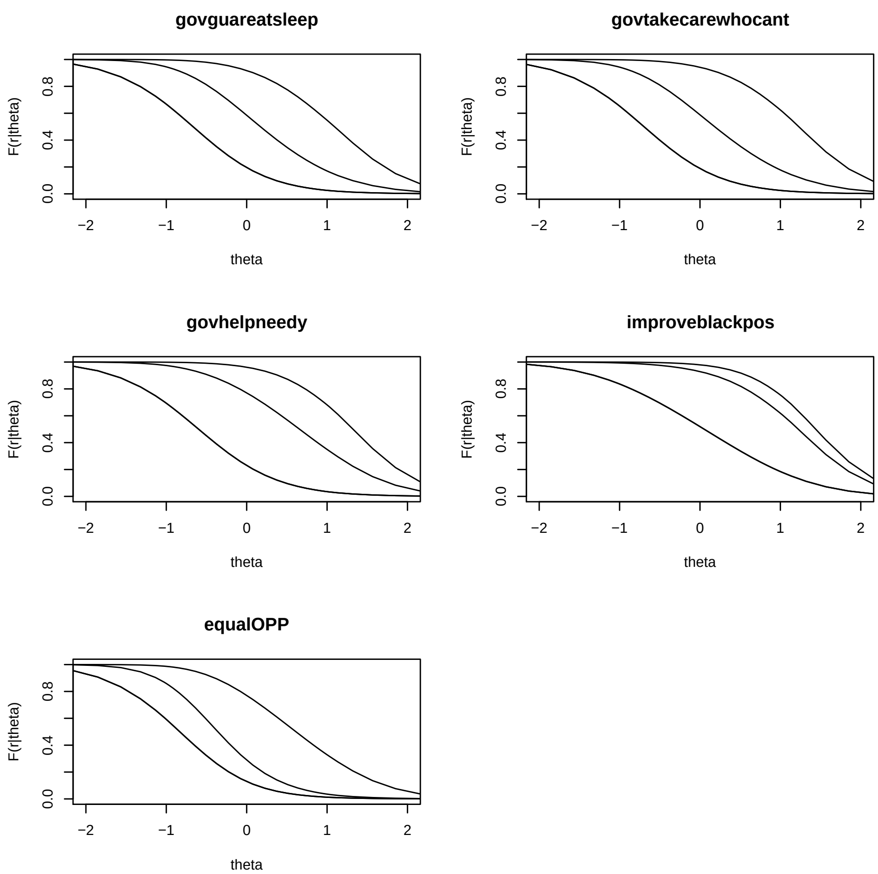
```
> Ftab <- res$Fab
> plot(Ftab)
```

This produces standard plot with default options



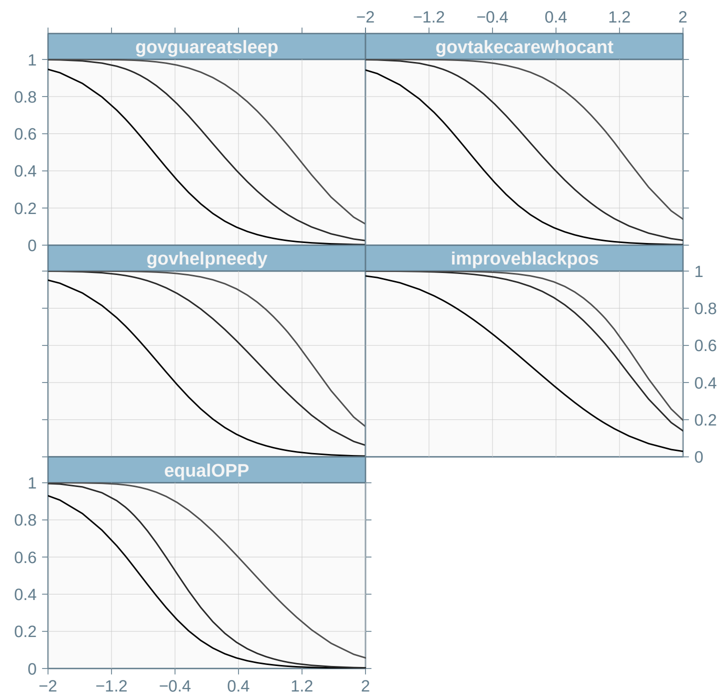
We can rescale this figure setting xscale option

```
> plot(Ftab, xscale = c(-2, 2))
```



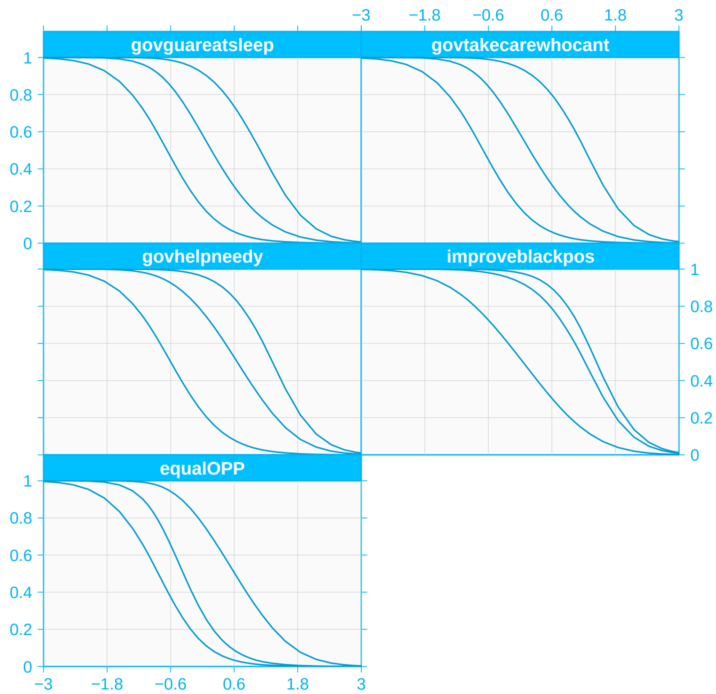
Grid plots require type option. We present the second of them

```
> plot(Ftab, xscale = c(-2, 2), type = "grid")
```



Grid graphics has also a set of color types i.e. *gray,black,red,blue* . They can be modified by option *coltype="blue"*

```
> plot(Ftab, xscale = c(-3, 3), type = "grid", coltype = "blue")
```



13. Conclusion

Conclusions ..

References

Knuth, D. E. (1992): *Literate Programming*. Center for the Study of Language and Information.
Leisch, F. (2003): "Sweave, Part II: Package Vignettes," *R News*, 3(2), 21–24.