

Kuntavaalit ja aakkosjärjestys

Seija Sirkiä

10 toukokuuta 2017

Okei okei, 1.3 miljoonaa ääntä annettiin ja aakkosjärjestyksessä ensimmäinen puolikas 33618 ehdokkaasta sai 36 000 ääntä enemmän kun loppupuolikas. Onkos se nyt sitten iso ero vai ei? Sattumaa? Mitä se tarkoittaa yhden ehdokkaan ja listan kohdalla?

Alkusäädöt:

```
library(tidyverse)

## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr

## Conflicts with tidy packages -----

## filter(): dplyr, stats
## lag():      dplyr, stats

library(stringr)

if(!file.exists("kv-2017_teat_maa.csv")) {
  if(!file.exists("kv-2017_ehd_maa.csv.zip"))
    download.file("http://tulospalvelu.vaalit.fi/KV-2017/kv-2017_ehd_maa.csv.zip")
  unzip("kv-2017_ehd_maa.csv.zip")
} # trololoo se purettu file on siis eri niminen

kv_2017_teat_maa <- read_delim("kv-2017_teat_maa.csv",
                              ";", escape_double = FALSE, col_names = FALSE,
                              locale = locale(encoding = "ISO-8859-1"),
                              trim_ws = TRUE)

## Parsed with column specification:
## cols(
##   .default = col_character(),
##   X8 = col_integer(),
##   X20 = col_integer(),
##   X29 = col_integer(),
##   X39 = col_integer(),
##   X44 = col_logical(),
##   X45 = col_double()
## )

## See spec(...) for full column specifications.

d <- kv_2017_teat_maa %>%
  filter(X4=="K") %>%
  select(X3,X10,X12,X15,X35,X19,X18) %>%
  transmute(kunta=X3,
```

```

    lista=X10,
    puolue=X12,
    kl=str_c(kunta,lista,sep="."),
    ehdokas=as.numeric(X15),
    aania=as.numeric(X35),
    nimi=str_c(X19,X18,sep=" ")) %>%
group_by(kl) %>%
arrange(ehdokas) %>% #oleellista, että saan kertymät oikein
mutate(rank=rank(ehdokas),
       namerank=rank(nimi),
       n=n(), #listan koko
       relrank=cume_dist(ehdokas),
       votesum=sum(aania), #listan summa
       voteprop=aania/votesum,
       cumvoteprop=cumsum(aania)/votesum) %>%
filter(n>1) %>% # en aio katella yhden ehdokkaan listoja
arrange(kunta,lista,ehdokas) %>% #ihan vaan omaksi iloksi
ungroup()

#vielä vähän listatietoa:
listat <- d %>%
  select(kl,rank,namerank) %>%
  group_by(kl) %>%
  summarise(rank_agree=all(rank==namerank),
            agree_prop=mean(rank==namerank))

d <- d %>% left_join(listat)

```

```
## Joining, by = "kl"
```

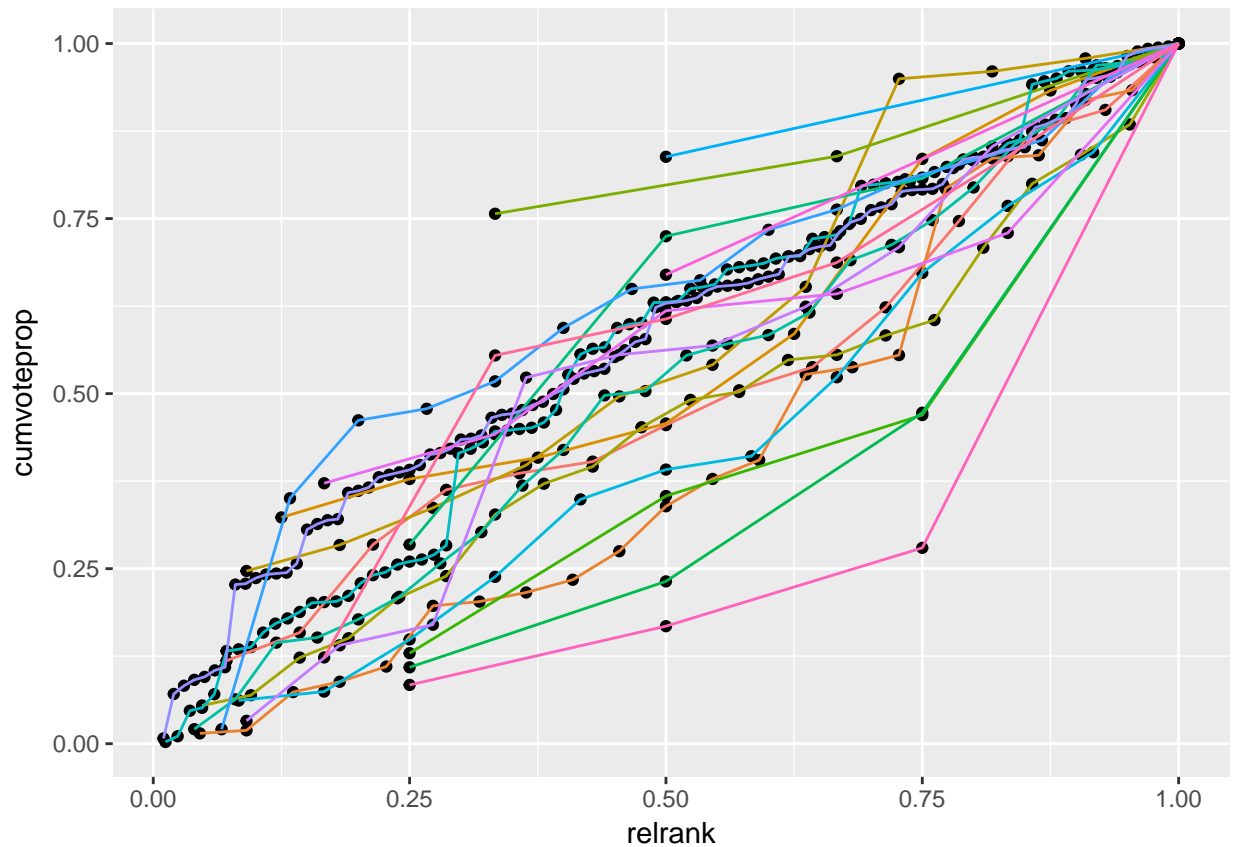
Tuossa siis eväät siihen että voi katella missä listoissa aakkosjärjestys ei päde ja miksi. Plus siitä saan kokoelman jonka avulla poimia listoja tuossa hetikohta. (Ja anteeksi puolikielisyys, vaihtui heti kun menin kirjoittamaan rank kerran.)

Ja noita hämäriä kertymäfunktioita laskin siksi että pääsisin piirtämään tuollaisia gini-käyrän näköisiä vehtaimia. Tämmösiä:

```

d %>%
  filter(kl %in% sample(listat$kl,20)) %>% #jos joku tietää paremman tavan tohon niin kertoo?
  ggplot(aes(relrank,cumvoteprop)) + geom_point()+geom_line(aes(col=kl),show.legend = FALSE)

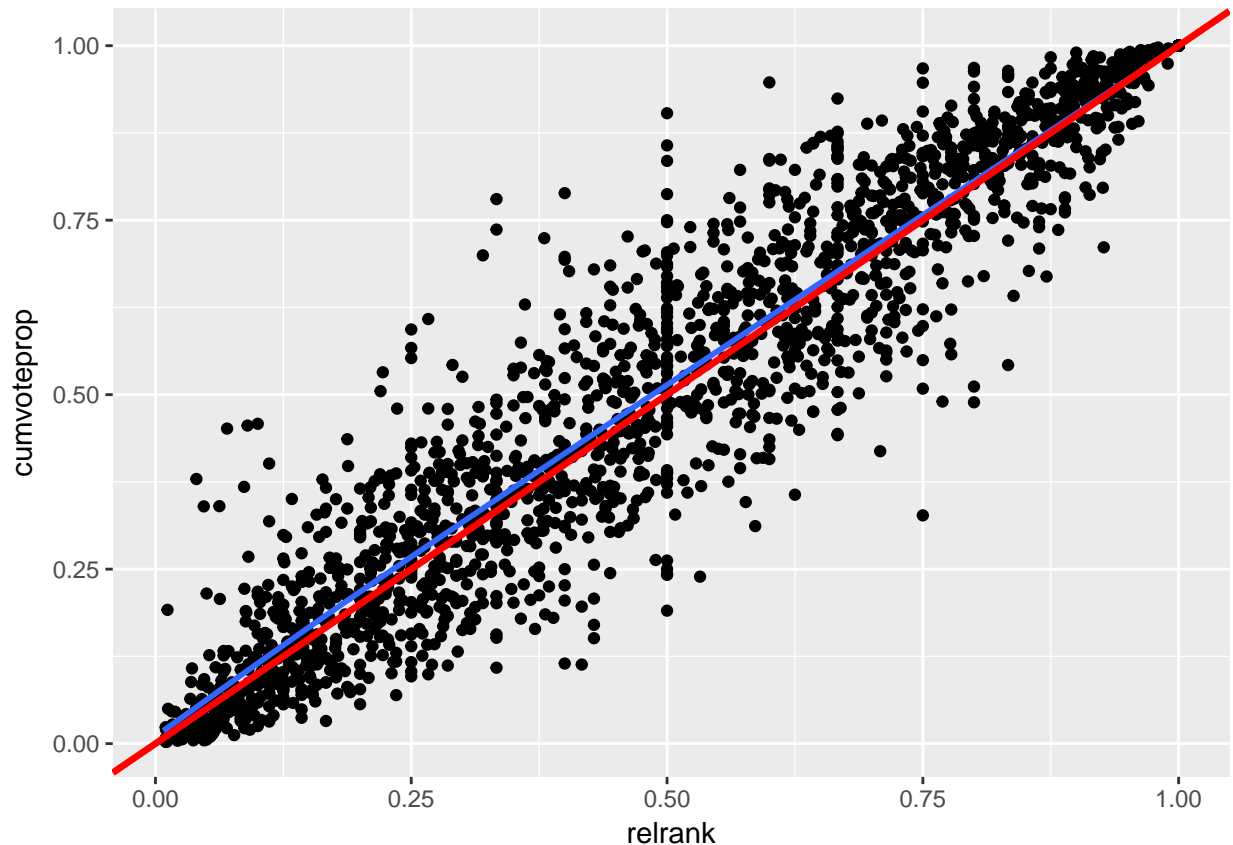
```



Jos listan järjestyksellä ei olisi väliä, jakautuminen olisi tasasta, ja noi menis kauniisti diagonaalia pitkin (plus satunnaisvaihtelu). Jos alkupää saa enemmän ääniä, noitten pitäis taipua diagonaalista ylöspäin. Ja ehkä, emmätiä? Ei noista mitään nää. Kyllä se siellä on, kun oikein tarkkaan tihrustaa koko datasta (pisteitä vaan 2000, sininen smotheri koko datasta):

```
ggplot(d,aes(relrank,cumvote)) + geom_point(data=sample_n(d,2000)) +  
  geom_smooth() + geom_abline(intercept=0,slope=1,col="red",lwd=1.2)
```

```
## `geom_smooth()` using method = 'gam'
```



Meh.

Mutta se merkitsevyys. Tähän ei oo oikein standarditestiä, kun kunnat on eri kokoisia ja listat on eri mittaisia ja se puoluekannatuskin ja kaikkee. Mutta empiiriset p-arvot onkin aina paljon kivampia. Siis.

Ehdokkaat on saaneet ne äänimäärät mitä ovat. Aakkosjärjestäminen koko maan yli on ikään kuin satunnainen järjestys, tai sen luulisi olevan. Satunnaisessa järjestyksessä tulee luonnollisestikin joku muu kuin tasatulos alku- ja loppupään välille. Mutta miten iso ero sellaisessa satunnaisessa järjestämisessä on tavallista? No, kokeillaan!

Ensinnäkin se järjestäminen. Tässä kohtaa saa valita tekeekö tämän kaikille listoille, vai vain niille joissa aakkosjärjestys on sama kuin ehdokasjärjestys.

```
if(TRUE) {
  d <- d %>% filter(rank_agree) %>%
    mutate(gnamerank=rank(nimi))
} else {
  d <- d %>% mutate(gnamerank=rank(nimi))
}
```

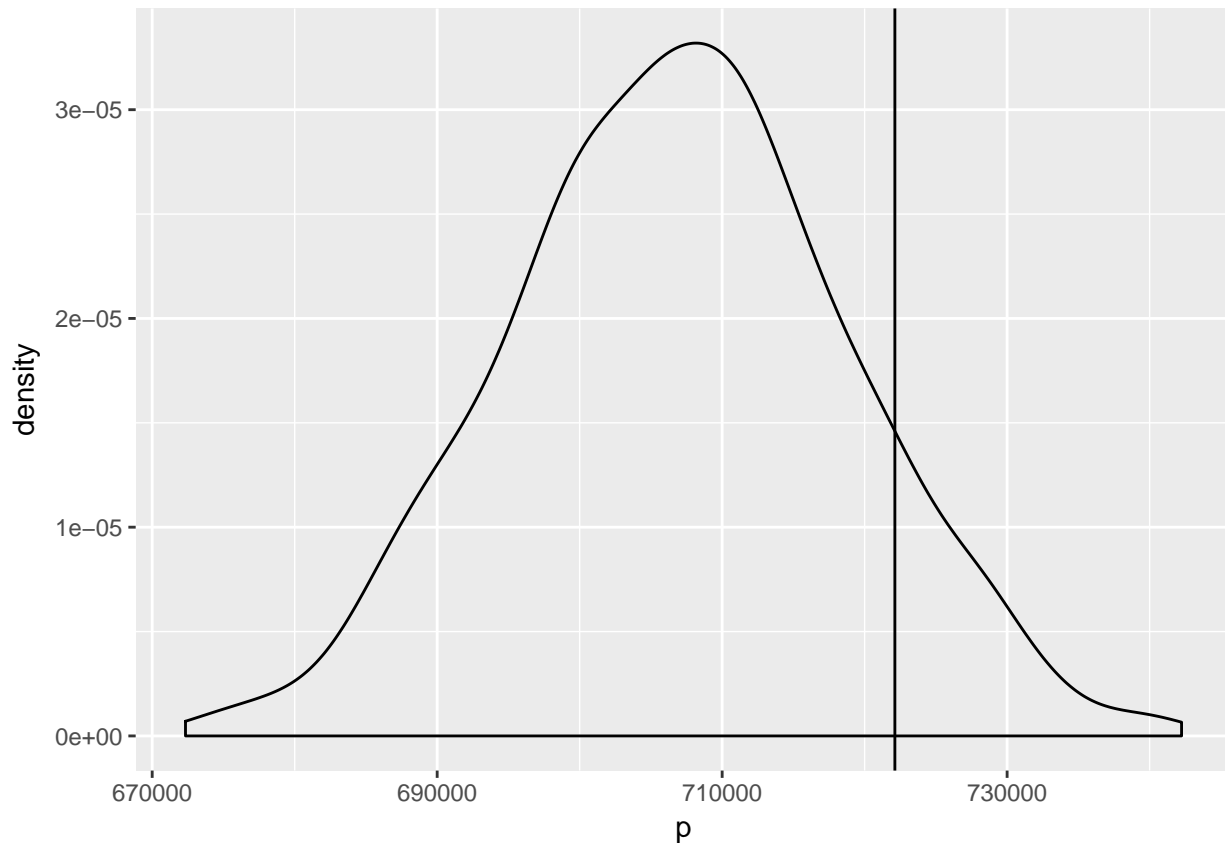
Sitten:

```
tuningp <- 0.5
Ts <- (d %>%
  mutate(smallrank=gnamerank<=tuningp*nrow(d)) %>%
  group_by(smallrank) %>%
  summarise(n=n(), vsum=sum(aania)) %>%
  filter(smallrank))$vsum
Ts
```

```
## [1] 722122
p <- rep(0,1000)
for(i in seq_along(p)) p[i] <-sum(sample_frac(d,tuningp)$aania)
summary(p)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 672300 698900 707000 706900 714700 742200

tibble(p=p) %>% ggplot(aes(p))+
  geom_density()+
  geom_vline(xintercept = Ts)
```



Whee! Ei se oo mitenkään erikoinen ero. Mutta jos samaa kokeilee ekalle viidennekselle (tuningp=0.2) niin sitten ollaan ihan jakauman reunalla.

Eli okei, kyllä se nyt vaan näyttäis niin olevan. Noh, paljonko listan eka saa hyötyä siitä että on listan eka?

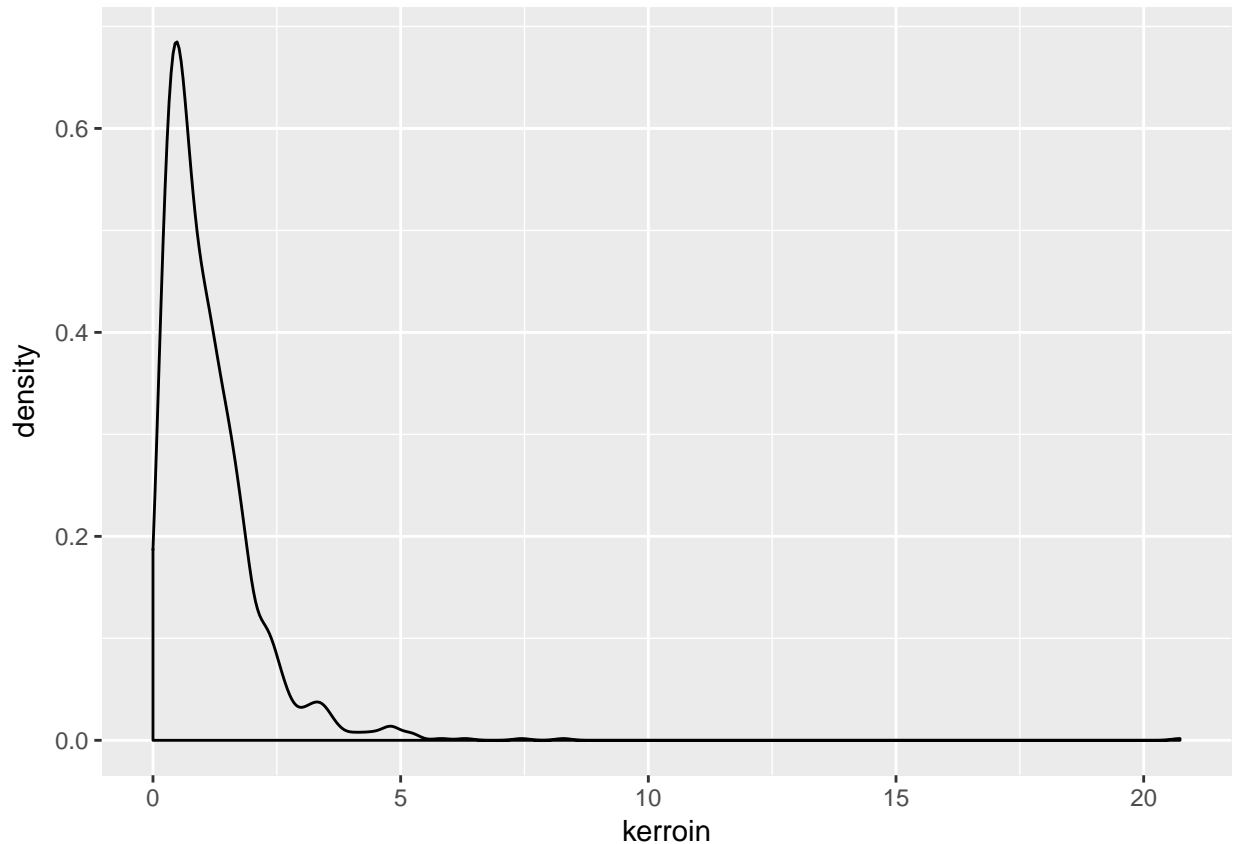
```
tmp <- d %>% filter(rank==1) %>%
  mutate(kerroin=aania/(votesum/n))
#eli monikokertainen äänimäärä oli verrattuna tasaosuuteen listalla
median(tmp$kerroin)
```

```
## [1] 0.8690347
ecdf(tmp$kerroin)(1)
```

```
## [1] 0.5658263
mean(tmp$kerroin)
```

```
## [1] 1.11047
```

```
ggplot(tmp)+geom_density(aes(kerroin))
```



Mmm... jaa että eipä juuri saa hyötyä. Tai siis jotkut saa. Vai ei? Joskushan luonnollinen ääniharava vaan on listan eka, ja toi jakauma tietty venyy tonne oikeelle kun jos lista on pitkä ja suosittu, niin ekan on mahdollista kerätä kauhea kerroin (Vapaavuori: 40x). Eli tarkoittaako tää nyt mitään? Emmätiiä!

Mutta sitä alkupäähän painottumista siis tapahtuu. Fine, mutta miten?? Ilmeisesti jotenkin sen ver-ran viehkosti että konkreettisesti ensimmäisen ehdokkaan kokema hyöty ei ole kovin kummoinen. Mutta ensimmäisten joukkoa kannattaa veikata kokonaisuudessaan, mieluummin kuin häntäpäätä.

Oma veikkaukseni selityksestä on anchoring effect: <https://en.wikipedia.org/wiki/Anchoring> Se voisi mennä niin, että äänestäjä fiksusti valitsee ensin puolueen (koska se vaalitapa on se mikä on) ja sitten käy valkkaamaan ehdokasta. Lukee nimet ja ammatit ja muut tiedot, katsoo kuvat, *aloittaen ensimmäisestä*. Anchoringin takia muuten kyseiselle äänestäjälle samanlaisista ehdokkaista miellyttävämmäksi vaihtoehdoksi jää se joka tuli vastaan ensin. Näin siis muuten “rationaalisesti” tehty valinta vienosti vääntyy alkupään suuntaan, ja siksi se efekti on niin epämääräinen, mutta selvä.