Applied Analysis 6

Rohan Hore

August 2024

1 Effect of open review policy on reviewers behavior towards a journal paper

The data set consists of data from the peer review process for papers submitted to academic journals. When a paper is submitted, the journal editor sends it to potential reviewers. Each reviewer can choose to accept or decline the invitation to review the paper. If the invitation is accepted, then they need to (1) write a review (which can be any length) and (2) choose a recommendation for what the journal should do with the paper(accept / request minor revisions / request major revisions / reject).

Typically, the reviewers are anonymous in the peer review process (i.e., the author will not see the names of the reviewers). However, the journals in this study implemented an open review policy several years ago, meaning that reviewers can choose to attach their names to the review. The goal of this paper is to examine changes in reviewer behavior that resulted from this change in policy. The data set contains data from years before and after this change was implemented. The data and analysis scripts from the paper were downloaded from nature article. The data set contains the following variables:

- `id` and `journal` are unique identifiers for the paper and for the journal it was submitted to.
- `invitation.date` and `year` indicate the date/year that the reviewer was invited to review the paper.
- `open.review` indicates whether the journal is o ering an open review option at the time of this paper
- `review.complete` indicates whether the reviewer submitted the review.
- `name.published` indicates whether the reviewer chose to publish their name.
- `recommendation` is what the reviewer recommended for the paper: Accept, Minor revisions, Major revisions, or Reject
- `accepted` indicates whether the reviewer accepted the invitation to review the paper (note: this does not mean that the reviewer recommends acceptance of the paper).
- `review.time` is the number of days between when the reviewer was invited to review, and when the review was submitted.
- `polarity` and `subjectivity` are variables computed via natural language processing. polarity takes values in $[-1, 1]$, where positive and negative values indicate positive and negative sentiments (e.g.,great or terrible). subjectivity takes values in $[0, 1]$, with larger values indicating an opinion(subjective) while smaller values indicate factual information (objective).
- `nchar` is the length of the submitted review (# of characters).
- `reviewer.status` takes values Professor, Dr, and other, recording whether the reviewer is a professor/faculty, or they have their PhD but are not a professor/faculty, or they do not have a PhD.
- `gender` is the gender of the reviewer. This information is not provided by the reviewer, but was imputed based on the name of the reviewer.

Data Reading

```r
#packages to load
library(lmerTest)
library(ordinal)
library(lubridate)
library(dplyr)
library(tidyr)

# Data upload and preparation
round1 <- read.csv("RevData.csv")
round1$id <- as.character(round1$id)
round1$journal <- factor(round1$journal, labels=c("Journal 1","Journal 2", "Journal 3",
                "Journal 4", "Journal 5"))
round1$open.review <- factor(round1$open.review, labels=c("No", "Yes"))
round1$review.complete <- factor(round1$review.complete, labels=c("No", "Yes"))
round1$name.published <- factor(round1$name.published, labels=c("No", "Yes"))
round1$recommendation <- factor(round1$recommendation, labels=c("Reject", "Major revisions",
                    "Minor revisions", "Accept"))
round1$accepted <- factor(round1$accepted, labels=c("No", "Yes"))
round1$reviewer.status <- factor(round1$reviewer.status, labels=c("Professor", "Other", "Dr."))
round1$gender <- factor(round1$gender, labels=c("Female", "Male", "Uncertain"))
```

```
id   journal invitation.date year open.review review.complete name.published
1 405 Journal 1      2010-01-01    0          No             Yes             No
2 405 Journal 1      2010-01-01    0          No             Yes             No
3 406 Journal 1      2010-01-01    0          No             Yes             No
4 406 Journal 1      2010-01-01    0          No              No             No
5 406 Journal 1      2010-01-01    0          No             Yes             No
6 407 Journal 1      2010-02-01    0          No              No             No

   recommendation accepted review.time   polarity subjectivity nchar reviewer.status
1          Reject      Yes          28 0.12838763    0.4085349  4110       Professor
2 Major revisions      Yes          16 0.08102662    0.4350710  4797           Other
3          Reject      Yes           9 0.10333333    0.4083333   687             Dr.
4            <NA>       No          NA 0.00000000    0.0000000     0             Dr.
5          Reject      Yes          39 0.13453609    0.5527891  3904             Dr.
6            <NA>       No          NA 0.00000000    0.0000000     0       Professor

     gender
1 Uncertain
2      Male
3      Male
4      Male
5      Male
6      Male
```

2 Possible Questions

Problem 1

Based on the model summary below, it was argued that

> the pure effect of the open review condition was not statistically significant. Furthermore, although several referee characteristics had an effect on the willingness of reviewing, only the interaction effect with the "other" status was significant.

How would you strengthen this statistical analysis?

Table 1 Mixed-effects logistic model on the acceptance of editors' invitation by referees

Fixed effects	Estimate	Std. error	z-value	p-value
(Intercept)	−0.193	0.214	−0.901	0.368
Open review	−0.025	0.073	−0.343	0.713
Status: Other	−0.476	0.050	−9.476	<0.001
Status: Dr	−0.135	0.030	−4.436	<0.001
Gender: Male	0.277	0.049	5.643	<0.001
Gender: Uncertain	0.338	0.055	6.164	<0.001
Year	−0.121	0.008	−14.415	<0.001
Open review × Status: Other	0.278	0.069	4.020	<0.001
Open review × Status: Dr	0.012	0.042	0.279	0.781
Open review × Gender: Male	−0.014	0.062	−0.219	0.827
Open review × Gender: Uncertain	0.005	0.070	0.074	0.941
Std. Dev. of random effects:				
Submission (intercept)	0.491			
Journal (intercept)	0.463			
No. of observations	62,790.0			
Log likelihood	−38,311.9			
AIC	76,649.8			

The reference class for the referees' status is "Professor", while for gender is "Female"

Figure 1: Model summary from paper

Answer: A better solution is to remove all the interaction terms that involve open year effect, and then compare that reduced model with the full model to make the comparison more reasonable.

Problem 2

The authors fit a gaussian linear mixed model for polarity and subjectivity with open review, the recommendation by referees, the (log of) the number of characters of the report, the year, and the gender and status of the referees (along with their interactions) respectively were included as fixed effects along with the random effects corresponding to submission and journal IDs, where

- polarity denotes the tone of the report was mainly negative or positive (varying in the $[-1, 1]$, with larger numbers indicating a more positive tone).

- subjectivity denotes whether the style used in the reports was predominantly objective (takes value in $[0,1]$, higher numbers indicating more subjective reports.

Do you have any suggestions towards improving this model?

Answer: Both are bounded responses, and might have clear peaks on the extreme boundary values. Maybe gaussian models is not a better choice here. We can try some transformation on the variables.

Problem 3

Attempt to reproduce Figure 2 of the paper. Based on visual inspection alone, comment on whether the degree of smoothing provided by the authors' Loess lines appears appropriate.
Answer:

```
Journals=c("Journal 1","Journal 2","Journal 3",
                "Journal 4","Journal 5")
acceptances_journal=invitations_journal=proportion_journal=list()
dates=head(sort(unique(round1$invitation.date)),-6)
for(i in 1:length(Journals)){
  journal_id=Journals[i]
  journal_data <- round1[round1$journal==journal_id,]
  acceptances_journal[[i]] = invitations_journal[[i]] = rep(0,length(dates))
  for(j in 1:length(dates)){
    acceptances_journal[[i]][j]<- as.numeric(table(journal_data[
      journal_data$invitation.date==dates[j],9])[2])
    invitations_journal[[i]][j]<- length(journal_data[
      journal_data$invitation.date==dates[j],9])
  }
  proportion_journal[[i]]<-acceptances_journal[[i]]/invitations_journal[[i]]
}
#plotting
dates=as.Date(dates);index=1:length(dates)
plot(dates,proportion_journal[[1]],ty="l",col=1,ylim=c(0.0,0.9),xlab="Date",
     ylab="proportion of referees accepting the invitation")
for (i in 2:length(Journals)){
  lines(dates,proportion_journal[[i]],col=i)
}
for(i in 1:length(Journals)){
  loess_model=loess(proportion_journal[[i]]~index,span=0.4)
  smoothed=predict(loess_model)
  nonNA_index=which(!is.na(proportion_journal[[i]]))
  smoothed_data=rep(NaN,length(index))
  smoothed_data[nonNA_index]=smoothed
  lines(dates,smoothed_data,col=i,lwd=2)
}
legend(x="bottomleft",legend=Journals,lty=1,col=1:5,lwd=2,cex=0.8)
```

In the figure, thicker curves denote the smoothed curves (loess smoothing as in the original paper). We see that in the later years, say $2014-2017$, we see a lot of variability in the proportion data which hasn't been captured due to higher grade of smoothing. In the original paper, smoothing to a greater extent has done injustice to the existing wiggliness in the proportion data, specially for Journal 2 and Journal 4.

Problem 4

In Table 1 of the paper, the authors used a logistic regression model with interactions to examine the effects of the open review policy on the acceptance probability of review invitations. An alternative approach is to run a logistic regression on each of the 9 subgroups separately (3 status levels $*$ 3 gender categories). For simplicity, in this question let's omit the Year variable and the random effects terms of journal and submission in both approaches.

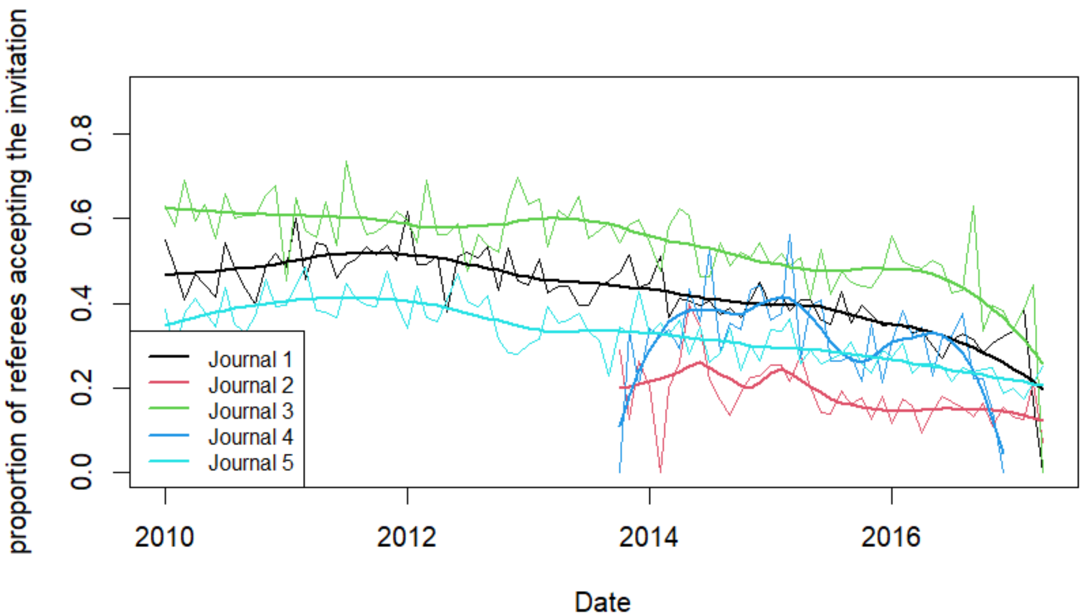

Figure 2: Plot 1

Can we find a regression model with interactions that has the same model assumptions as a set of simple logistic regression models for each of the 9 subgroups separately? If yes, will the estimates and confidence intervals of the open review effect on each subgroup be different from the two approaches? Provide an analytical justification and also check your conclusions numerically.

Answer: Let's define O_i to be the indicator variable denoting open review policy was implemented or not. Similarly, we define y_i to be the indicator random variable indicating whether the invitation to review was accepted or not. We are omitting the year variable and all the random effects of submission, journal for this problem. Thus, the set of individual logistic regression model on each of the 9 subgroups is given by

$$y_i^{s,g} \sim \text{Bin}(1, p_i^{s,g}) \quad \text{logit}(p_i^{s,g}) = \alpha^{s,g} + \alpha_o^{s,g} O_i^{s,g} \quad \text{for subgroup having status } s \text{ and gender } j$$

In total we have 18 parameters $\{\alpha^{s,g}, \alpha_o^{s,g}\}_{s,g}$ where s represents status and g represents gender, each having three levels.

Now, we consider the logistic regression model with all interactions between open effect, status and gender, where status and gender are both treated as factor variables here.

$$\begin{aligned} logit(p_i) = \beta_0 + \beta_{S1}1_{S_i=1} + \beta_{S2}1_{S_i=2} + \beta_{G1}1_{G_i=1} + \beta_{G2}1_{G_i=2} + \beta_{S1G1}1_{S_i=1,G_i=1} + \\ \beta_{S1G2}1_{S_i=1,G_i=2} + \beta_{S2G1}1_{S_i=2,G_i=1} + \beta_{S2G2}1_{S_i=2,G_i=2} \\ + (\beta_o + \beta_{OS1}1_{S_i=1} + \beta_{OS2}1_{S_i=2} + \beta_{OG1}1_{G_i=1} + \beta_{OG2}1_{G_i=2} + \beta_{OS1G1}1_{S_i=1,G_i=1} + \\ \beta_{OS1G2}1_{S_i=1,G_i=2} + \beta_{OS2G1}1_{S_i=2,G_i=1} + \beta_{OS2G2}1_{S_i=2,G_i=2})O_i \end{aligned}$$

This full interaction model also contains 18 parameters. To check whether the models are equivalent, we just need to find a one-one correspondence between the α parameters and β parameters in order to ensure that the mean under both models are same in each subgroup.

$$
\begin{aligned}
&\alpha^{0,0} = \beta_0, \quad \alpha_o^{0,0} = \beta_o \\
&\alpha^{1,0} = \beta_0 + \beta_{S1}, \quad \alpha_o^{1,0} = \beta_o + \beta_{OS1} \\
&\alpha^{2,0} = \beta_0 + \beta_{S2}, \quad \alpha_o^{2,0} = \beta_o + \beta_{OS2} \\
&\alpha^{0,1} = \beta_0 + \beta_{G1}, \quad \alpha_o^{0,1} = \beta_o + \beta_{OG1} \\
&\alpha^{0,2} = \beta_0 + \beta_{G2}, \quad \alpha_o^{0,2} = \beta_o + \beta_{OG2} \\
&\alpha^{1,1} = \beta_0 + \beta_{G1} + \beta_{S1} + \beta_{S1G1}, \quad \alpha_o^{1,1} = \beta_o + \beta_{OG1} + \beta_{OS1} + \beta_{OS1G1} \\
&\alpha^{1,2} = \beta_0 + \beta_{G2} + \beta_{S1} + \beta_{S1G2}, \quad \alpha_o^{1,2} = \beta_o + \beta_{OG2} + \beta_{OS1} + \beta_{OS1G2} \\
&\alpha^{2,1} = \beta_0 + \beta_{G1} + \beta_{S2} + \beta_{S2G1}, \quad \alpha_o^{2,1} = \beta_o + \beta_{OG1} + \beta_{OS2} + \beta_{OS2G1} \\
&\alpha^{2,2} = \beta_0 + \beta_{G2} + \beta_{S2} + \beta_{S2G2}, \quad \alpha_o^{2,2} = \beta_o + \beta_{OG2} + \beta_{OS2} + \beta_{OS2G2}
\end{aligned}
$$

This is a invertible linear system, which proves the claim.

Summary from 9 individual logistics:

```r
statuses = c("Professor", "Other", "Dr.")
genders = c("Female", "Male", "Uncertain")
coeff_matrix=matrix(ncol=4,nrow=9)
lower_intervals=matrix(ncol=4,nrow=9)
upper_intervals=matrix(ncol=4,nrow=9)
id=1
for(i in 1:length(statuses)){
  for(j in 1:length(genders)){
    subgroup_data = round1[(round1$reviewer.status==statuses[i])&(round1$gender==
                                                              genders[j]),]
    subgroup_model = glm(accepted~open.review,data=subgroup_data,family="binomial")
    coeff_matrix[id,1]=lower_intervals[id,1]=upper_intervals[id,1]=statuses[i]
    coeff_matrix[id,2]=lower_intervals[id,2]=upper_intervals[id,2]=genders[j]
    coeff_matrix[id,3:4]= coefficients(subgroup_model)
    Int=confint(subgroup_model)
    lower_intervals[id,3:4]=as.vector(Int[,1])
    upper_intervals[id,3:4]=as.vector(Int[,2])
    id=id+1
  }
}
coeff_matrix
```

```
      [,1]        [,2]        [,3]                  [,4]
 [1,] "Professor" "Female"    "-0.619789395988211"  "-0.267513799011092"
 [2,] "Professor" "Male"      "-0.0174911960113082" "-0.52277832708905"
 [3,] "Professor" "Uncertain" "0.118535683384868"   "-0.409446683379068"
 [4,] "Other"     "Female"    "-0.87762159071353"   "-0.144029656815583"
 [5,] "Other"     "Male"      "-0.966938656762998"  "-0.2992779986006"
 [6,] "Other"     "Uncertain" "-0.533846274756363"  "-0.588158893483611"
 [7,] "Dr."       "Female"    "-0.622195444587608"  "-0.531854050969155"
 [8,] "Dr."       "Male"      "-0.339700260127888"  "-0.471293265257916"
 [9,] "Dr."       "Uncertain" "-0.257544978031515"  "-0.411170283010504"
```

Summary from the big logistic model:

```r
whole_model=glm(accepted~open.review*reviewer.status*gender,family="binomial",data=round1)
coeff_wholemodel=coefficients(whole_model)
as.data.frame(coeff_wholemodel)
```

```
(Intercept) -0.619789396
open.reviewYes -0.267513799
reviewer.statusOther -0.257832195
```

```
reviewer.statusDr. -0.002406049
genderMale 0.602298200
genderUncertain 0.738325079
open.reviewYes:reviewer.statusOther 0.123484142
open.reviewYes:reviewer.statusDr. -0.264340252
open.reviewYes:genderMale -0.255264528
open.reviewYes:genderUncertain -0.141932891
reviewer.statusOther:genderMale -0.691615266
reviewer.statusDr.:genderMale -0.319803016
reviewer.statusOther:genderUncertain -0.394549763
reviewer.statusDr.:genderUncertain -0.373674613
open.reviewYes:reviewer.statusOther:genderMale 0.100016186
open.reviewYes:reviewer.statusDr.:genderMale 0.315825314
open.reviewYes:reviewer.statusOther:genderUncertain -0.302196346
open.reviewYes:reviewer.statusDr.:genderUncertain 0.262616659
```

Question: Do the coefficients match?

```
    #Subgroup:female & professor
c(coeff_wholemodel[1],coeff_wholemodel[2])
 coeff_matrix[1,3:4]
# #Subgroup:male & professor
 c(coeff_wholemodel[1]+coeff_wholemodel[5],coeff_wholemodel[2]+coeff_wholemodel[9])
 coeff_matrix[2,3:4]
# #Subgroup:female & Doctor
 c(coeff_wholemodel[1]+coeff_wholemodel[4],coeff_wholemodel[2]+coeff_wholemodel[8])
 coeff_matrix[7,3:4]
#subgroup:male & Doctor
c(coeff_wholemodel[1]+coeff_wholemodel[4]+coeff_wholemodel[5]+coeff_wholemodel[12],
   coeff_wholemodel[2]+coeff_wholemodel[9]+coeff_wholemodel[8]+coeff_wholemodel[16])
coeff_matrix[8,3:4]
```

```
     (Intercept) open.reviewYes
    -0.6197894     -0.2675138
[1] "-0.619789395988211" "-0.267513799011092"
   (Intercept) open.reviewYes
    -0.0174912     -0.5227783
[1] "-0.0174911960113082" "-0.52277832708905"
   (Intercept) open.reviewYes
    -0.6221954     -0.5318541
[1] "-0.622195444587608" "-0.531854050969155"
   (Intercept) open.reviewYes
    -0.3397003     -0.4712933
[1] "-0.339700260127888" "-0.471293265257916"
```

Note: Similarly, check for confidence intervals?

Problem 5

Answer the same questions for the cumulative-logit model in Table 2 of the paper. Here, the response is recommendation which is treated as an ordered categorical variable in the paper; here we compare the full interaction cumulative-logit model with the cumulative-logit model on each subgroup separately, while we ignore the other variables.

Answer: The outcome variable recommendation is an ordinal variable, which has four levels; we can encode them suitably as $1, 2, 3$ and 4. The cumulative logit model on individual subgroups can be written as

$$logit(P(y_i^{s,g} <= j)) = \alpha_j^{s,g} + \alpha_o^{s,g} O_i^{s,g} \quad \text{for } j = 1, 2, 3$$

Like before, consider the cumulative logit model with all interactions. We encode the status and gender variable as $0, 1$ and 2. Now, the full interaction cumulative logit is written as

$$logit(P(y_i <= j)) = \beta_j + \beta_{S1}1_{S_i=1} + \beta_{S2}1_{S_i=2} + \beta_{G1}1_{G_i=1} + \beta_{G2}1_{G_i=2} + \beta_{S1G1}\ 1_{S_i=1,G_i=1}+$$
$$\beta_{S1G2}\ 1_{S_i=1,G_i=2} + \beta_{S2G1}\ 1_{S_i=2,G_i=1} + \beta_{S2G2}\ 1_{S_i=2,G_i=2}$$
$$+ (\beta_o + \beta_{OS1}\ 1_{S_i=1} + \beta_{OS2}\ 1_{S_i=2} + \beta_{OG1}\ 1_{G_i=1} + \beta_{OG2}\ 1_{G_i=2} + \beta_{OS1G1}\ 1_{S_i=1,G_i=1}+$$
$$\beta_{OS1G2}\ 1_{S_i=1,G_i=2} + \beta_{OS2G1}\ 1_{S_i=2,G_i=1} + \beta_{OS2G2}\ 1_{S_i=2,G_i=2})O_i \quad \text{for } j = 1, 2, 3$$

These two models are not same.For the full interaction cumulative logit model, for any subgroup, we have

$$logit(P(y_i <= j|S_i, G_i, O_i)) - logit(P(y_i <= k|S_i, G_i, O_i)) = \beta_j - \beta_k,$$

while for the subgroup-specific models we get

$$logit(P(y_i <= j|S_i, G_i, O_i)) - logit(P(y_i <= k|S_i, G_i, O_i)) = \alpha_j^{s,g} - \alpha_k^{s,g}.$$

Further, the set of cumulative logit models in total have $9 * 4 = 36$ free parameters, whereas our cumulative logit model with full interactions, only have 20 parameters. So, we can't build an one-one correspondence between the two model parameters like before.

Problem 6

As the open review policy is not randomized, the open review effect is confounded with year/time. The paper adjusts for the confounding year effect by adding a linear fixed effect term of year in their regression models. Assuming that the year effect is linear can be a strong assumption. For instance, our reproduced plot for proportion of accepted papers clearly suggests that the year effect could be non-linear.

In this question, we will use only the data on 3 journals Journals 1, 3, and 5 from years $2010 - 2014$ (before the pilot study starts for Journal 3/5). We focus on estimating the policy effect on review time(days) for Journal 1. Instead of assuming a shared linear effect of year as in Table 3, we assume that the Year effect (mean review time differences across years, after controlling for all other variables) is the same for all 3 journals. Perform an analysis to estimate the average effect (averaged across the reviewers who have accepted and completed the review) of the open review policy on the review time for Journal 1 after adjusting for Year and test whether the average effect is 0 or not.

Answer: For this problem, we will only focus on Journal $1, 3$ and 5 and the data spanning from year 2010 to year 2014. Our goal is to estimate average effect of the open review policy on the review time for Journal 1 after adjusting for year.

For this problem, we only look at the records where review is complete. Towards that, we first pre-process the data accordingly, and consider the following gaussian linear mixed model

$$\text{Review\_time}_i = \beta + \beta_O O_i + \sum_{j=1}^{4} \beta_{Yj} 1_{Y_i=j} + u_{\text{Id}_i} + v_{\text{Journal}_i} + \epsilon_i$$

In above model, u is the random intercept corresponding to submission and v is random intercept corresponding to a journal. Since the linearity in year effect is a strong assumption, we treat year as a factor variable. Also, observe that in between the years 2010 and 2014, there was no instance of open review policy implemented for journal 3 and 5. Thus, the above β_O is the effect of open review policy for Journal 1.

```
id=(round1$journal %in% c("Journal 1","Journal 3","Journal 5")) &
  (round1$year %in% 0:4) & round1$review.complete=="Yes"& round1$accepted=="Yes"
data_prob5=round1[id,]
data_prob5$year=as.factor(data_prob5$year)
model <- lmer(review.time ~ open.review  + year + (1 | id) + (1 | journal), data=data_prob5)
(model_summary=summary(model))
```

```
  Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: review.time ~ open.review + year + (1 | id) + (1 | journal)
   Data: data_prob5

REML criterion at convergence: 95392.9

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.6436 -0.5214 -0.1073  0.3486  8.3523

Random effects:
 Groups   Name        Variance Std.Dev.
 id       (Intercept)  83.33    9.129
 journal  (Intercept) 172.70   13.142
 Residual             419.73   20.487
Number of obs: 10551, groups:  id, 5045; journal, 3

Fixed effects:
                Estimate Std. Error        df t value Pr(>|t|)
(Intercept)      31.1970     7.6106    2.0190   4.099  0.05381 .
open.reviewYes    6.1354     0.9668 5125.4620   6.346 2.40e-10 ***
year1             1.1350     0.8376 4524.2889   1.355  0.17544
year2             1.9784     0.8438 4287.3982   2.345  0.01909 *
year3            -2.2915     0.8533 4277.7013  -2.686  0.00727 **
year4            -3.4580     0.8610 4373.4789  -4.016 6.01e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) opn.rY year1  year2  year3
open.revwYs  0.001
year1       -0.055  0.026
year2       -0.055 -0.361  0.491
year3       -0.055 -0.387  0.482  0.632
year4       -0.054 -0.447  0.477  0.647  0.655
```

We see that the estimate of open review policy for journal 1 after adjusting for year is 6.1354.

Now, we can also see the p-value reported in the summary which suggests that the effect is significant. We can also do a LRT test for the same.

```
  reduced_model <- lmer(review.time ~ year + (1|id) + (1|journal),data= data_prob5)
redmodel_summary <- summary(reduced_model)
LRT_stat = 2*(model_summary$logLik-redmodel_summary$logLik)
pchisq(LRT_stat,1,lower.tail = FALSE)
```

```
  'log Lik.' 1.050806e-10 (df=9)
```

We see that our LRT test also suggests that the effect of open review policy for journal 1 is significant.

We might add the fixed effect of status and gender and their interaction in the model, and do the above analysis on the more general model.

```
full.model <- lmer(review.time ~ open.review  + year + reviewer.status +gender +
                   open.review:reviewer.status + open.review:gender +
                   (1 | id) + (1 | journal), data=data_prob5)
full.model.summary=summary(full.model)
reduced.model <- lmer(review.time ~ year + reviewer.status + gender +
```

```
                         (1|id) + (1|journal),data= data_prob5)
reduced.model.summary <- summary(reduced.model)
LRT_stat = 2*(full.model.summary$logLik-reduced.model.summary$logLik)
pchisq(LRT_stat,5,lower.tail = FALSE)
```

```
'log Lik.' 7.3784e-11 (df=17)
```

The p-value suggests that the effect of open.review is significant.

Problem 7

In this question, we will examine how the probability that a potential reviewer accepts the review invitation varies among papers in each journal. Call

$$p_j = \text{the probability that an invited reviewer accepts to review paper } j$$

and, we assume that it is a property of the paper (and the journal it was submitted to), but not dependent on reviewer characteristics.

You may model the reviewer acceptance/non-acceptance data for each paper j using either a binomial or negative binomial model, with success probability p_j. For each journal, assess the variation in p_j across papers. For which journal(s) is there strong evidence that p_j is not constant across papers? Which journals appear to have greatest variability in p_j?

Answer: For simplicity of analysis, we will model the acceptance/rejection data as a Binomial model. Hence, we assume n_j the number of invites sent by the journal for paper j is fixed and known. Then, if we call y_j to be the number of accepted invites for paper j, we model it as

$$y_j \sim \text{Bin}(n_j, p_j) \quad \text{for each paper } j$$

The maximum likelihood estimate $\hat{p}_j^{MLE}$ can be computed as the proportions for each paper

$$\hat{p}_j^{MLE} = \frac{y_j}{n_j}$$

```
#unique paper ids
unique_papers = unique(round1$id)
#assembling the data paper-wise
paperwise_data = as.data.frame(matrix(0,nrow=length(unique_papers),ncol=5))
for(i in 1:length(unique_papers)){
  paperwise_data[i,1]=unique_papers[i]
  paperwise_data[i,2]=as.character(unique(round1[round1$id==
                                                   unique_papers[i],]$journal))
  tab = table(round1[round1$id==unique_papers[i],]$accepted)
  paperwise_data[i,3]=as.vector(tab)[2]
  paperwise_data[i,4]=sum(as.vector(tab))
  paperwise_data[i,5]=paperwise_data[i,3]/paperwise_data[i,4]
}
colnames(paperwise_data)=c("Id","Journal","Accepted","Invitations","Proportions")
head(paperwise_data)

#proportions of acceptance for each journal
journal1_probs=paperwise_data[paperwise_data[,2]=="Journal 1",5]
journal2_probs=paperwise_data[paperwise_data[,2]=="Journal 2",5]
journal3_probs=paperwise_data[paperwise_data[,2]=="Journal 3",5]
journal4_probs=paperwise_data[paperwise_data[,2]=="Journal 4",5]
journal5_probs=paperwise_data[paperwise_data[,2]=="Journal 5",5]

#histogram of the proportions of acceptance
```

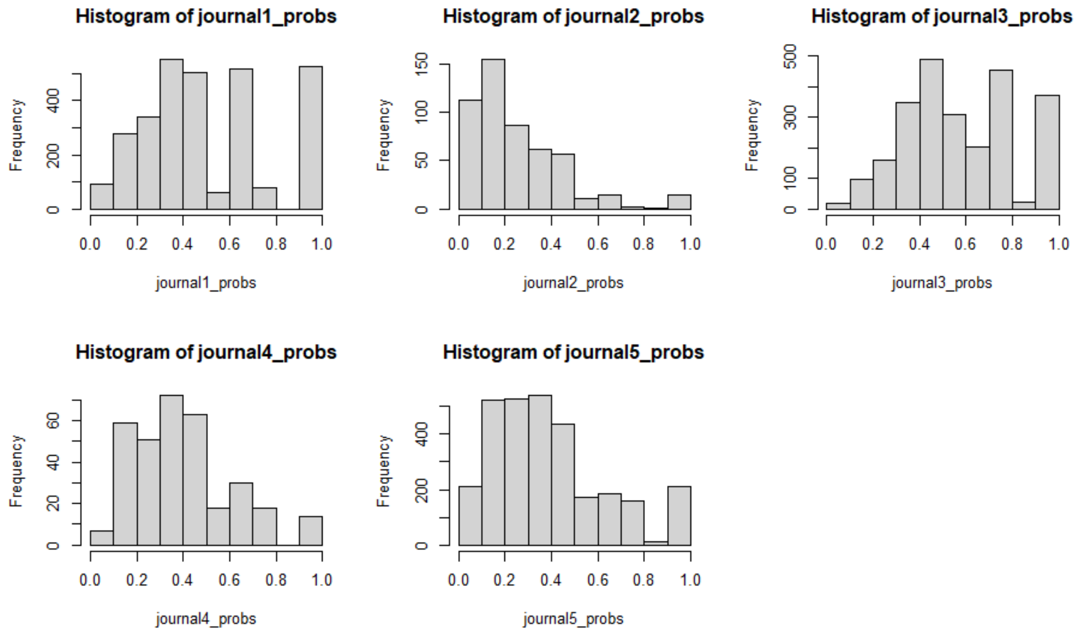

Figure 3: histograms for MLE estimates

```r
par(mfrow=c(2,3))
hist(journal1_probs)
hist(journal2_probs)
hist(journal3_probs)
hist(journal4_probs)
hist(journal5_probs)

#standard deviations of proportion estimates
sd(journal1_probs)
sd(journal2_probs)
sd(journal3_probs)
sd(journal4_probs)
sd(journal5_probs)
```

```
[1] 0.2798388, 0.210145, 0.2391466, 0.2223551, 0.2520773
```

A simple histogram plot of those can give a picture about the variations in p_j across papers.

Few observations: We see that there is a clear spike at 1 in the histogram of proportion estimates for journal 1,3 and 5. For journal 2,4 and 5, the histogram is right skewed, having most of the mass at the low proportions. Journal 1 and 5 seems to have the greatest variability in p_j. This is because there is a spike at 1 for both of these journals and rest of the mass is distributed mostly at the low proportions.

Problem 8

Using an Empirical Bayes approach, or otherwise, obtain an approximate posterior mean and 90% credible interval for each p_j. Compare the posterior mean estimates with the maximum likelihood estimates.

Answer: Now, using empirical Bayes approach, we will obtain an approximate posterior mean and 90% credible interval for each p_j. The above histogram shows remarkable variability across journals and also intuitively, the probability of acceptance should depend on the journal,the paper has been submitted to.

$$y_j \sim Bin(n_j, p_j) \quad p_j \sim Beta(a_i, b_i) \quad \text{if paper } j \text{ was submitted to journal } i$$

This means marginally,

$$y_j \sim BetaBinom(n_j, a_i, b_i) \quad \text{if paper } j \text{ was submitted to journal } i$$

Thus, the estimates $\hat{a}_i$ and $\hat{b}_i$ for each journal i can be found by maximizing this marginal Beta-Binomial likelihood. Since the data of different journals are independent, we can analyze each journal separately. Once, we have the estimates of a_i and b_i for every i, we can compute the posterior means. Observe that if paper j was submitted to journal i,then the posterior distribution of $p_j|y_j, \hat{a}_i, \hat{b}_i$ can be written as

$$p_j|y_j, \hat{a}_i, \hat{b}_i \sim Beta(\hat{a}_i + y_j, \hat{b}_i + n_j - y_j)$$

Thus, the posterior estimates of p_j can be written as

$$\hat{p}_j^{eb} = \frac{\hat{a}_i + y_j}{\hat{a}_i + \hat{b}_i + n_j}.$$

```
    library(extraDistr)
Journals=c("Journal 1","Journal 2","Journal 3","Journal 4","Journal 5")
estimates=list();confidence_interval=list()
prior_means=rep(0,length(Journals))
par(mfrow=c(1,2))
for(i in 1:length(Journals)){
  journal_id = Journals[i]
  acceptance_journalwise=paperwise_data[paperwise_data[,2]==journal_id,]
  accepted_number = acceptance_journalwise[,3];invited_number=
    acceptance_journalwise[,4]
  marginal_loglikelihood=function(par){
    a=par[1];b=par[2]
    sum=0
    for(i in 1:length(accepted_number)){
      sum=sum-dbbinom(accepted_number[i],size=invited_number[i],alpha=a,beta=b,
                      log=TRUE)
    }
    return(sum)
  }
  optimized=optim(par=c(1,2),fn=marginal_loglikelihood)
  a_hat=optimized$par[1];b_hat=optimized$par[2]
  print(paste("Journal:",journal_id,"a_hat:",a_hat,"b_hat:",b_hat))
  prior_means[i]=a_hat/(a_hat+b_hat)
  par1=a_hat+accepted_number
  par2=b_hat+invited_number-accepted_number
  upper=qbeta(0.95,par1,par2);lower=qbeta(0.05,par1,par2)
  confidence_interval[[i]]=as.data.frame(cbind(upper,lower))
  colnames(confidence_interval[[i]])=c("upper","Lower")
  mle_est = accepted_number/invited_number
  eb_est = (accepted_number+a_hat)/(invited_number+a_hat+b_hat)
  estimates[[i]]=as.data.frame(cbind(mle_est,eb_est))
  colnames(estimates[[i]])=c("MlE Estimate","EB Estimate")
  #if(i==5){par(mfrow=c(1,2))}
  hist(mle_est,main=paste("histogram for",journal_id),xlim=c(0,1))
  hist(eb_est,main=paste("histogram for",journal_id),xlim=c(0,1))
}

prior_means
```

Estimated prior means:

```
0.4339464 0.1983784 0.5523608 0.3452261 0.3424187
```

As expected, the EB method has shrink-ed the MLE estimates towards the estimated prior mean as expected. In order to compute the 90% credible intervals for each p_j, we will be using the quantiles of our estimated posterior distribution $p_j|y_j, \hat{a}_i, \hat{b}_i$.

```r
#posterior estimates: Journal 1
head(estimates[[1]][2])
#confidence Intervals: Journal 1
head(confidence_interval[[1]])
```

```
  EB Estimate
1 0.4883817
2 0.4659761
3 0.4268140
4 0.4883817
5 0.4455361
6 0.3937242

      upper     lower
1 0.6655661 0.3124992
2 0.6400315 0.2955543
3 0.5940596 0.2667128
4 0.6655661 0.3124992
5 0.6162459 0.2803818
6 0.5539407 0.2430607
```

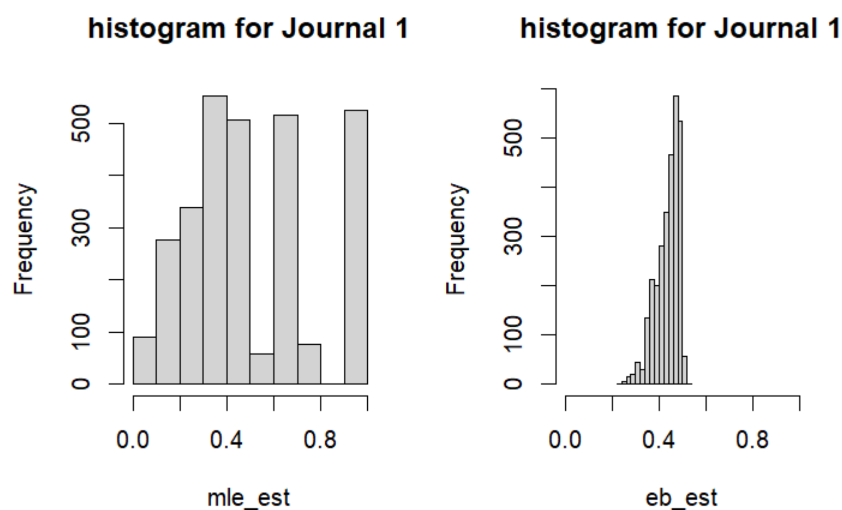

Figure 4: Compare MLE and EB estimates for Journal 1

Note: What about the peaks near 1 for most journals? Can we do better?

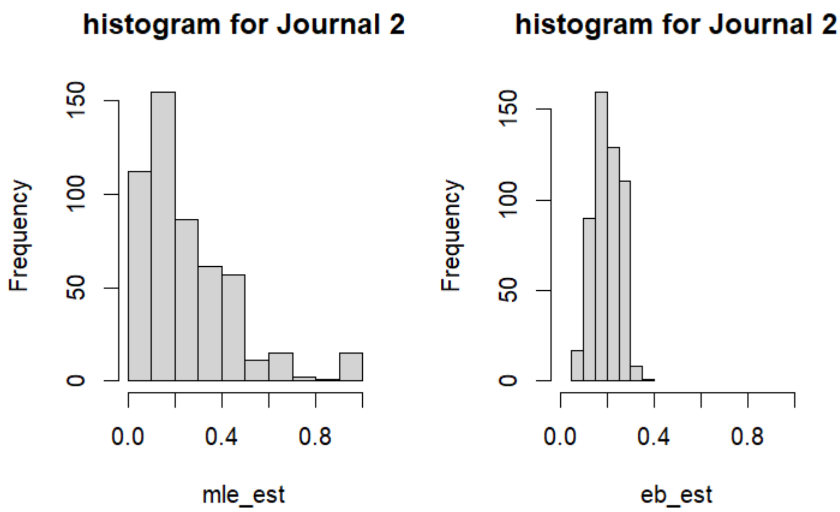

Figure 5: Compare MLE and EB estimates for Journal 2

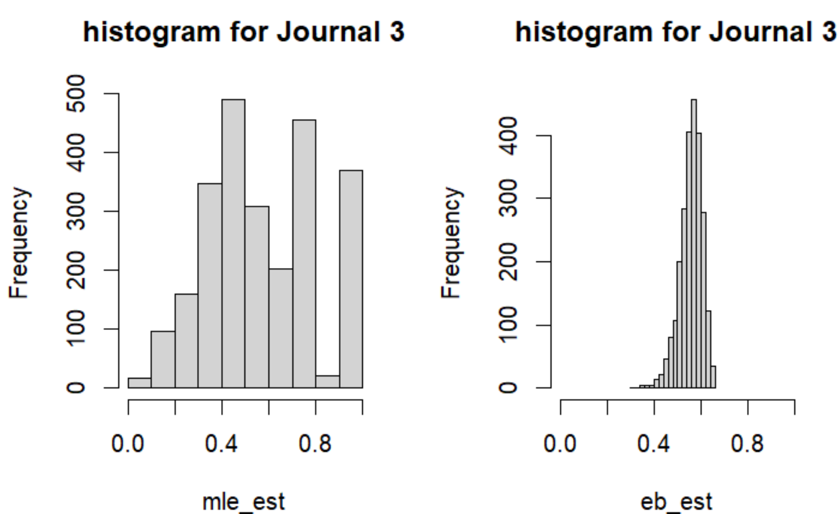

Figure 6: Compare MLE and EB estimates for Journal 3

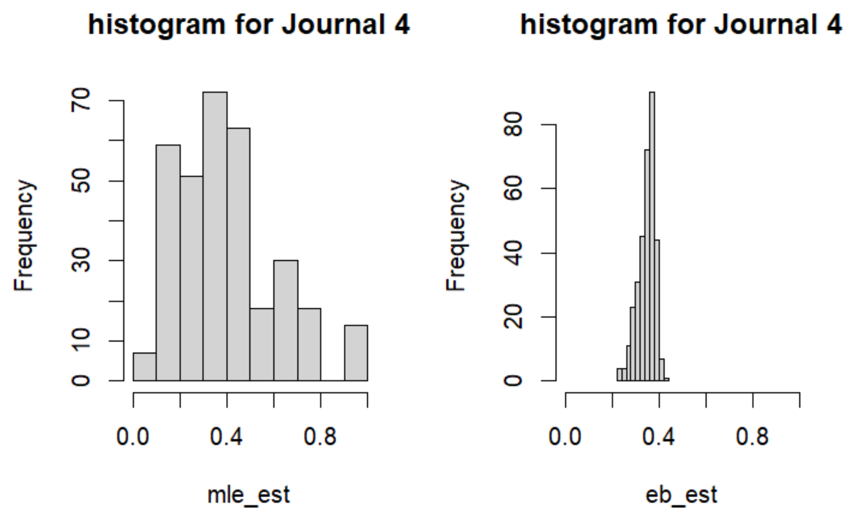

Figure 7: Compare MLE and EB estimates for Journal 4

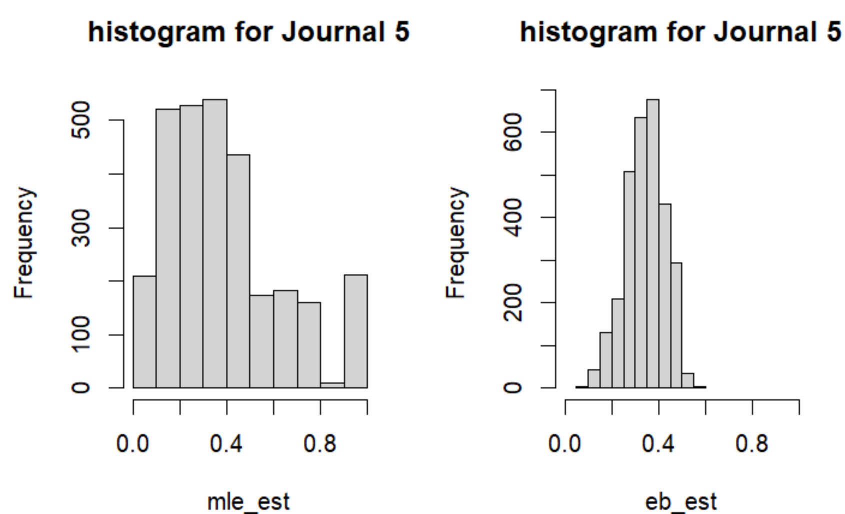

Figure 8: Compare MLE and EB estimates for Journal 5