Power simulation for linear mixed-effects models

Nora Wickelmaier

February 6, 2023

Power analysis by simulation

Why simulation?

- Simulation is at the heart of statistical inference
- Inference: Compare the data with the output of a statistical model
- If data look different from model output, reject model (or its assumptions)
- Simulation forces us to **specify a data model** and to attach meaning to its components
- Model should not be totally unrealistic for those aspects of the world we want to learn about

Specify the model including the effect of interest

(1) Choose statistical model according to its assumptions
- Binomial test $\rightarrow$ binomial distribution $\rightarrow$ `rbinom()`
- t test $\rightarrow$ normal distribution $\rightarrow$ `rnorm()`
- ...

(2) Fix unknown quantities
- Standard deviations, correlations, ...

(3) Specify the effect of interest
- Smallest effect size of interest (SESOI)

Estimating power with simulation

Pseudo code

```
Set sample size
replicate
{
  Draw sample from model with minimal relevant effect
  Test null hypothesis
}
Determine proportion of significant results
```

Sample size calculation

- Adjust n until desired power (0.8 or 0.95) is reached
- To be on the safe side, assume higher variation, less (or more) correlation, and smaller interesting effects (what results can we expect, if . . .)

Generalized linear models

- A generalized linear model is defined by

$$g(E(y)) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k,$$

  where $g()$ is the link function that links the mean to the linear predictor. The response y is assumed to be independent and to follow a distribution from the exponential family
- In R, a GLM is fitted by

```
glm(y ~ x1 + x2 + ... + xk, family(link), data)
```

Binomial regression

- Logit or probit models are special cases of GLMs for binomial response variables
- Artificial example: congenital eye disease

Logit model

$$\log \frac{p}{1-p} = \beta_0 + \beta_1 AGE$$

Probit model

$$\Phi^{-1}(p) = \beta_0 + \beta_1 AGE$$

Logit model / logistic regression

- We want to model the probability of y with the logistic function

$$p(y) = \frac{1}{1+e^{-y}} \quad \text{with } y \sim Binom(n, p)$$

- How do we get the logit model $\log \frac{p}{1-p} = \beta_0 + \beta_1 x$ from that?

$$\begin{aligned}
\log\left(\frac{p}{1-p}\right) = \log(p) - \log(1-p) &= \log\left(\frac{1}{1+e^{-y}}\right) - \log\left(1 - \frac{1}{1+e^{-y}}\right) \\
&= \log(1) - \log(1+e^{-y}) - \log(e^{-y}) + \log(1+e^{-y}) \\
&= -\log(e^{-y}) \\
&= y := \beta_0 + \beta_1 x
\end{aligned}$$

with $1 - \frac{1}{1+e^{-y}} = \frac{e^{-y}}{1+e^{-y}}$

Refresher: Logarithm rules

Product: $\log(xy) = \log x + \log y$

Quotient: $\log \frac{x}{y} = \log x - \log y$

Power: $\log \left(x^p\right) = p \log x$

Root: $\log \sqrt[p]{x} = \frac{\log x}{p}$

Meta-information and Memory: Nicole's experiment

Task:

Participants read 40 sentences that can be either true (30) or false (10) which are presented in different font colors

Study 1:

Recognition Memory: "Did you read this statement?"

DV – correct answer: correct recall

IVs – condition (control vs. high vs. low discriminability)

– truthvalue (true vs. false)

Study 2:

Source Memory: "Was this statement true, false, or new?"

DV – correct answer: true sentence classified as true

IVs – condition (high vs. low discriminability)

– truthvalue (true vs. false)

Base

The friends sit around a coffee table.

Low

The friends sit around a coffee table.

High

The friends sit around a coffee table.

Meta-information and Memory: Nicole's experiment

Hypotheses

Study 1 Recognition Memory:
Significant interaction effect between truthvalue and discriminability: Recognition memory will be higher for true than false statements when a discriminability task is present

Study 2 Source Memory:
Significant interaction effect between truthvalue and discriminability: Participants in high discriminability condition will more often classify statements correctly as true or false than in the low discriminability condition

Meta-information and Memory: Nicole's experiment

Hypotheses

Study 1 Recognition Memory:
Significant interaction effect between truthvalue and discriminability: Recognition memory will be higher for true than false statements when a discriminability task is present

Study 2 Source Memory:
Significant interaction effect between truthvalue and discriminability: Participants in high discriminability condition will more often classify statements correctly as true or false than in the low discriminability condition

→ We have the data of the first study and want to calculate the **number of participants** we need for the second study

Meta-information and Memory: Nicole's experiment

Model

- In order to test the interaction, we fit the following model:

$$log\left(\frac{P(correct)}{P(false)}\right) = \beta_0 + \beta_1 truthval + \beta_2 cond + \beta_3(truthval \times cond) + \upsilon_0 + \eta_0$$

  with $\upsilon_0 \sim N(0, \sigma_\upsilon)$ and $\eta_0 \sim N(0, \sigma_\eta)$, all random effects i.i.d.

- This model can be fitted in R with

```
lme4::glmer(correctAnswer ~ truthvalue * condition +
            (1|participantID) + (1|itemID),
            data, family = binomial)
```

Meta-information and Memory: Nicole's experiment

Model

- In order to test the interaction, we fit the following model:

$$log\left(\frac{P(correct)}{P(false)}\right) = \beta_0 + \beta_1 truthval + \beta_2 cond + \beta_3(truthval \times cond) + \upsilon_0 + \eta_0$$

  with $\upsilon_0 \sim N(0, \sigma_\upsilon)$ and $\eta_0 \sim N(0, \sigma_\eta)$, all random effects i.i.d.

- This model can be fitted in R with

```
lme4::glmer(correctAnswer ~ truthvalue * condition +
            (1|participantID) + (1|itemID),
            data, family = binomial)
```

→ How many parameters does this model have?

Estimates from Study 1

Random effects	Variance	Std.Dev.
itemID	0.29	0.54
participantID	0.58	0.76

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.62	0.15	11.15	0.00
truthvaluetrue	-0.02	0.12	-0.16	0.87
conditionhigh	-0.50	0.19	-2.59	0.01
conditionlow	-0.64	0.19	-3.35	0.00
truthvaluetrue:conditionhigh	0.39	0.16	2.44	0.01
truthvaluetrue:conditionlow	0.40	0.16	2.48	0.01

Estimates from Study 1

Random effects	Variance	Std.Dev.
itemID	0.29	0.54
participantID	0.58	0.76

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.62	0.15	11.15	0.00
truthvaluetrue	-0.02	0.12	-0.16	0.87
conditionhigh	-0.50	0.19	-2.59	0.01
conditionlow	-0.64	0.19	-3.35	0.00
truthvaluetrue:conditionhigh	0.39	0.16	2.44	0.01
truthvaluetrue:conditionlow	0.40	0.16	2.48	0.01

→ Parameters are on the logit scale

Model predictions

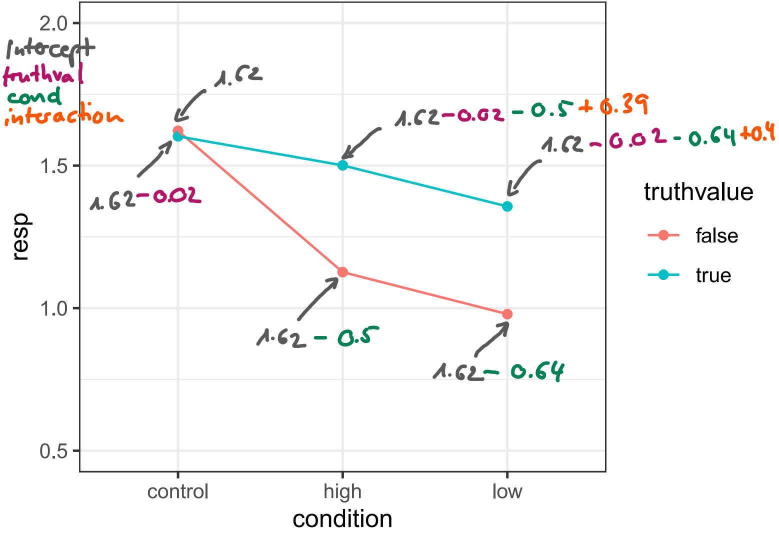
truthvalue	condition	prob	logit
false	control	0.84	1.62
true	control	0.83	1.60
false	high	0.76	1.13
true	high	0.82	1.50
false	low	0.73	0.98
true	low	0.80	1.36

Predictions on the logit scale

	est
(Intercept)	1.62
truthvaltrue	-0.02
condhigh	-0.50
condlow	-0.64
truthvaltrue:condhigh	0.39
truthvaltrue:condlow	0.40

Get predictions:

```
predict(glm1,
        re.form = NA)
```

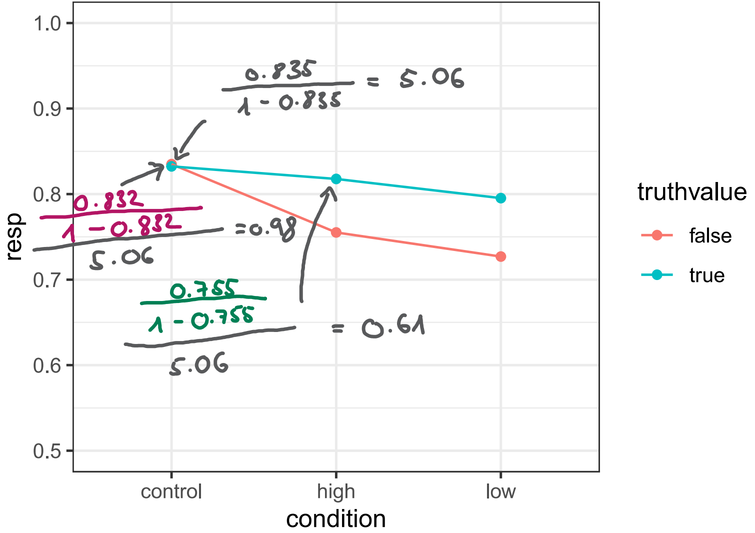


Predictions on the probability scale

	OR
(Intercept)	5.06
truthvaltrue	0.98
condhigh	0.61
condlow	0.53
truthvaltrue:condhigh	1.48
truthvaltrue:condlow	1.49

Get predictions:

```
predict(glm1,
        re.form = NA,
        type = "resp")
```

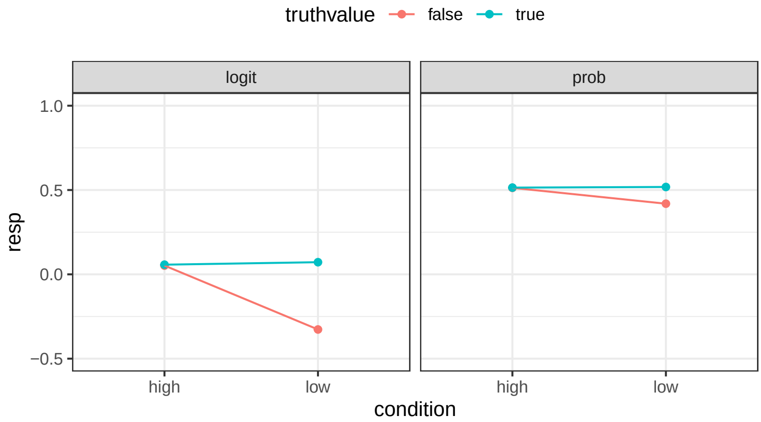


Data simulation for Study 2

```
n <- 200    # try different values

# Create data frame
dat <- data.frame(
    id          = factor(rep(1:n, each = 40)),
    item        = factor(paste(rep(1:5, each = 40), 1:40, sep = ":")),
    condition   = factor(rep(c("high", "low"), each = nitem)),
    truthvalue = factor(rep(c("true", "false"), c(30, 10)))
)

# Do some checks
xtabs( ~ id + item, dat)
xtabs( ~ condition + truthvalue, dat)
xtabs( ~ condition + truthvalue + id, dat)
```

Effect size of interest

How do I find suitable numbers for the model parameters?
Let us simulate some data with minimal error and plot the results.



	"true"
(Intercept)	0.00
truthvaltrue	0.05
condlow	-0.40
truthvaltrue:condlow	0.40

Data simulation for Study 2

```r
ran <- c("id.(Intercept)"   = 0.6,
         "item.(Intercept)" = 0.8)
fix <- c("(Intercept)"                      = 0,
         "truthvaluetrue"                   = 0.05,
         "conditionlow"                     = -0.4,
         "truthvaluetrue:conditionlow" = 0.4)

# Simulate data
sim <- simulate( ~ truthvalue * condition + (1|id) + (1|item),
                newdata   = dat,
                newparams = list(beta = fix, theta = ran),
                family    = binomial,
                nsim      = 20) # should be at least 400!
```

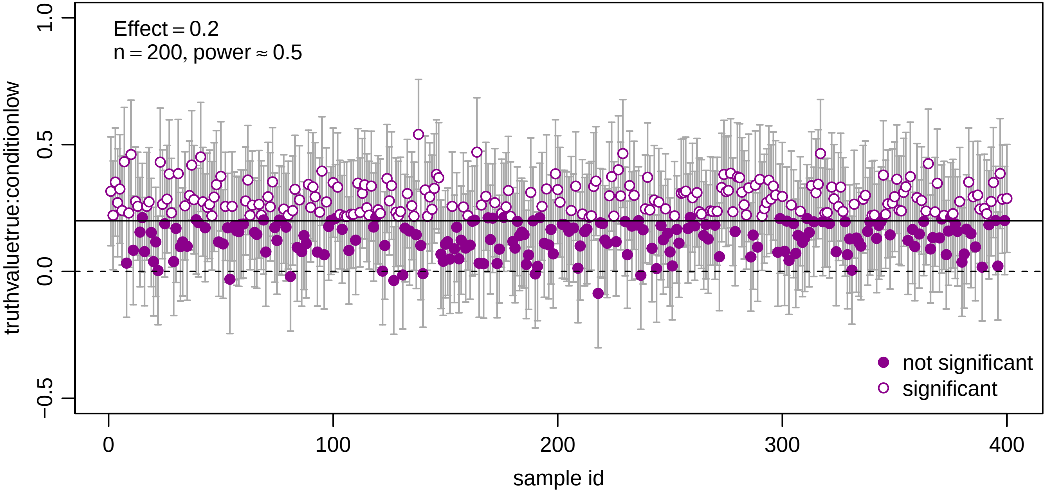
Power simulation

```
pval <- numeric(ncol(sim))

for (i in seq_len(ncol(sim))) {
  m1 <- glmer(sim[, i] ~ truthvalue * condition +
                (1|id) + (1|item), dat, family = binomial)
  pval[i] <- summary(m1)$coef["truthvaluetrue:conditionlow", "Pr(>|z|)"]
}

# Power
mean(pval < 0.05)
hist(pval)
```

Parameter recovery

```
par_rev <- replicate(400, {
  sim <- simulate( ~ truthvalue*condition + (1|id) + (1|item),
                  newdata = dat,
                  newparams = list(beta = c(0, 0.01, -0.2, 0.2),
                                   theta = c(0.5, 0.5)),
                  family = binomial)[,1]
  glmer(sim ~ truthvalue * condition + (1|id) + (1|item),
        data = dat, family = binomial)
  })
mean(sapply(par_rev, fixef)[1,]); mean(sapply(par_rev, fixef)[2,])
mean(sapply(par_rev, fixef)[3,]); mean(sapply(par_rev, fixef)[4,])

mean(sqrt(sapply(par_rev, function(x) unlist(VarCorr(x)))[1,]))
mean(sqrt(sapply(par_rev, function(x) unlist(VarCorr(x)))[2,]))
```

Estimation of inflated effect



Estimation of inflated effect

```
# Parameter estimates and confidence intervals
int <- sapply(par_rev,
  function(x) fixef(x)["truthvaluetrue:conditionlow"])
ci <- sapply(par_rev,
  function(x) confint(x, method = "Wald")["truthvaluetrue:conditionlow",])
# p values
p <- sapply(par_rev,
  function(x) summary(x)$coef["truthvaluetrue:conditionlow", "Pr(>|z|)"])

# Power
mean(p < 0.05)
hist(p)
# Inflation of effect
summary(int[p < 0.05])
```

Results of Study 2

	est	se	z	p
(Intercept)	-0.53	0.08	-6.66	0.00
truthvaltrue	1.34	0.07	18.72	0.00
condlow	-0.21	0.11	-1.90	0.06
truthvaltrue:condlow	0.08	0.10	0.74	0.46

References

Wickelmaier, F. (2022).Simulating the power of statistical tests: A collection of R examples. *ArXiv*. https://arxiv.org/abs/2110.09836