Power Priors for Replication Studies

Samuel Pawel*, **Frederik Aust**†, **Leonhard Held***, **Eric-Jan Wagenmakers**†
* Department of Biostatistics, University of Zurich
† Department of Psychological Methods, University of Amsterdam
E-mail: samuel.pawel@uzh.ch

May 25, 2022

This is a preprint which has not yet been peer reviewed.

Abstract

Power priors are used for incorporating historical data in Bayesian analyses by taking the likelihood of the historical data raised to the power α as the prior distribution for the model parameters. The power parameter α is typically unknown and assigned a prior distribution, most commonly a beta distribution. Here, we give a novel theoretical result on the resulting marginal posterior distribution of α in case of the the normal and binomial model. Counterintuitively, when the current data perfectly mirror the historical data and the sample sizes from both data sets become arbitrarily large, the marginal posterior of α does not converge to a point mass at $\alpha = 1$. This implies that a complete pooling of historical and current data is impossible if a power prior with beta prior for α is used. The result explains the often empirically observed phenomenon that the marginal posterior of α hardly changes from its prior when there is no conflict between historical and current data.

Keywords:

1 Introduction

Power priors are a class of prior distributions which can be used for incorporation of historical data in Bayesian analysis of current data (Chen and Ibrahim, 2000). The basic idea is to use the likelihood of the historical data raised to the power of $\alpha \in [0, 1]$ as the prior distribution for the model parameters θ. This leads to the posterior distribution of θ being informed by both the current and the historical data, typically resulting in an information gain compared to an analysis of the current data in isolation. The power parameter α determines how much the historical data are discounted, thus representing a quantitative compromise between the extreme positions of completely trusting ($\alpha = 1$) and completely ignoring them ($\alpha = 0$).

Usually, the power parameter α is unknown and assigned a prior distribution. In this case, the marginal posterior density of the model parameters θ based on the current data D and the historical data D_0 is given by

$$f(\theta \mid D, D_0) = \int_0^1 f(\theta \mid D, D_0, \alpha) f(\alpha \mid D, D_0) \mathrm{d}\alpha,$$

that is, the posterior of θ based on a fixed power parameter α averaged over the marginal posterior density of α. The marginal posterior distribution of α thus determines how much pooling between the two data sets occurs, and intuition suggests that it should become increasingly concentrated at $\alpha = 1$ when there is no conflict between the data sets, and their sample size increases. Here, we show that this is not the case, at least not in the typical situation when α is assigned a beta prior distribution and the data have normal or binomial likelihood. Instead, there is is a limiting posterior distribution which is hardly different from the prior.

Table 1: Historical and current data on the comparison between Fidaxcomicin and Vancomicion with respect to their effect on *Clostridium difficile*-associated diarrhoea in adults. Data taken from the meta-analysis of Nelson et al. (2017).

Study	Risk (Fidaxomicin)	Risk (Vancomicin)	Risk ratio (95% CI)
Cornely et al. (2012)	$\frac{193}{270} = 71.5\%$	$\frac{163}{265} = 61.5\%$	1.16 (1.04, 1.31)
Louie et al. (2011)	$\frac{214}{302} = 70.9\%$	$\frac{198}{327} = 60.6\%$	1.17 (1.04, 1.31)

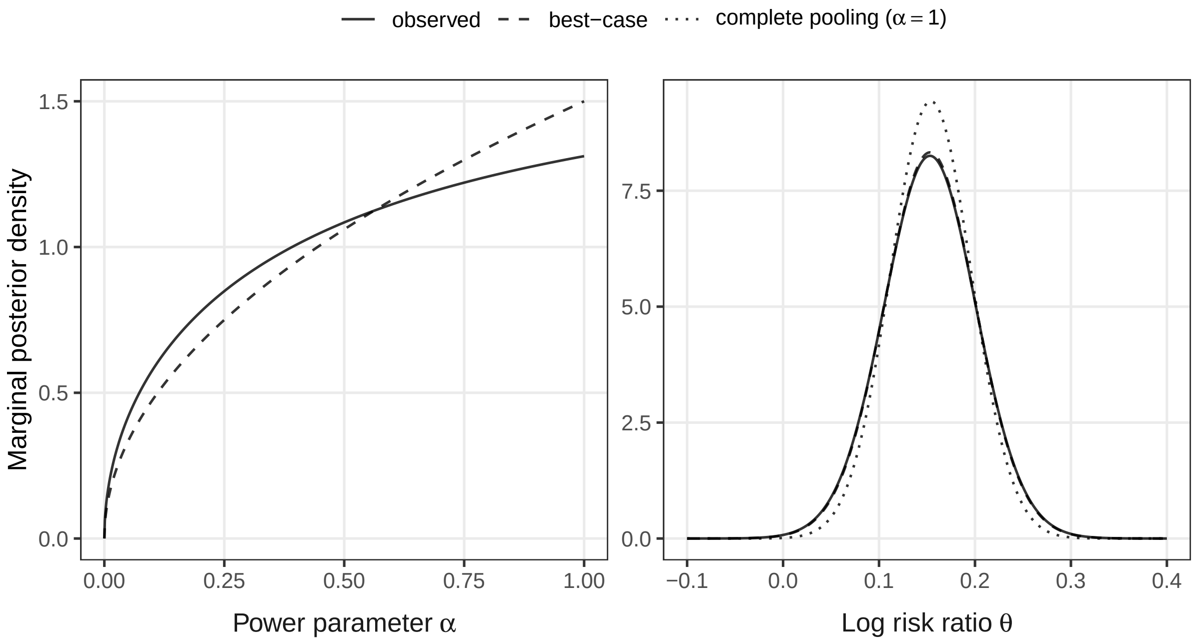

Figure 1: Power prior modeling of current data $D = \{\hat{\theta} = 0.15, \sigma = 0.06\}$ from Cornely et al. (2012) and historical data $D_0 = \{\hat{\theta}_0 = 0.16, \sigma_0 = 0.06\}$ from Louie et al. (2011). A $\alpha \sim \text{Be}(1, 1)$ prior is used for the power parameter. Marginal posterior densities are computed by numerical integration.

2 Parameter estimates under normality

Let $D = \{\hat{\theta}, \sigma^2\}$ and $D_0 = \{\hat{\theta}_0, \sigma_0^2\}$ denote estimates of an unknown parameter θ and their (assumed to be known) standard errors obtained from current and historical data, respectively. For both estimates, assume a normal likelihood centered around the parameter θ and with variance equal to the squared standard error. The default "normalized" version of the power prior (Duan et al., 2005; Neuenschwander et al., 2009) is obtained via updating of an initial prior for the parameter θ by the likelihood of the historical data raised to the power α. The prior for the power parameter α is then marginally assigned. Here and henceforth we will use an initial uniform prior for the parameter $f(\theta) \propto 1$ and a beta prior for the power parameter $\alpha \sim \mathrm{Be}(p, q)$. This choice leads to the normalized power prior

$$f(\theta, \alpha \mid D_0) = \mathrm{N}(\theta \mid \hat{\theta}_0, \sigma_0^2/\alpha)\, \mathrm{Be}(\alpha \mid p, q) \tag{1}$$

with $\mathrm{N}(\cdot \mid m, v)$ the normal density function and $\mathrm{Be}(\cdot \mid p, q)$ the beta density function. Combining (1) with the likelihood of the current data produces a joint posterior for parameter θ and power parameter α, from which a marginal posterior for α can be obtained by integrating out θ

$$f(\alpha \mid D, D_0) = \frac{\mathrm{N}(\hat{\theta} \mid \hat{\theta}_0, \sigma^2 + \sigma_0^2/\alpha)\, \mathrm{Be}(\alpha \mid p, q)}{\int_0^1 \mathrm{N}(\hat{\theta} \mid \hat{\theta}_0, \sigma^2 + \sigma_0^2/\alpha')\, \mathrm{Be}(\alpha' \mid p, q) \mathrm{d}\alpha'}. \tag{2}$$

The integral in the denominator of (2) is in general not available in closed form, except in certain important situations which we will discuss in the following.

The first situation occurs when the current data perfectly mirror the historical data in the sense that both parameter estimates are equivalent ($\hat{\theta} = \hat{\theta}_0$). In this case, several terms cancel in (2) so that the integral be represented in terms of the hypergeometric function ${}_2F_1(a, b, c; z) = \{\int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tz)^{-a}\mathrm{d}t\}/\mathrm{B}(b, c-b)$ with $\mathrm{B}(x, y)$ the beta function (Abramowitz and Stegun, 1965, chapter 15), i. e.,

$$f(\alpha \mid D, D_0, \hat{\theta} = \hat{\theta}_0) = \frac{(\alpha/c + 1)^{-1/2}\, \mathrm{Be}(\alpha \mid p + 1/2, q)}{{}_2F_1(1/2, p + 1/2, p + q + 1/2; -1/c)}. \tag{3}$$

where $c = \sigma_0^2/\sigma^2$ is the relative variance. The distribution (3) is close to a $\mathrm{Be}(p + 1/2, q)$ distribution which is hardly different from the $\mathrm{Be}(p, q)$ distribution despite perfect compatibility of both data sets. Importantly, the marginal posterior (3) does not depend on the actual value of the standard errors σ and σ_0 but only the variance ratio c. This means that (3) holds for finite standard errors but also in the idealized mathematical situation where both standard errors go equally fast to zero (i. e., infinite sample size), but with possibly different starting values ($c \neq 1$). Typically, the historical data are predetermined and only the standard error of the current study can be changed. It is therefore interesting to study the behavior of (3) for $c \to \infty$, i. e., the current standard error σ goes to zero while the historical standard error σ_0 remains fixed, reflecting an arbitrary increase of the current sample size. In that case it is straightforward to see from the power series representation of the hypergeometric function that

$$\lim_{c \to \infty} {}_2F_1(1/2, p + 1/2, p + q + 1/2; -1/c) = \lim_{c \to \infty} 1 + \mathcal{O}(1/c) = 1.$$

Hence, the limiting posterior density is

$$\lim_{c \to \infty} f(\alpha \mid D, D_0, \hat{\theta} = \hat{\theta}_0) = \mathrm{Be}(\alpha \mid p + 1/2, q), \tag{4}$$

that is, again a beta density but with updated success parameter $p+1/2$, so just slightly different from the prior.

The second situation in which the marginal posterior (2) is available in closed form is the limiting case when the current standard error σ goes to zero while the historical standard error σ_0 remains fixed (but the parameter estimates can take different values). In this case, the integral in (2) can be represented by the confluent hypergeometric function $M(a, b, z) = \{\int_0^1 \exp(zt) t^{a-1}(1-t)^{b-a-1}\,\mathrm{d}t\}/\mathrm{B}(b-a, a)$ (Abramowitz and Stegun, 1965, chapter 13) so that the marginal posterior is given by

$$\lim_{\sigma \downarrow 0} f(\alpha \mid D, D_0) = \mathrm{Be}(\alpha \mid p + 1/2, q)\, \frac{\exp\left\{-\alpha\,(\hat{\theta} - \hat{\theta}_0)^2/(2\sigma_0^2)\right\}}{M\{p + 1/2, p + q + 1/2, -(\hat{\theta} - \hat{\theta}_0)^2/(2\sigma_0^2)\}}. \tag{5}$$

The distribution (5) reduces to (4) when the parameter estimates are equal ($\hat{\theta} = \hat{\theta}_0$) since then the right fraction becomes one, which can be shown using the power series representation of the confluent hypergeometric function. Figure 2 shows the distribution (5) for different values of $|\hat{\theta} - \hat{\theta}_0|/\sigma_0$.

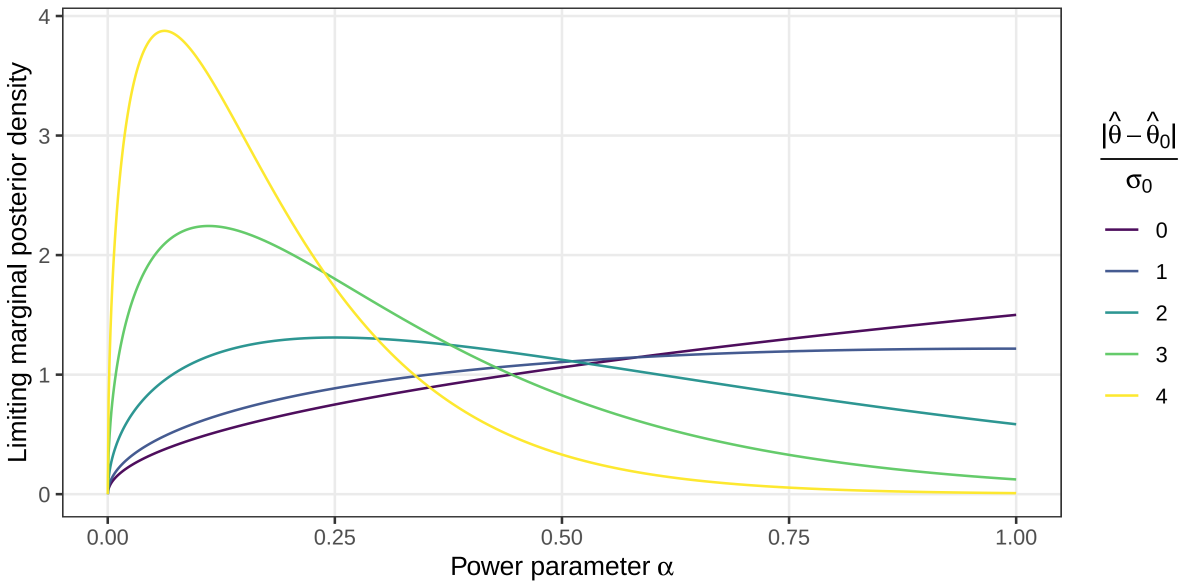

Figure 2: Limiting marginal posterior distribution of power parameter α based on $\alpha \sim \mathrm{Be}(1, 1)$ prior and when the current standard error goes to zero ($\sigma \downarrow 0$), for different values of the parameter difference standardized by the historical standard error $|\hat{\theta} - \hat{\theta}_0|/\sigma_0$.

One can see that when the parameter estimates become more different (larger $|\hat{\theta}_0 - \hat{\theta}|$) the limiting distribution (5) will be increasingly shifted towards smaller values of α indicating more incompatibility among the data sets. This shift is amplified by decreasing historical standard errors σ_0, meaning that the posterior can become arbitrarily peaked by increasing the sample size of the historical study. In contrast, when the parameter estimates are the same ($\hat{\theta} = \hat{\theta}_0$) the historical standard error σ_0 does not influence the posterior. Since the limiting marginal posterior for α can in the best case be a $\mathrm{Be}(p + 1/2, q)$, this implies that a complete pooling of historical and current data ($\alpha = 1$) can never be achieved. However, the fact that conflict between the current and historical data can make the marginal posterior arbitrarily peaked at $\alpha = 0$ implies that at least a complete discounting ($\alpha = 0$) is possible.

3 Binomial model

We will now show that the previous result holds also approximately for binary data. Let $D = \{x, n\}$ and $D_0 = \{x_0, n_0\}$ denote the number of successes and number of total trials from a current data and a historical data set, respectively. For each of them, assume a binomial likelihood with success probability θ, and let $\hat{\theta} = x/n$ and $\hat{\theta}_0 = x_0/n_0$ denote the respective maximum likelihood estimates. Assigning an improper initial beta prior $\theta \sim \text{Be}(0, 0)$ for the success probability and a beta prior $\alpha \sim \text{Be}(p, q)$ for the power parameter leads to the normalized power prior

$$f(\theta, \alpha \,|\, D_0) = \text{Be}\{\theta \,|\, \alpha\, n_0\, \hat{\theta}_0, \alpha\, n_0(1 - \hat{\theta}_0)\}\, \text{Be}(\alpha \,|\, p, q). \tag{6}$$

Combining the prior (6) with the likelihood of the current data leads to a joint posterior distribution for θ and α, from which a marginal posterior for α can be obtained by integrating out θ, i. e.,

$$f(\alpha \,|\, D, D_0) = \frac{\text{BeBin}\{x \,|\, n, \alpha\, n_0\, \hat{\theta}_0, \alpha\, n_0(1 - \hat{\theta}_0)\}\, \text{Be}(\alpha \,|\, p, q)}{\int_0^1 \text{BeBin}\{x \,|\, n, \alpha'\, n_0\, \hat{\theta}_0, \alpha'\, n_0(1 - \hat{\theta}_0)\}\, \text{Be}(\alpha' \,|\, p, q)\, \mathrm{d}\alpha'} \tag{7}$$

with $\text{Bin}(\cdot \,|\, n, \theta)$ the binomial probability mass function and $\text{BeBin}(\cdot \,|\, n, p, q)$ the Beta-binomial probability mass function.

As for the normal model, the integral in the denominator of (7) is generally not available in closed form. Yet again, it is possible to obtain a closed form expression when the probability estimates from both studies are equivalent ($\hat{\theta} = \hat{\theta}_0$). Application of Stirling's approximation $\text{Be}(x, y) \approx \sqrt{2\pi} x^{x-1/2} y^{y-1/2} (x + y)^{-(x+y-1/2)}$ to both beta functions in the probability mass function of the beta-binomial leads then to

$$\text{BeBin}\{x \,|\, n, \alpha\, n_0\, \hat{\theta}, \alpha\, n_0(1 - \hat{\theta})\} \approx \binom{n}{x} \hat{\theta}^{\hat{\theta} n} (1 - \hat{\theta})^{\hat{\theta} n} \sqrt{\frac{\alpha n_0}{n + \alpha n_0}}. \tag{8}$$

Using the approximation (8) in numerator and denominator of (7) produces the same expression (**??**) as for normal data but with with $c = n/n_0$.

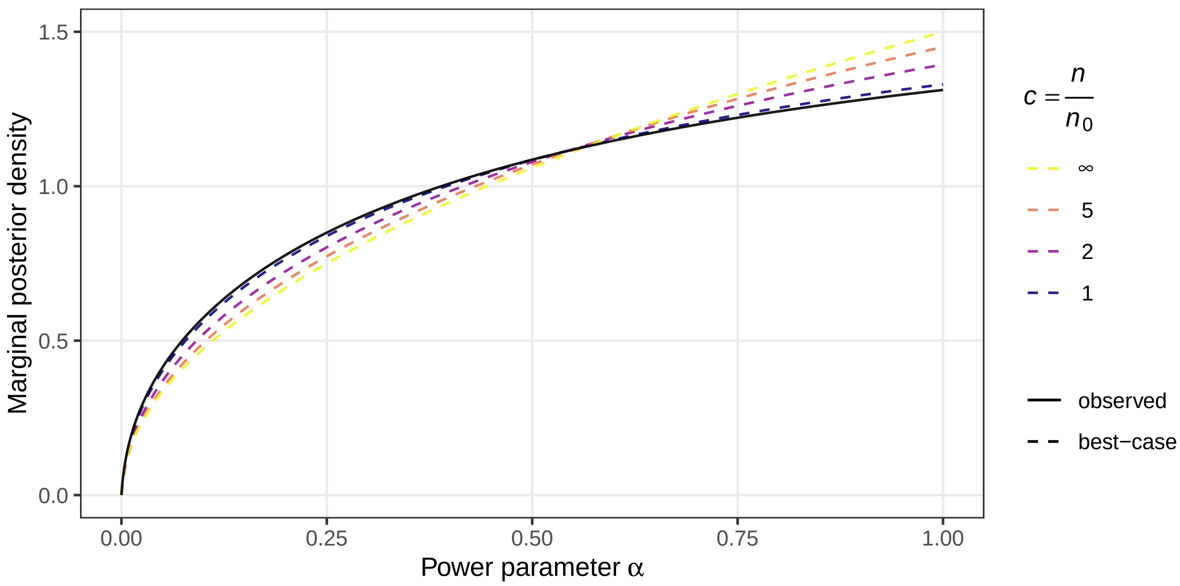

Figure 3: Marginal posterior of power parameter α for binomial data and based on $\alpha \sim \text{Be}(1, 1)$ prior. The dashed lines show the marginal posterior for the historical data $D_0 = \{x_0 = 214, n_0 = 302\}$ from Louie et al. (2011) and hypothetical current data $D = \{x = c \times x_0, n = c \times n_0\}$ mirroring original data but with relative sample sizes $c = n/n_0$. The black solid line shows the posterior for the actually observed current data $D = \{x = 193, n = 270\}$ from Cornely et al. (2012) computed by numerical integration.

4 Discussion

We showed that power priors in normal and binomial models combined with beta priors assigned to the power parameter α have undesirable and counterintuitive properties. Specifically, in the best-case scenario when the current data perfectly mirror the historical data and the sample sizes from both data sets become arbitrarily large, the marginal posterior of α does not converge to a point mass at $\alpha = 1$ but approaches a $\alpha \sim \text{Be}(p + 1/2, q)$ distribution, hardly differing from the prior $\alpha \sim \text{Be}(p, q)$. The result implies that a complete pooling of historical and current data can never be achieved. This drawback of the power prior is not only mathematical curiosity but a real concern in statistical modeling of medical data as illustrated by the case studies from Louie et al. (2011) and Cornely et al. (2012).

Uninformative priors, such as the typically assigned uniform $\alpha \sim \text{Be}(1, 1)$, hence very much limit the amount of possible borrowing since the resulting best-case posterior still gives substantial mass to small values of α. Data analysts who want to enable strong borrowing thus need to specify informative priors that give more mass to larger values of α. A pragmatic alternative is to specify α via an empirical Bayes approach as proposed by Gravestock and Held (2017), which permits complete pooling of both data sets.

We only studied the limiting marginal posterior of α in the normal and binomial models combined with beta priors on α, yet we conjecture that the issue is more fundamental and also present in other types of models. However, this will likely be more difficult to establish as marginal posteriors are typically not available in closed form for more complex models.

For the normal model, there is an exact correspondence between power parameter models with fixed α and hierarchical (random-effects meta-analysis) models with fixed heterogeneity variance (Chen and Ibrahim, 2006). This connection may provide an intuition for why the counterintuitive result occurs: Precisely estimating a heterogeneity variance from two observations alone (the historical and current data sets) seems like an impossible task as the "unit of information" is the number of data sets and not the number of samples within a data set. We will report in a future study about the precise connection between power parameter and hierarchical models when power parameter and the heterogeneity variance are random.

References

Abramowitz, M. and Stegun, I. A., editors (1965). *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables.* Dover Publications, Inc., New York.

Chen, M.-H. and Ibrahim, J. G. (2000). Power prior distributions for regression models. *Statistical Science*, 15(1). doi:10.1214/ss/1009212673.

Chen, M.-H. and Ibrahim, J. G. (2006). The relationship between the power prior and hierarchical models. *Bayesian Analysis*, 1(3). doi:10.1214/06-ba118.

Cornely, O. A., Crook, D. W., Esposito, R., Poirier, A., Somero, M. S., Weiss, K., Sears, P., and Gorbach, S. (2012). Fidaxomicin versus vancomycin for infection with clostridium difficile in europe, canada, and the USA: a double-blind, non-inferiority, randomised controlled trial. *The Lancet Infectious Diseases*, 12(4):281–289. doi:10.1016/s1473-3099(11)70374-7.

Duan, Y., Ye, K., and Smith, E. P. (2005). Evaluating water quality using power priors to incorporate historical information. *Environmetrics*, 17(1):95–106. doi:10.1002/env.752.

Gravestock, I. and Held, L. (2017). Adaptive power priors with empirical Bayes for clinical trials. *Pharmaceutical Statistics*, 16(5):349–360. doi:10.1002/pst.1814.

Louie, T. J., Miller, M. A., Mullane, K. M., Weiss, K., Lentnek, A., Golan, Y., Gorbach, S., Sears, P., and Shue, Y.-K. (2011). Fidaxomicin versus vancomycin for clostridium difficile nfection. *New England Journal of Medicine*, 364(5):422–431. doi:10.1056/nejmoa0910812.

Nelson, R. L., Suda, K. J., and Evans, C. T. (2017). Antibiotic treatment for clostridium difficile-associated diarrhoea in adults. *Cochrane Database of Systematic Reviews*, 2017(3). doi:10.1002/14651858.cd004610.pub5.

Neuenschwander, B., Branson, M., and Spiegelhalter, D. J. (2009). A note on the power prior. *Statistics in Medicine*, 28(28):3562–3566. doi:10.1002/sim.3722.

```r
## print R sessionInfo to see system information and package versions
## used to compile the manuscript
sessionInfo()

## R version 4.2.0 (2022-04-22)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 20.04.4 LTS
##
## Matrix products: default
## BLAS:   /usr/lib/x86_64-linux-gnu/openblas-pthread/libblas.so.3
## LAPACK: /usr/lib/x86_64-linux-gnu/openblas-pthread/liblapack.so.3
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=en_US.UTF-8        LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=en_US.UTF-8    LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=en_US.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] hypergeo_1.2-13 ggplot2_3.3.6   knitr_1.39
##
## loaded via a namespace (and not attached):
##  [1] pillar_1.7.0      compiler_4.2.0    highr_0.9         tools_4.2.0
##  [5] digest_0.6.29     viridisLite_0.4.0 evaluate_0.15     lifecycle_1.0.1
##  [9] tibble_3.1.7      gtable_0.3.0      pkgconfig_2.0.3   rlang_1.0.2
```

```
## [13] cli_3.3.0          DBI_1.1.2          xfun_0.31          withr_2.5.0
## [17] dplyr_1.0.9        stringr_1.4.0      generics_0.1.2     vctrs_0.4.1
## [21] contfrac_1.1-12    elliptic_1.4-0     grid_4.2.0         tidyselect_1.1.2
## [25] glue_1.6.2         deSolve_1.32       R6_2.5.1           fansi_1.0.3
## [29] purrr_0.3.4        farver_2.1.0       magrittr_2.0.3     scales_1.2.0
## [33] ellipsis_0.3.2     MASS_7.3-56        assertthat_0.2.1   colorspace_2.0-3
## [37] labeling_0.4.2     utf8_1.2.2         stringi_1.7.6      munsell_0.5.0
## [41] crayon_1.5.1
```