Bayesian Sample Size Justification

Aaron R. Caldwell

06 May, 2021

Introduction

In this study we collected data on 300 sport and exercise science research articles (100 from 3 journals). Based on the work of Büttner et al. (2020), we anticipated at least 150 (50%) of the articles would include a hypothesis that was tested. Further, based on the work of Fanelli (2010), Scheel, Schijen, and Lakens (2020), and Büttner et al. (2020) we hypothesized that the percentage of articles that find support for their hypothesis was greater than 80%.

Hypothesis

For this study, we hypothesized that the rate of positive results (i.e., studies that find at least partial support for their hypothesis) was greater than 80%. Therefore, the null hypothesis (H_0) was that the proportion of positive results was less than .8 and our alternative was greater than .8. There was no other effect being estimated in this study therefore the intercept of the model is what will be tested.

$H_0 : Intercept \leq .8$

$H_1 : Intercept > .8$

Prior Choice

The prior we selected for this analysis was informed by the previous studies assuming the true positive rate is approximately 85% (Fanelli 2010). However, we would like to avoid "spiking" the prior in favor of our hypothesis and therefore want a skeptical prior. Based on the work of Scheel, Schijen, and Lakens (2020) and Büttner et al. (2020) the estimated positive rates in original research investigations ranged from 82%-92%, and even some fields included in the survey by Fanelli (2010) observed rates at low as ~70%. Therefore, we selected a prior of $\beta(17,3)$, and is visualized it below. This prior is centered around .85, but includes the possibility of higher (.9) and much lower (.7) proportions as compatible parameter estimates.
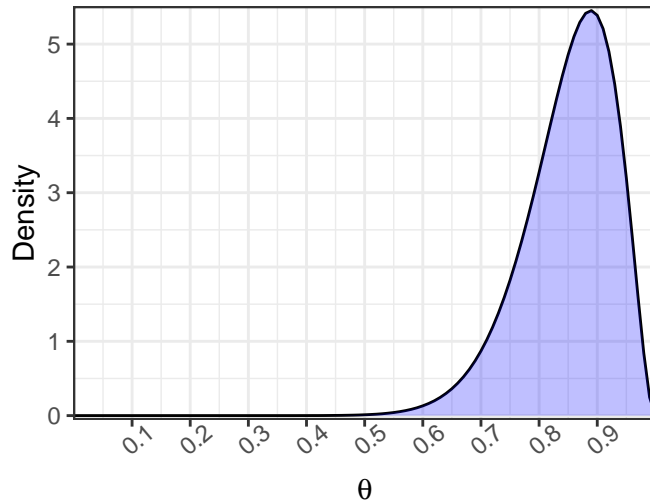
Data Analysis Example

Below, we have incorporated this prior (`prior_1`) and then analyzed this with the `brm` function (saved as `m_test`).

```
#Set prior
prior_1 = set_prior("beta(17, 3)", class = "b", lb = 0, ub = 1)

#Generate test data
test_df = data.frame(run = 1,
                     pos = rbinom(1, 150, .85),
                     N = rep(150, 1)) %>%
  mutate(rate = pos/N)

#Build model
m_test <- brm(
  pos | trials(N) ~ 0 + Intercept,
  family = binomial(link = "identity"),
  prior = prior_1,
  data = test_df,
  sample_prior = TRUE,
  iter = 1e4,
  cores = 4,
  refresh = 0
)
```

We can then visualize the performance of the prior and the posterior from this model.
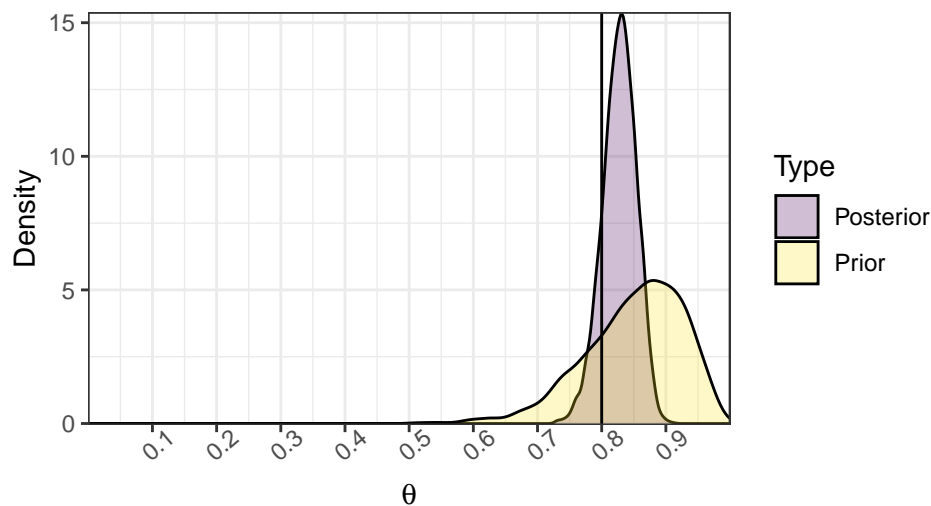
In addition, the hypothesis was tested with the `hypothesis` function and the posterior compatibility intervals (C.I.).

```
h_test <- hypothesis(m_test, "Intercept > 0.8")
knitr::kable(h_test$hypothesis, caption = "Hypothesis Test")
```

Table 1: Hypothesis Test

Hypothesis	Estimate	Est.Error	CI.Lower	CI.Upper	Evid.Ratio	Post.Prob	Star
(Intercept)-(0.8) > 0	0.0273387	0.0263142	-0.0177693	0.067925	5.688963	0.8505	

```
test_pos = posterior_interval(m_test,
                    prob = .95)
knitr::kable(test_pos, caption = "95% Posterior C.I.")
```

Table 2: 95% Posterior C.I.

	2.5%	97.5%
b_Intercept	0.7726666	0.8743508
prior_b	0.6761030	0.9658665
lp___	-5.6546314	-3.0958643

From the simulated scenario we find that given the data the hypothesis that the true positive result rate is greater than 80% is 5.69 times more likely than the true value being less than 80%.

References

Büttner, Fionn, Elaine Toomey, Shane McClean, Mark Roe, and Eamonn Delahunt. 2020. "Are Questionable Research Practices Facilitating New Discoveries in Sport and Exercise Medicine? The Proportion of Supported Hypotheses Is Implausibly High." *British Journal of Sports Medicine*, July, bjsports–2019–101863. https://doi.org/10.1136/bjsports-2019-101863.

Fanelli, Daniele. 2010. "'Positive' Results Increase down the Hierarchy of the Sciences." Edited by Enrico Scalas. *PLoS ONE* 5 (4): e10068. https://doi.org/10.1371/journal.pone.0010068.

Scheel, Anne M., Mitchell Schijen, and Daniel Lakens. 2020. "An Excess of Positive Results: Comparing the Standard Psychology Literature with Registered Reports," February. https://doi.org/10.31234/osf.io/p6e9c.