hybridpower: An R package for Bayesian-classical power analysis

Joonsuk Park

Independent Scholar

and

Jolynn Pek

The Ohio State University

Joonsuk Park, Independent Scholar; Jolynn Pek, Department of Psychology, The Ohio State University

Author note: Correspondence concerning this article should be addressed to Joonsuk Park, park.1952@osu.edu. Parts of this work have been presented at the Annual Meeting of the Society for Multivariate Experimental Psychology in Albuquerque, New Mexico by Joonsuk Park.

HybridPower: An R package for Bayesian-classical hybrid power analysis

Statistical power is central to research design and sample size determination (Cohen, 1988). For any specific data analytic procedure (e.g., one-sample t-test), statistical power $(1-\beta)$, Type I error (α), sample size (N), and the population effect size (θ) and nuisance parameters are functions of one another. The *classical* perspective to study design rests on varying inputs of this function to obtain N or power. Sample size for a study can be determined by selecting a desired level of power (e.g., $[1-\beta] = .80$), specifying α, and selecting values of the effect size and nuisance parameters. Alternatively, power can be calculated from specifying N, α, and values of the effect size and nuisance parameters. Power calculations used to be provided in tables, but modern computers promote the use of power calculators and specialized software.

There are several software packages that can determine sample sizes or compute power. Example software are PROC POWER in SAS, SamplePower in SPSS, the pwr package in R (Champely, 2018), the pwr2ppl package in R (Aberson, 2019), the MOTE package in R (Buchanan, Gillenwaters, Scofield, & Valentine, 2019), certain functions in the MBESS package in R (Kelley, 2007), and G*power (Faul, Erdfelder, Lang, & Buchner, 2007; Faul, Erdfelder, Buchner, & Lang, 2009). These power calculators approach statistical power from the classical perspective described above (see also Cohen, 1988; Lipsey, 1990). Beyond classical power calculations, the MBESS R package can also calculate power from the classical perspective and the accuracy in parameter estimation (AIPE) perspective (e.g., Kelley & Maxwell, 2003; Maxwell, Kelley, & Rausch, 2008). The AIPE perspective determines sample size by specifying a confidence interval with particular width and coverage probability (e.g., $[1-\alpha] = .95$). The AIPE perspective avoids the unrealistic assumption made by the classical approach that the specified (unknown) population effect size value is correct. Stated differently, a vulnerability of power calculated from the classical approach is local optimality (Du & Wang, 2016) where power estimates are valid if and only if the specified inputs (e.g., θ) are perfectly correct.

Although promising, the AIPE has not been readily applied in practice, leaving classical power the dominant approach to study design. To address uncertainty inherent in not knowing the precise value of the unknown effect size θ in classical power analysis, Pek and Park (2019) developed the *Bayesian-classical hybrid* approach to power analysis. This new approach combines classical power analysis with the Bayesian idea of expressing uncertainty about the effect size with a prior distribution, $\pi(\theta)$, where θ is the unknown population effect size of interest. Within the Bayesian literature, this prior is called a design prior (Beavers & Stamey, 2012; O'Hagan & Stevens, 2001; Wang & Gelfand, 2002) because it formalizes the sampling space of effect sizes considered in study design. Note that the uncertainty represented in $\pi(\theta)$ is epistemic and due to lack of knowledge. With full information on the population, $\pi(\theta)$ will converge to a point mass θ. By incorporating a design prior in classical power analysis, the Bayesian-classical hybrid approach formally expresses uncertainty about not knowing the true value of θ, resulting in a distribution of power estimates (cf., the classical approach yields a single value of power).

The Bayesian-classical hybrid approach to power analysis incorporates two sources of uncertainty: (a) sampling variability that is expressed under the classical approach with the sampling distribution of the test statistic, and (b) uncertainty of not knowing the true value of θ that is manifest in a distribution of power values. Bayesian-classical power is calculated by sampling effect sizes from a design prior $\pi(\theta)$, computing power for each sampled value of θ, and collecting the power estimates to construct a distribution. With a distribution of power estimates, sample size can be determined from the mean power value (called assurance; Chen, Fraser, & Cuddeback, 2018; O'Hagan & Stevens, 2001; O'Hagan, Stevens, & Campbell, 2005), or other quantities such as the median, quartiles, or specific quantiles of the power distribution. We recommend visualizing such power distributions to examine its shape and dispersion that informs on the uncertainty about power. When researchers consider the variability in power distributions when determining sample size of a study, they are taking into account sampling variability and uncertainty about the value of θ .

The purpose of this paper is to introduce an R package, `HybridPower`, which was developed to facilitate the easy application of power analysis in the classical and Bayesian-classical hybrid perspectives. The R statistical language is a freely available software environment for statistical computing, which is widely used by social and behavioral scientists (cf. licensed software such as SPSS, STATA and SAS). Some aspects of `HybridPower` can also be executed via `RShiny`, which allows users unfamiliar with R to make use of the functions in `HybridPower` through a web application. The package supports popular NHST procedures including t-tests, linear models (i.e., ANOVA and simple linear regression), correlations, and categorical data analysis (e.g., 2×2 tables). The package also provides new functionalities that have not been available in traditional power analysis tools, including power analysis procedures for non-parametric tests such as the sign test. In the following sections, we introduce the package and illustrate how to compute power with `HybridPower` from the classical and Bayesian-classical hybrid approaches for selected examples.

The paper is organized as follows. First, we describe the structure of `HybridPower`, demonstrating features of the package using the one-sample t-test. In this section we review how classical power and Bayesian-classical hybrid power is computed and summarized with `HybridPower`. Next, we illustrate how `HybridPower` computes classical power for example procedures such as the two-sample t-test, one-way ANOVA, and correlation. Then, we show how to obtain Bayesian-classical hybrid power estimates for three popular designs: (a) factorial ANOVA, (b) simple linear regression, and (c) test of proportions from a two-way table. We then provide a list of procedures where Bayesian-classical hybrid power can be computed using the package and outline future extensions to the package.

Computing Power with `HybridPower`

Structure of the Package

`HybridPower` is written such that different procedures are implemented as different classes within the package. For example, calculations of power for the one-sample t-test is implemented within the class `HybridPowerTtest`. Other classes and their required inputs are summarized in Table X. The package can compute classical power as well as Bayesian-classical hybrid power for the same procedure. The a package also includes additional functions to summarize distributions of power estimates calculated from the Bayesian-classical hybrid perspective that we describe below.

Before computing power for a particular procedure, an object that contains information required to compute power is first created as an instance within a class or procedure. Suppose that this object is called `x`; then the `inputs` will contain information such as sample size, the Type I error rate α, and specified values of effect sizes and nuisance parameters. What is specified in `inputs` is specific to the procedure. For the same procedure, the inputs are different for classical power versus Bayesian-classical power. Information for both classical and Bayesian-classical power, however, can be combined into a single object. Below, we demonstrate these three versions of inputs for the one-sample t-test.

Classical power. To compute classical power for the one-sample t-test, the object `x_classical` with inputs is declared with the following code:

```
x_classical <- HybridPowerTtest$input(
    ns = seq(10, 90, 10),
    d = 0.5)
```

where `ns` contains a vector of sample sizes, and `d` is the the unknown effect size in the scale of Cohen's d. Classical power is then computed by calling the function `classical_power()` with the following code:

```
x_classical$classical_power()
```

which will return power for the sample sizes of $N = 10, 20, 30, 40, 50, 60, 70, 80$, and 90 as

```
[1] 0.2928286 0.5644829 0.7539627 0.8693979 0.9338976
[6] 0.9677886 0.9847848 0.9929987 0.9968496
```

Bayesian-classical power. Continuing with the one-sample t-test, the inputs required for Bayesian-classical power are sample sizes `ns`, the number of draws from the design prior `n_prior`, the functional form of the prior distribution `prior`, and the parameters of the prior distribution. Given that the prior distribution is normal, the parameters of this prior distribution are the mean or the expected effect size in the scale of Cohen's *d* in `prior_mu` and the standard deviation of this effect size in the scale of Cohen's *d* in `prior_sigma`. These inputs are declared with the following code:

```
x_hybrid <- HybridPowerTtest$input(
  ns = seq(10, 90, 10),
  n_prior = 1000,
  prior = 'normal',
  prior_mu = 0.5,
  prior_sigma = 0.1)
```

When `hybrid_power()` is called with the following code:

```
x_hybrid$hybrid_power()
```

A distribution of power estimates for each sample size N will be generated. In this example, this distribution for each sample size N is 1,000 as declared in `n_prior = 1000`. The large number of power values can be numerically summarized by calling the function `assurances()`, which outputs the mean power values for each N.

```
x_hybrid$assurances()
```

The assurances or the means of the distributions associated with $N = 10, 20, 30, 40, 50, 60, 70, 80,$ and 90 are .30, .58, .73, .83, .88, .93, .95, .96, and .99, respectively.

Note that the calculated power values and their sample sizes can be saved into a data frame that we label `output`. This data frame will have two variables called `n` and `power` that contain values of the specified sample sizes and the calculated power, respectively. In the example above, the data frame will have 9 levels of N, each with 1,000 estimates of power, resulting in a data set of 9000 rows and 2 columns. Applying function `quantiles()` on `output` can output values associated with the specified quantiles of the distribution of power values. We provide example code showing how to subset `ouput` for a specific sample size. Additionally, calling the function plot() and applying it out output will generate box plots for each N to visually summarize the distribution. The code below shows how the functions can be used to summarize the distribution of power values:

```
output <- x_hybrid$hybrid_power()
quantile(out$power[which(out$n=='10')],probs=c(0,.25,.5,.75,1))
quantile(out$power[which(out$n=='20')],probs=c(0,.25,.5,.75,1))
x_hybrid$boxplot(output)
```

The quantiles specified in the code above relates to the five number summary that includes the minimum (0th percentile), the first quartile or 25th percentile, the median or 50th percentile, the third quartile or 75th percentile, and the maximum or 100th percentile.

```
> quantile(output$power[which(out$n=='10')],probs=c(0,.25,.5,.75,1))
        0%        25%        50%        75%       100%
0.07592246 0.22440293 0.28561685 0.35129475 0.62081951
> quantile(output$power[which(out$n=='20')],probs=c(0,.25,.5,.75,1))
       0%       25%       50%       75%      100%
0.1150023 0.4461660 0.5563449 0.6709731 0.9237371
```

These values are visualized in the box plot in Figure 1, where the edges of the boxes communicate the first and third quartiles ($Q1$ and $Q3$ respectively), the bar in the center of the box depicts the median, and the diamond in the box represents the mean or assurance. The whiskers extend to $1.5 \times IQR$, where the IQR is the interquartile range computed as $Q3 - Q1$, and outliers beyond the whiskers are represented as dots.

Classical and Bayesian-classical power. When the researcher intends to compute both classical and Bayesian-classical power, the inputs for each perspective can be combined using the following code:

```
x_both <- HybridPowerTtest$input(
  ns = seq(10, 90, 10),
```

```
d = 0.5,
n_prior = 1000,
prior = 'normal',
prior_mu = 0.5,
prior_sigma = 0.1)
```

Given this input, both perspectives to computing power can be obtained by calling the functions of `classical_power()` and `hybrid_power()`. Similarly, functions to summarize the distribution of power values from the Bayesian-classical hybrid approach can be called using the following code:

```
x_both$classical_power()
x_both$assurances()
output <- x_both$hybrid_power()
x_both$boxplot(output)
quantile(out$power[which(out$n=='90')],probs=c(0,.25,.5,.75,1))
```

In the sections to follow, we highlight different classes or procedures implemented in `HybridPower`, providing examples of their inputs.

Classical Power Examples

Bayesian-Classical Hybrid Power Examples

Procedures

One-sample t-test

Paired samples t-test

Independent samples t-test (equal variances)

Independent samples t-test (unequal variances) One-way ANOVA (equal variances)

One-way ANOVA (unequal variances)

Two-Way ANOVA (equal variances)

Correlation

Simple linear regression

Test of proportions - 1-way table

2-way tables

McNemar's test

Future Developments

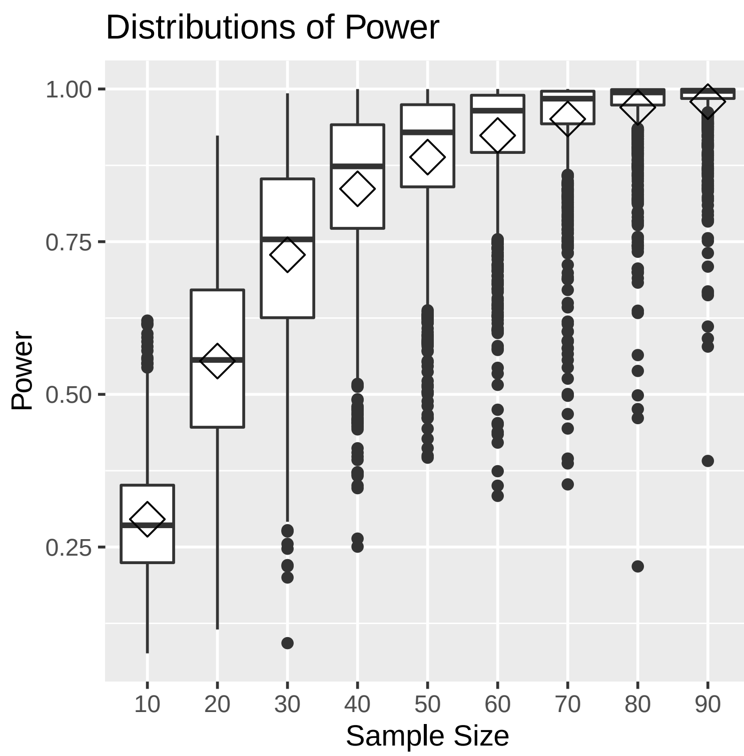

Figure 1: Box plots for distributions of power values for the one-sample t-test computed from the Bayesian-classical hybrid approach. The design prior is a normal distribution with mean .5 and standard deviation .01. Estimates are obtained with Monte Carlo simulation of 1,000 draws per sample size N.

References

Aberson, C. L. (2019). *pwr2ppl: Power analysis for common designs. r package version 0.1.* [Computer software]. Retrieved from https://CRAN.R-project.org/package=pwr2ppl

Beavers, D. P., & Stamey, J. D. (2012). Bayesian sample size determination for binary regression with a misclassified covariate and no gold standard. *Computational Statistics & Data Analysis*, *56*(8), 2574–2582. doi: 10.1016/j.csda.2012.02.014

Buchanan, E. M., Gillenwaters, A. M., Scofield, J. E., & Valentine, K. D. (2019). *Mote: R package version 1.02.* [Computer software]. Retrieved from https://CRAN.R-project.org/package=MOTE

Champely, S. (2018). *pwr: Basic functions for power analysis.* [Computer software]. Retrieved from `https://CRAN.R-project.org/package=pwr`

Chen, D.-G., Fraser, M. W., & Cuddeback, G. S. (2018). Assurance in intervention research: A Bayesian perspective on statistical power. *Journal of the Society for Social Work and Research*, *9*(1), 159–173. doi: 10.1086/696239

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Academic Press.

Du, H., & Wang, L. (2016). A Bayesian power analysis procedure considering uncertainty in effect size estimates from a meta-analysis. *Multivariate Behavioral Research*, *51*(5), 589–605. doi: 10.1080/00273171.2016.1191324

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. doi: 10.3758/brm.41.4.1149

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. doi: 10.3758/bf03193146

Kelley, K. (2007). Methods for the behavioral, educational, and social sciences: An r package. *Behavior Research Methods*, *39*(4), 979–984. doi: 10.3758/bf03192993

Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods*, *8*(3), 305–321. doi: 10.1037/1082-989X.8.3.305

Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research*. Newbury Park, CA: Sage.

Maxwell, S. E., Kelley, K., & Rausch, J. R. (2008). Sample size planning for statistical power and accuracy in parameter estimation. *Annual Review of Psychology*, *59*, 537–563. doi: 10.1146/annurev.psych.59.103006.093735

O'Hagan, A., & Stevens, J. W. (2001). Bayesian assessment of sample size for clinical trials of cost-effectiveness. *Medical Decision Making*, *21*(3), 219–230. doi: 10.1177/0272989x0102100307

O'Hagan, A., Stevens, J. W., & Campbell, M. J. (2005). Assurance in clinical trial design. *Pharmaceutical Statistics*, *4*(3), 187–201. doi: 10.1002/pst.175

Pek, J., & Park, J. (2019). Complexities in power analysis: Quantifying uncertainties with a Bayesian-classical hybrid approach. *Psychological Methods*(4), 590–605. doi: 10.1037/met0000208

Wang, F., & Gelfand, A. E. (2002). A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models. *Statistical Science*, *17*(2), 193–208. doi: 10.1214/ss/1030550861