BayesPPD: An R Package for Bayesian Sample Size Determination Using the Power and Normalized Power Prior for Generalized Linear Models

by Yueqi Shen, Matthew A. Psioda and Joseph G. Ibrahim

Abstract The R package **BayesPPD** (Bayesian Power Prior Design) supports Bayesian power and type I error calculation and model fitting after incorporating historical data with the power prior and the normalized power prior for generalized linear models (GLM). The package accommodates summary level data or subject level data with covariate information. It supports use of multiple historical datasets as well as design without historical data. Supported distributions for responses include normal, binary (Bernoulli/binomial), Poisson and exponential. The power parameter can be fixed or modeled as random using a normalized power prior for each of these distributions. In addition, the package supports the use of arbitrary sampling priors for computing Bayesian power and type I error rates. In addition to describing the statistical methodology and functions implemented in the package to enable sample size determination (SSD), we also demonstrate the use of **BayesPPD** in two comprehensive case studies.

Introduction: BayesPPD

There has been increasing interest over the past few decades in incorporating historical data in clinical trials, particularly on controls (Pocock, 1976; Neuenschwander et al., 2010; Viele et al., 2014). Use of historical data can increase effective sample size, potentially leading to more accurate point estimates and increased power (Neuenschwander et al., 2010; Viele et al., 2014). Bayesian methods provide a natural mechanism for information borrowing through the use of informative priors. Some popular informative priors for Bayesian clinical trial design include the power prior (Chen and Ibrahim, 2000), the normalized power prior (Duan et al., 2006), the commensurate power prior (Hobbs et al., 2011), and the robust meta-analytic-predictive prior (Schmidli et al., 2014).

Some advantages of the power prior include its easy construction, its natural way of incorporating historical data, its intuitive interpretation, and its desirable theoretical properties (Ibrahim et al., 2015). For example, Ibrahim et al. (2003) show that the power prior is an optimal class of informative priors in the sense that it minimizes a convex sum of the Kullback–Leibler (KL) divergences between two posterior densities, in which one density is based on no incorporation of historical data, and the other density is based on pooling the historical and current data. Duan et al. (2006) propose a modification of the power prior, the normalized power prior, which adds a normalizing constant component when the power parameter is modeled as random. The normalizing constant poses computational challenges in the presence of covariates, because it is analytically intractable except in the case of the normal linear model (Carvalho and Ibrahim, 2021). We address this challenge by utilizing the PWK estimator (Wang et al., 2018) to approximate the normalizing constant for use with generalized linear models. We also develop a novel way of incorporating the approximation of the normalizing constant into the Markov chain Monte Carlo (MCMC) algorithm.

There is a growing literature on Bayesian sample size determination, including the works of Rahme and Joseph (1998), Simon (1999), Wang and Gelfand (2002), De Santis (2007), M'Lan et al. (2006) and Joseph et al. (2008). We consider the simulation-based method developed in Chen et al. (2011) and Psioda and Ibrahim (2019), which extends the the fitting and sampling priors of Wang and Gelfand (2002) with a focus on controlling the type I error rate and calculating power. In addition, our package supports the use of arbitrary sampling priors for computing Bayesian power and type I error rates, and has specific features for GLMs that semi-automatically generate sampling priors from historical data.

The R package **BayesPPD** (Bayesian Power Prior Design) (Shen et al., 2022) supports Bayesian clinical trial design after incorporating historical data with the power prior and the normalized power prior. **BayesPPD** has two categories of functions: functions for model fitting and functions for Bayesian power and type I error rate estimation. The package accommodates summary level data or subject level data with covariate information for normal, binary (Bernoulli/binomial), Poisson and exponential models. It supports use of multiple historical datasets and design without historical data.

Several Bayesian clinical trial design packages are available on the Comprehensive R Archive Network (CRAN), such as **BDP2**, **ph2bayes** and **gsbDesign** (Kopp-Schneider et al., 2018; Nagashima, 2018; Gerber and Gsponer, 2016). However, these packages do not accommodate the incorporation of historical data and are limited to normal and binary endpoints. The **RBesT** package (Weber, 2021) accounts for historical data using the meta-analytic-predictive prior. Commercial software for clinical trial design such as **FACTS**, **East** and **ADDPLAN** (LLC Consultants, 2014; Cytel Software Corporation, 2014; Wassmer and Eisebitt, 2005) do not implement the power prior, to our knowledge. The **BayesCTDesign** (Eggleston et al., 2019) package supports two-arm randomized Bayesian trial design using historical control data with the power prior, but it does not allow covariates, nor does it allow the power parameter to be treated as random. The **NPP** (Han et al., 2021) package implements the normalized power prior for two group cases for Bernoulli, normal, multinomial and Poisson models, as well as for the normal linear model. It does not support generalized linear models, nor does it include functions for sample size determination. The **bayesDP** (Balcome et al., 2021) package implements the power prior where the power parameter is determined by a discounting function estimated based on a measure of prior-data conflict. Thus, this approach is not fully Bayesian, and the package must be used in conjunction with the package **bayesCT** (Chandereng et al., 2020) for trial design. While **bayesDP** supports two-arm trials for binomial, normal and survival models as well as linear and logistic regression models, **BayesPPD** allows covariates for Bernoulli/binomial, normal, Poisson and exponential models with several choices of link functions. The **BayesPPD** package is a comprehensive resource that supports Bayesian analysis and design using the power prior and normalized power prior.

Another advantage of **BayesPPD** is its computational speed. **BayesPPD** implements MCMC algorithms with **Rcpp** (Eddelbuettel and Francois, 2011) without recourse to asymptotics. For most sample sizes, functions for analysis take only a few seconds to run. Functions for design for two group cases run in seconds for fixed a_0, and generally run in less an hour for random a_0, depending on the desired level of precision (e.g., number of simulated datasets). In the presence of covariates, functions for design are more computation-intensive; an approximation method based on asymptotic theory has been implemented to help users obtain a rough estimate of the desired sample size before fine-tuning using the MCMC-based method.

This article is organized as follows. We first describe the methods implemented by the package. We then provide details on how to use **BayesPPD** for different data scenarios and model needs. We also present two case studies with example code, one with covariates and one without. The article is concluded with a brief discussion.

Theoretical framework

Basic formulation of the power prior

Let D denote data from the current study and D_0 denote data from a historical study. Let θ denote model parameters and $L(\theta|D)$ denote a general likelihood function associated with a given outcome model, such as a linear model, generalized linear model (GLM), survival model, or random effects model. Following Chen and Ibrahim (2000), the power prior is formulated as

$$\pi(\theta|D_0, a_0) \propto L(\theta|D_0)^{a_0}\pi_0(\theta).$$

where $0 \leq a_0 \leq 1$ is a discounting parameter for the historical data likelihood, and $\pi_0(\theta)$ is the initial prior for θ. The parameter a_0 allows researchers to control the influence of the historical data on the posterior distribution. When $a_0 = 0$, historical information is discarded and the power prior becomes equivalent to the initial prior $\pi_0(\theta)$. When $a_0 = 1$, the power prior corresponds to the posterior distribution of θ given the historical data and the initial prior. When a_0 is treated as fixed, sensitivity analysis can be performed to determine an appropriate a_0 value. When a_0 is treated as random, priors such as the beta distribution can be specified. The choice of a_0 is discussed in, for example, Ibrahim et al. (2015) and Psioda and Ibrahim (2018).

The power prior can easily accommodate multiple historical datasets. Suppose there are K historical datasets denoted by D_{0k} for $k = 1, \cdots, K$ and let $D_0 = (D_{01}, \cdots, D_{0K})$. The power prior becomes

$$\pi(\theta|D_0, a_0) \propto \prod_{k=1}^{K} L(\theta|D_{0k})^{a_{0k}}\pi_0(\theta),$$

where $a_0 = (a_{01}, \cdots, a_{0K})'$ and $0 \leq a_{0k} \leq 1$ for $k = 1, \cdots, K$.

The normalized power prior

Modeling a_0 as random allows one to represent uncertainty in how much the historical data should be discounted. The simplest power prior that allows this is the *joint power prior* (Chen and Ibrahim, 2000) which is given by

$$\pi(\theta, a_0|D_0) \propto L(\theta|D_0)^{a_0}\pi_0(\theta)\pi_0(a_0).$$

Neuenschwander et al. (2009) point out that this formulation is not ideal because the normalizing constant,

$$c(a_0) = \int L(\theta|D_0)^{a_0}\pi_0(\theta)d\theta,$$

for $L(\theta|D_0)^{a_0}\pi_0(\theta)$ is not incorporated and thus $\pi_0(a_0)$ is not actually the marginal prior for a_0. In fact, Duan et al. (2006) point out that this formulation of the power prior does not obey the likelihood principle. Duan et al. (2006) proposed a modification of the power prior, the *normalized power prior*, which is given by

$$\pi(\theta, a_0|D_0) = \pi(\theta|D_0, a_0)\pi(a_0) = \frac{L(\theta|D_0)^{a_0}\pi_0(\theta)}{c(a_0)}\pi_0(a_0),$$

where $\pi_0(a_0)$ is the initial prior for a_0. The normalized power prior specifies a conditional prior for θ given a_0 and a marginal prior for a_0. The normalizing constant,

$$c(a_0) = \int L(\theta|D_0)^{a_0}\pi_0(\theta)d\theta,$$

is often analytically intractable and requires Monte Carlo methods for estimation. When a_0 is modeled as random, the normalized power prior is implemented in **BayesPPD** using a beta initial prior on a_0, for which the user must specify values of the two shape parameters that define the beta density. The package supports the inclusion of multiple historical datasets when a_0 is modeled as random.

The power prior for generalized linear models

The power prior can easily accommodate covariates. Let y_i denote the response variable and x_i denote a p-dimensional vector of covariates for subject $i = 1, \cdots, n$. Denote $\tilde{\beta} = (\beta_0, \beta)$, where β_0 is the intercept and $\beta = (\beta_1, \cdots, \beta_p)'$ is a p-dimensional vector of regression coefficients. We assume the GLM of $y_i|x_i$ is given by

$$f(y_i|x_i, \tilde{\beta}, \tau) = \exp\{\alpha_i^{-1}(\tau)(y_i g(\beta_0 + x_i'\beta) - \psi(g(\beta_0 + x_i'\beta))) + \phi(y_i, \tau)\},$$

where τ is a scale parameter and g is a monotone differentiable link function. In particular, **BayesPPD** allows the distribution of $y_i|x_i$ to be normal, Bernoulli, binomial, Poisson or exponential. Note that for Bernoulli, binomial, Poisson and exponential regression models, τ is equal to 1.

Let $D_{0k} = \{(y_{0ki}, x_{0ki}), i = 1, \cdots, n_{0k}\}$ denote the k-th historical dataset, where y_{0ki} is the response variable for historical subject i and x_{0ki} is the p-dimensional vector of covariates for historical subject i. By default, **BayesPPD** assumes the historical data consists of control group subjects only. Therefore, the historical covariate matrix does not have the treatment indicator variable, while the current covariate matrix does. The package also allows the historical data to be used to inform the treatment effect parameter; then the historical and current covariate matrices will both have the treatment indicator.

The GLM for $y_{0ki}|x_{0ki}$ is

$$f(y_{0ki}|x_{0ki}, \tilde{\beta}, \tau_{0k}) = \exp\{\alpha_{0i}^{-1}(\tau_{0k})(y_{0ki} g(\beta_0 + x_{0ki}'\beta) - \psi(g(\beta_0 + x_{0ki}'\beta))) + \phi(y_{0ki}, \tau_{0k})\},$$

where τ_{0k} is the scale parameter for the k-th historical dataset. Note that the precision parameter is assumed to be unshared. The historical data likelihood for K historical datasets is $L(\tilde{\beta}, \tau_{01}, \cdots, \tau_{0K}|D_0) \propto \prod_{k=1}^{K}\prod_{i=1}^{n_{0k}} f(y_{0ki}|x_{0ki}, \tilde{\beta}, \tau_{0k})$. The power prior for GLMs with fixed $a_0 = (a_{01}, \cdots, a_{0K})'$ is

$$\pi(\tilde{\beta}, \tau_{01}, \cdots, \tau_{0K}|D_0, a_0) \propto \prod_{k=1}^{K}\left\{L(\tilde{\beta}, \tau_{0k}|D_{0k})^{a_{0k}}\pi_0(\tau_{0k})\right\}\pi_0(\tilde{\beta}).$$

When a_0 is modeled as random, we assume $\tau_{01}, \cdots, \tau_{0K} = \tau$ for computational simplicity. The

normalized power prior for GLMs with a random a_0 vector is given by

$$\pi(\tilde{\beta}, \tau, a_0|D_0) = \frac{\prod_{k=1}^{K} L(\tilde{\beta}, \tau|D_{0k})^{a_{0k}} \pi_0(\tilde{\beta})\pi_0(\tau)}{\int_0^\infty \int_{\mathbb{R}^p} \prod_{k=1}^{K} L(\tilde{\beta}, \tau|D_{0k})^{a_{0k}} \pi_0(\tilde{\beta})\pi_0(\tau) d\tilde{\beta} d\tau} \pi_0(a_0).$$

Estimating the normalizing constant for GLMs

The normalizing constant $c(a_0)$ in the normalized power prior for GLMs is analytically intractable except for normal linear regression models. For other types of regression models, we approximate the normalizing constant with the partition weighted kernel (PWK) estimator proposed by Wang et al. (2018). The PWK estimator requires MCMC samples from the posterior distribution (based on a discounted historical data likelihood with fixed a_0 value), which we obtain using the slice sampler (Neal, 2003), and the known kernel function for computing the normalizing constant. The authors first impose a working parameter space, defined as the space where the kernel value is bounded away from zero. As stated in Wang et al. (2018), the PWK estimator is constructed by first partitioning the working parameter space and then estimating the marginal likelihood by a weighted average of the kernel values evaluated at a MCMC sample for each partition, where the weights are assigned locally using a representative kernel value in each partitioned subset. The PWK estimator has been shown to have desirable properties, including being consistent and having finite variance (Wang et al., 2018).

The function `normalizing.constant` in our package computes a vector of coefficients that defines a function $f(a_0)$ that approximates the normalizing constant for GLMs with random a_0. Suppose there are K historical datasets. Basic usage of the `normalizing.constant` function entails the following steps:

1. The user inputs a grid of M rows and K columns of potential values for a_0.
2. For each row of a_0 values in the grid, the function obtains M samples for β from the power prior associated with the current values of a_0 using the slice sampler. Note that τ is not applicable here because the models implemented using the PWK estimator do not have scale parameters.
3. For each of the M sets of posterior samples, the PWK algorithm (Wang et al., 2018) is used to estimate the log of the normalizing constant $d_1, \cdots, d_M$ for the normalized power prior.
4. At this point, one has a dataset with outcomes $d_1, \cdots, d_M$ and predictors corresponding to the rows of the a_0 grid matrix. A polynomial regression is employed to estimate a function $d = f(a_0)$ based on these quantities. The degree of the polynomial regression is determined by the algorithm to ensure $R^2 > 0.99$.
5. The `normalizing.constant` function returns the vector of coefficients from the polynomial regression model, which the user must input into the analysis or design function for GLMs with a_0 modeled as random (`glm.random.a0` and `power.glm.random.a0`).

In the Examples section below, we demonstrate computing the normalizing constant for one historical dataset with three covariates. Due to computational intensity, the `normalizing.constant` function has not been evaluated for accuracy for high dimensional β (e.g., dimension > 10) or high dimensional a_0 (e.g., dimension > 5).

Sample size determination

Hypotheses for two group models

Following Chen et al. (2011), for two group models (i.e., treatment and control group with no covariates), denote the parameter for the treatment group by μ_t and the parameter for the control group by μ_c. For example, for binomial models, μ_t and μ_c are the probability of having some outcome (e.g., tumor response) for the treatment and control group, respectively. Let τ_c denote the nuisance parameters for the control group in the model. For normal models, τ_c is a vector of precision parameters. For K historical datasets $D_0 = (D_{01}, \cdots, D_{0K})'$ with fixed a_0, we assume each historical dataset D_{0k} has a precision parameter τ_{c0k}. When a_0 is modeled as random, the historical and current datasets are assumed to have the same precision parameter, in which case τ_c reduces to a scalar. The precision parameter of the treatment group is denoted by τ_t.

We consider the following power prior for (μ_c, τ_c) given multiple historical datasets D_0

$$\pi(\mu_c, \tau_c|D_0, a_0) \propto \prod_{k=1}^{K} \left[L(\mu_c|D_{0k}, \tau_c)^{a_{0k}}\right] \pi_0(\mu_c)\pi_0(\tau_c),$$

where $a_0 = (a_{01}, \cdots, a_{0K})'$, $0 \leq a_{0k} \leq 1$ for $k = 1, \cdots, K$, $L(\mu_c|D_{0k}, \tau_c)$ is the historical data likelihood, and $\pi_0(\mu_c)$ and $\pi_0(\tau_c)$ are the initial priors. To model a_0 as random, we consider the normalized power prior

$$\pi(\mu_c, \tau_c, a_0|D_0) \propto \frac{\prod_{k=1}^{K} [L(\mu_c|D_{0k}, \tau_c)^{a_{0k}}]\, \pi_0(\mu_c)\pi_0(\tau_c)}{c(a_0)} \pi_0(a_0),$$

where

$$c(a_0) = \int_0^\infty \int_{-\infty}^\infty \prod_{k=1}^{K} [L(\mu_c|D_{0k}, \tau_c)^{a_{0k}}] \pi_0(\mu_c)\pi_0(\tau_c) d\mu_c d\tau_c.$$

For models other than the exponential model, the power / type I error calculation algorithm assumes the null and alternative hypotheses are given by

$$H_0 : \mu_t - \mu_c \geq \delta$$

and

$$H_1 : \mu_t - \mu_c < \delta,$$

where δ is a prespecified constant. To test hypotheses of the opposite direction, i.e., $H_0 : \mu_t - \mu_c \leq \delta$ and $H_1 : \mu_t - \mu_c > \delta$, one can set the parameter `nullspace.ineq` to "<".

For positive continuous data assumed to follow exponential distribution, the hypotheses are given by

$$H_0 : \mu_t / \mu_c \geq \delta$$

and

$$H_1 : \mu_t / \mu_c < \delta,$$

where μ_t and μ_c are the hazards for the treatment and the control group, respectively.

Definition of Bayesian type I error rate and power

Let Θ_0 and Θ_1 denote the parameter spaces corresponding to H_0 and H_1. Let $y^{(n)}$ denote the simulated current data associated with a sample size of n and let $\theta = (\mu_t, \mu_c, \tau_c)$ denote the model parameters. Let $\pi^{(s)}(\theta)$ denote the sampling prior and let $\pi^{(f)}(\theta)$ denote the fitting prior. The sampling prior is used to generate the hypothetical data while the fitting prior is used to fit the model after the data is generated. Let $\pi_0^{(s)}(\theta)$ denote a sampling prior that only puts mass in the null region, i.e., $\theta \subset \Theta_0$. Let $\pi_1^{(s)}(\theta)$ denote a sampling prior that only puts mass in the alternative region, i.e., $\theta \subset \Theta_1$. To determine Bayesian sample size, we estimate the quantity

$$\beta_{sj}^{(n)} = E_s[I\{P(\mu_t - \mu_c < \delta|y^{(n)}, \pi^{(f)}) \geq \gamma\}], \tag{1}$$

where $j = 0$ or 1, corresponding to the expectation taken with respect to $\pi_0^{(s)}(\theta)$ or $\pi_1^{(s)}(\theta)$. The constant $\gamma > 0$ is a prespecified posterior probability threshold for rejecting the null hypothesis (e.g., 0.975). The probability is computed with respect to the posterior distribution given the simulated data $y^{(n)}$ and the fitting prior $\pi^{(f)}(\theta)$, and the expectation is taken with respect to the marginal distribution of $y^{(n)}$ defined based on the sampling prior $\pi^{(s)}(\theta)$. Then $\beta_{s0}^{(n)}$ corresponding to $\pi^{(s)}(\theta) = \pi_0^{(s)}(\theta)$ is the Bayesian type I error rate, while $\beta_{s1}^{(n)}$ corresponding to $\pi^{(s)}(\theta) = \pi_1^{(s)}(\theta)$ is the Bayesian power. Note that Bayesian type I error rate and power can be equivalently defined as weighted averages of the quantities based on fixed values of θ with weights determined by the sampling priors (Psioda and Ibrahim, 2018). For given $\alpha_0 > 0$ and $\alpha_1 > 0$, we can compute $n_{\alpha_0} = \min\{n : \beta_{s0}^{(n)} \leq \alpha_0\}$ and $n_{\alpha_1} = \min\{n : \beta_{s1}^{(n)} \geq 1 - \alpha_1\}$. Then, the sample size is taken to be $\max\{n_{\alpha_0}, n_{\alpha_1}\}$. Common choices of α_0 and α_1 include $\alpha_0 = 0.05$ and $\alpha_1 = 0.2$. These choices guarantee that the Bayesian type I error rate is at most 0.05 and the Bayesian power is at least 0.8.

Estimation of Bayesian type I error rate and power

In this section, we discuss the simulation-based procedure used to estimate the Bayesian type I error rate and power. Let N denote the number of simulated trials. To compute $\beta_{sj}^{(n)}$, the following algorithm is used for each simulated trial b:

- Step 1: Generate $\theta^{(b)} \sim \pi_j^{(s)}(\theta)$.
- Step 2: Generate $y^{(b)} \sim f(y^{(b)}|\theta^{(b)})$.

- Step 3: Estimate the posterior distribution $\pi(\theta|y^{(b)}, D_0, a_0)$ and the posterior probability $P(\mu_t - \mu_c < \delta|y^{(b)}, \pi^{(f)}, D_0, a_0)$.
- Step 4: Compute the indicator $r^{(b)} = I\{P(\mu_t - \mu_c < \delta|y^{(b)}, \pi^{(f)}, D_0, a_0) \geq \gamma\}$.

Then the estimate of $\beta_{sj}^{(n)}$ is $\frac{1}{N}\sum_{b=1}^{N} r^{(b)}$.

Specification for regression models

For regression models, we assume the first column of the covariate matrix is the treatment indicator, and the corresponding parameter is β_1, which, for example, corresponds to a difference in means for the linear regression model and a log hazard ratio for the exponential regression model. The hypotheses are given by

$$H_0 : \beta_1 \geq \delta$$

and

$$H_1 : \beta_1 < \delta.$$

The definition of $\beta_{sj}^{(n)}$ and the algorithm change accordingly.

Prior distributions

Two group cases

For two group models, continuous responses of the control group are assumed to follow $N(\mu_c, \tau_c^{-1})$. Each historical dataset D_{0k} is assumed to have a different precision parameter τ_{c0k}. The initial prior for the μ_c is the uniform improper prior. The initial prior for τ_c is the Jeffery's prior, τ_c^{-1}, and the initial prior for τ_{c0k} is τ_{c0k}^{-1}. Posterior samples of μ_c, τ_c and τ_{c0k}'s (if historical data is given) are obtained through Gibbs sampling. When a_0 is modeled as random, the historical datasets are assumed to have the same precision parameter τ_c as the current dataset for computational simplicity. The initial prior for τ_c is the Jeffery's prior, τ_c^{-1}. Posterior samples of a_0 are obtained through slice sampling.

For binary, count or positive continuous data, a single response from the control group is assumed to follow Bernoulli(μ_c), Poisson(μ_c) or exponential(rate=μ_c), respectively. A beta initial prior is used for μ_c for Bernoulli data, and a gamma prior is used for Poisson and exponential data. The user can specify the hyperparameters. When a_0 is modeled as random, posterior samples of a_0 are obtained through slice sampling. The conditional posterior distributions of μ_c given a_0 have closed form solutions.

When computing the power or the type I error rate, treatment group data are simulated and posterior samples of μ_t (and τ_t for normal data) are obtained using basic Bayesian models. The priors used for μ_t are the same as the initial priors used for μ_c. For normal data, the prior for τ_t is the Jeffery's prior, τ_t^{-1}.

GLM cases

For GLMs, a continuous response y_i is assumed to follow $N(\beta_0 + x_i'\beta, \tau^{-1})$. Each historical dataset D_{0k} is assumed to have a different precision parameter τ_k. The initial prior for τ is the Jeffery's prior, τ^{-1}, and the initial prior for τ_k is τ_k^{-1}. Posterior samples of β_0 and β are obtained through Gibbs sampling. For all other types of data, a link function must be applied. Posterior samples of β_0 and β are obtained through slice sampling. When a_0 is fixed, the initial prior for β_0 and β is the uniform improper prior. When a_0 is modeled as random, the historical datasets are assumed to have the same precision parameter τ as the current dataset. The initial prior for τ is the Jeffery's prior, τ^{-1}. Independent normal priors with mean zero and a user-specified variance are used for β. Here we use a proper initial prior for β to ensure the propriety of the normalized power prior. Posterior samples of a_0 are obtained through slice sampling. The normalizing constant of the normalized power prior is estimated using the PWK estimator.

	Two groups, fixed a_0	Two groups, random a_0	GLM, fixed a_0*	GLM, random a_0
Bernoulli/ Binomial	Numerical integration	Slice	Slice	Slice & PWK
Normal	Gibbs	Gibbs & Slice	Gibbs	Gibbs & Slice
Poisson	Numerical integration	Slice	Slice	Slice & PWK
Exponential	Numerical integration	Slice	Slice	Slice & PWK

Table 1: Estimation method used for each model and data type. Each column contains a type of model, and each row contains an outcome distribution. Gibbs sampling is used for normally distributed outcomes. Slice sampling is used when a_0 is modeled as random.
*Approximation method is available for sample size determination for fast implementation.

Using BayesPPD

Package overview

The **BayesPPD** package accommodates summary level data or subject level data with covariate information. It supports SSD for design applications with multiple historical datasets as well as with no historical data. Functions with names containing `"two.grp"` assume that the input data are sufficient statistics (e.g., sample mean) for independent and identically distributed treatment and control group data. Simulated control group data are analyzed using the power or normalized power prior and posterior samples of μ_c are returned. By default, functions with names containing `"glm"` assume that the historical control data include a covariate matrix X_0 and the current data include the same set of covariates with an additional column (the first column) of treatment indicator. The package assumes the historical data is composed of control group subjects only by default. If the user wants to use the historical data to inform the treatment effect, one can set `borrow.treat=TRUE` and include the treatment indicator in the historical covariate matrix. Simulated data are analyzed using the power or normalized power prior and posterior samples of the regression coefficients are returned. For each of two cases, the power parameter a_0 can be fixed or modeled as random, resulting in four model fitting functions, `two.grp.fixed.a0`, `two.grp,random.a0`, `glm.fixed.a0` and `glm.random.a0`. For each of the four model fitting functions, a companion function prefixed with `"power"` calculates power or type I error rate, given historical data and current data sample size. Supported distributions of responses include normal, binary (Bernoulli/binomial), Poisson and exponential. Since functions for sample size determination for GLMs are computationally intensive, an approximation method based on asymptotic theory has been implemented for the model with fixed a_0.

Table 1 shows the sampling methods used for each model and data distribution. Gibbs sampling is used for normally distributed data. Slice sampling (Neal, 2003) is used for all other data distributions, and for obtaining posterior samples of a_0 when a_0 is considered random. For two group models with fixed a_0, numerical integration is performed using the **RcppNumerical** package (Qiu et al., 2019). For GLMs with random a_0, the PWK estimator (Wang et al., 2018) is used to estimate the normalizing constant. The functions return S3 objects with `summary` methods implemented.

Two group cases

If one has current and/or historical control data for an application with no covariates and would like to obtain posterior samples of μ_c (and τ_c for normal data), one uses the function `two.grp.fixed.a0` or `two.grp.random.a0`. The user must specify the `data.type` (`"Normal"`, `"Bernoulli"`, `"Poisson"` or `"Exponential"`), the sum of responses `y.c`, the sample size `n.c` and the sample variance `v.c` (for normal data only) of the current control data. The optional `historical` argument is a matrix where the columns contain the sufficient statistics and each row represents a historical dataset. For `two.grp.fixed.a0`, `historical` must contain a column of a_0 values, one a_0 value for each historical dataset. For non-normal data, the user can specify `prior.mu.c.shape1` and `prior.mu.c.shape2`, the hyperparameters of the initial prior for μ_c.

When $a_0 = (a_{01}, \cdots, a_{0K})'$ is modeled as random, a beta prior is specified for a_0 with hyperparameters `prior.a0.shape1` and `prior.a0.shape2`. Posterior samples of a_0 are obtained through

slice sampling. The optional tuning parameters for the slice sampler include `lower.limits` and `upper.limits` which control the upper and lower limits of the parameters being sampled, as well as `slice.widths` which controls the width of each slice. The length of `lower.limits`, `upper.limits` and `slice.widths` should be at least equal to the number of parameters, i.e., the dimension of a_0. Their default values are 0, 1 and 0.1, respectively, for each a_{0k}.

For sample size determination, `power.two.grp.fixed.a0` and `power.two.grp.random.a0` compute the power or the type I error rate given the sample sizes of the treatment and control groups for the new study and other inputs. If a sampling prior with support in the null space is used, the value returned is a Bayesian type I error rate. If a sampling prior with support in the alternative space is used, the value returned is a Bayesian power. The arguments `samp.prior.mu.t` and `samp.prior.mu.c` contain vectors of samples for μ_t and μ_c, which are discrete approximations of the sampling priors. For normal data, arguments `samp.prior.var.t` and `samp.prior.var.c`, which contain samples for τ_t^{-1} and τ_c^{-1}, must also be provided. The argument `delta` specifies the constant that defines the boundary of the null hypothesis. The default value is zero. The argument `gamma` specifies the posterior probability threshold for rejecting the null hypothesis. The default value is 0.95.

GLM cases

If one has current and historical data for an application with covariates and would like to obtain posterior samples of β (and τ for normal data), one uses the function `glm.fixed.a0` or `glm.random.a0`. It is recommended that the covariates be transformed or standardized so that the estimation of β will be stable. The user must specify the `data.type`, the `data.link` (except for normal data), the vector of responses `y` and the matrix of covariates `x` where the first column should be the treatment indicator. Supported link functions include logit, probit, log, identity-positive, identity-probability and complementary log-log. If the data is binary and all covariates are discrete, the user can collapse the Bernoulli data into a binomial structure, which may result in a much faster slice sampler. In this case, the user needs to provide `n`, a vector of integers specifying the number of subjects who have a particular value of the covariate vector. The optional `historical` argument is a list of lists where each list contains information about a historical dataset with named elements `y0`, `x0` and `a0` (only for `glm.fixed.a0`). If `borrow.treat=FALSE` (the default), the historical covariate matrix `x0` should not have the treatment indicator. Apart from missing the treatment indicator, `x0` should have the same set of covariates in the same order as `x`. If `borrow.treat=TRUE`, `x0` should have the same set of covariates in the same order as `x`, where the first column of `x0` must be the treatment indicator. For non-normal data, slice sampling is used to obtain posterior samples of β, and the user can specify the `lower.limits`, `upper.limits` and `slice.widths` of the sampler. The length of `lower.limits`, `upper.limits` and `slice.widths` should be at least equal to the number of parameters, i.e., the dimension of β. A matrix of posterior samples of β is returned, where the first column contains posterior samples of the intercept and the second column contains posterior samples of β_1, the parameter for the treatment indicator.

When a_0 is modeled as random for non-normal data, the user must first use the function `normalizing.constant` to obtain the value of `a0.coefficients`, a vector of coefficients for a_0 necessary for estimating the normalizing constant for the normalized power prior. For the `grid` argument of `normalizing.constant`, the user inputs a grid of M rows and K columns of potential values for a_0 for K historical datasets. For example, one can choose the vector `v = c(0.1,0.25,0.5,0.75,1)` and use `expand.grid(a0_1=v,a0_2=v,a0_3=v)` when $K = 3$ to get a grid with $M = 5^3 = 125$ rows and three columns. If there are more than three historical datasets, the dimension of v can be reduced to limit the size of the grid. A large grid will increase runtime. If some of the coefficients are not estimable in the polynomial regression, the algorithm will product the error message, "some coefficients not defined because of singularities." To resolve the issue, the user can try increasing or decreasing the number of rows in the grid. Other possible causes include insufficient sample size of the historical data, insufficient number of iterations for the slice sampler, and near-zero grid values.

When a_0 is modeled as random, slice sampling is used for a_0 only for normal data, and the length of `lower.limits`, `upper.limits` and `slice.widths` should be equal to the dimension of a_0. For all other data types, slice sampling is used for β and a_0, and the length of those vectors should be equal to the dimension of β plus the dimension of a_0.

For sample size determination, `power.glm.fixed.a0` and `power.glm.random.a0` compute the power or the type I error given the total sample size (`data.size`) for the new study and other inputs. If historical datasets are provided, the algorithm samples with replacement from the historical covariates to construct the simulated datasets. Otherwise, the algorithm samples with replacement from `x.samples`. One of the arguments `historical` and `x.samples` must be provided. The argument `samp.prior.beta` contains a matrix of samples for β, which is a discrete approximation

of the sampling prior. For normal data, the argument `samp.prior.var` containing samples for τ^{-1} must also be provided. The average posterior means of the parameters are also returned.

Sampling priors

Our implementation in **BayesPPD** does not assume any particular distribution for the sampling priors. The user specifies discrete approximations of the sampling priors by providing a vector or a matrix of sample values and the algorithm samples with replacement from the vector or the matrix as the first step of data generation. For two group cases, the user simply specifies `samp.prior.mu.t` and `samp.prior.mu.c` which are vectors of samples for μ_t and μ_c. For normal data, arguments `samp.prior.var.t` and `samp.prior.var.c`, which contain samples for τ_t^{-1} and τ_c^{-1}, must also be provided. The second application example below demonstrates the use of point mass sampling priors for binary data.

For GLM cases, the user specifies `samp.prior.beta`, a matrix of samples for β. For normal data, the argument `samp.prior.var` containing samples for τ^{-1} must also be provided. For example, suppose one wants to compute the power for the hypotheses

$$H_0 : \beta_1 \geq 0$$

and

$$H_1 : \beta_1 < 0.$$

To approximate the sampling prior for β_1, one can simply sample from a truncated normal distribution with negative mean, so that the mass of the prior falls in the alternative space. Conversely, to compute the type I error rate, one can sample from a truncated normal distribution with positive mean, so that the mass of the prior falls in the null space. Next, to generate the sampling prior for the other parameters $(\beta_0, \beta_2, \cdots, \beta_p)$, one can use the posterior samples given the historical data as the discrete approximation to the sampling prior. The function `glm.fixed.a0` generates such posterior samples if the `current` argument is set to `FALSE` and $a_{0k} = 1$ for $k = 1, \cdots, K$. The second application example in this article illustrates this method for binary data with covariates. Psioda and Ibrahim (2018) discusses sampling prior elicitation in detail.

Approximation for GLMs

Because running `power.glm.fixed.a0` and `power.glm.random.a0` is potentially time-consuming, an approximation method based on asymptotic theory (Ibrahim et al., 2015) has been implemented for the model with fixed a_0. In order to attain the exact sample size needed for the desired power, the user can start with the approximation to get a rough estimate of the sample size required, using `power.glm.fixed.a0` with `approximate=TRUE`. The second application example below illustrates the use of the approximation method. For normal data, the closed form of the distribution of the MLE of β is derived and used to compute power. For other types of data, the Newton-Raphson algorithm is used. Only canonical links are allowed.

Examples

Design of a non-inferiority trial for medical devices

We first consider the non-inferiority design application of Chen et al. (2011) considering a model for binary outcomes for treatment and control groups with no covariates. The goal of that application was to design a trial to evaluate a new generation of drug-eluting stent (DES) ("test device") with the first generation of DES ("control device"). The primary endpoint is the 12-month Target Lesion Failure (TLF), defined as any of ischemia-driven revascularization of the target lesion (TLR), myocardial infarction (MI) (Q-wave and non-Q-wave) related to the target vessel, or (cardiac) death related to the target vessel. Historical information can be borrowed from two previously conducted trials involving the first generation of DES. Table 2 summarizes the historical data.

We will illustrate Bayesian SSD incorporating historical data using the power prior with fixed a_0 and the normalized power for a_0 modeled as random. Let $\mathbf{y}_t^{(n_t)} = (y_{t1}, \cdots, y_{tn_t})$ and $\mathbf{y}_c^{(n_c)} = (y_{c1}, \cdots, y_{cn_c})$ denote the responses from the current trial for the test device and the control device, respectively. The total sample size is $n = n_t + n_c$. We assume the i-th observation from the test group y_{ti} follows Bern(μ_t), and the i-th observation from the control group y_{ci} follows Bern(μ_c). Note that the notation used in our package is different from the notation used in Chen et al. (2011),

	12-Month TLF
	% TLF (# of failure/n_{0k})
Historical Trial 1	8.2% (44/535)
Historical Trial 2	10.9% (33/304)

Table 2: Summary of two historical trials involving the first generation of DES. The primary endpoint is the 12-month Target Lesion Failure (TLF). The pooled proportion for the two historical control datasets is 9.2%.

which assumes y_{ti} follows Bern(p_t) and $\mu_t = \log\left(\frac{p_t}{1-p_t}\right)$. The hypotheses for non-inferiority testing are

$$H_0 : \mu_t - \mu_c \geq \delta$$

and

$$H_1 : \mu_t - \mu_c < \delta,$$

where δ is a prespecified non-inferiority margin. We set $\delta = 4.1\%$. We choose beta($10^{-4}, 10^{-4}$) for the initial prior for μ_c, which performs similarly to the uniform improper initial prior for $\log\left(\frac{\mu_c}{1-\mu_c}\right)$ used in Chen et al. (2011) in terms of operating characteristics. We compute the Bayesian power and type I error defined in (1). Power is computed under the assumption that $\mu_t = \mu_c$ and type I error rate is computed under the assumption that $\mu_t = \mu_c + \delta$. For sampling priors, a point mass prior at $\mu_c = 9.2\%$ is used for $\pi^{(s)}(\mu_c)$ where 9.2% is the pooled proportion for the two historical control datasets, and a point mass prior at $\mu_t = \mu_c$ is used for $\pi^{(s)}(\mu_t)$. For all computations, we use $\frac{n_t}{n_c} = 3$, $N = 10,000$, and $\gamma = 0.95$, where N is the number of simulated trials and γ is a prespecified posterior probability threshold for rejecting the null hypothesis. For this example, we consider $n_t = 750$ and $a_{01} = a_{02} = 0.3$. Power can be calculated with following code in **BayesPPD**. The `historical` matrix is defined where each row represents a historical dataset, and the three columns represent the sum of responses, sample size and a_0, respectively, of the historical control data. Since point mass sampling priors are used for μ_t and μ_c, `samp.prior.mu.t` and `samp.prior.mu.c` are both scalars. For Bernoulli outcomes, beta initial priors are used for μ_t and μ_c, with hyperparameters specified by `prior.mu.t.shape1`, `prior.mu.t.shape2`, `prior.mu.c.shape1` and `prior.mu.c.shape2`.

```
historical <- matrix(0, ncol=3, nrow=2)
historical[1,] <- c(44, 535, 0.3)
historical[2,] <- c(33, 304, 0.3)

set.seed(1)
power <- power.two.grp.fixed.a0(data.type="Bernoulli",
      n.t=750, n.c=round(750/3), historical=historical,
      samp.prior.mu.t=0.092, samp.prior.mu.c=0.092,
      prior.mu.t.shape1=0.0001, prior.mu.t.shape2=0.0001,
      prior.mu.c.shape1=0.0001,prior.mu.c.shape2=0.0001,
      delta=0.041, N=10000)
power$power/type I error
[1] 0.8428
```

When a_0 is random, the normalized power prior is used and the priors for a_{01} and a_{02} are beta(1,1), as in Chen et al. (2011). We use the default settings for the upper limits, lower limits and slice widths for a_{01} and a_{02}. We run 20,000 iterations of the slice sampler. The same initial priors and sampling priors are used as in the fixed a_0 case. The code is shown below for $n_t = 750$.

```
historical <- matrix(0, ncol=2, nrow=2)
historical[1,] <- c(44, 535)
historical[2,] <- c(33, 304)

set.seed(1)
power <- power.two.grp.random.a0(data.type="Bernoulli",
      n.t=750, n.c=round(750/3),historical=historical,
      samp.prior.mu.t=0.092, samp.prior.mu.c=0.092,
      prior.mu.t.shape1=0.0001, prior.mu.t.shape2=0.0001,
      prior.mu.c.shape1=0.0001,prior.mu.c.shape2=0.0001,
```

Total sample size		1000	1080	1200	1280	1480
n_t		750	810	900	960	1110
n_c		250	270	300	320	370
		Power				
$a_0 = (0.3, 0.3)$	**BayesPPD**	0.843	0.858	0.889	0.898	0.924
	Chen et al. (2011)	0.840	0.856	0.884	0.892	0.923
Random a_0	**BayesPPD**	0.864	0.885	0.909	0.921	0.937
	Chen et al. (2011)	0.843	0.878	0.897	0.902	0.914
		Type I Error Rate				
$a_0 = (0.3, 0.3)$	**BayesPPD**	0.030	0.027	0.032	0.030	0.032
	Chen et al. (2011)	0.030	0.027	0.028	0.030	0.032
Random a_0	**BayesPPD**	0.032	0.027	0.031	0.031	0.031
	Chen et al. (2011)	0.038	0.031	0.029	0.036	0.039

Table 3: Power and type I error rate comparisons for Chen et al. (2011) and **BayesPPD** for a few different sample sizes. We use $N = 10,000$ simulated trials. Two models are considered, a power prior with a_0 fixed at 0.3 for both historical trials and the normalized power prior. We observe that the results provided by **BayesPPD** closely match the results provided in Chen et al. (2011).

```
      prior.a0.shape1=1,prior.a0.shape2=1,
      delta=0.041, gamma=0.95,
      nMC=20000, nBI=250, N=10000)
power$`power/type I error`
[1] 0.864
```

Table 3 compares power calculations from Chen et al. (2011) and **BayesPPD** for a few different sample sizes. We observe that the results provided by **BayesPPD** closely match the results provided in Chen et al. (2011).

Study of acquired immunodeficiency syndrome (AIDS)

Using data from two trials that study the effect of Zidovudine on AIDS, ACTG019 and ACTG036, we will demonstrate how **BayesPPD** can be used for coefficient estimation as well as power and type I error rate calculation for generalized linear models in designs that incorporate historical data.

Zidovudine (AZT) is an inhibitor of the replication of the human immunodeficiency virus (HIV). The ACTG019 study was a double-blind placebo-controlled clinical trial comparing AZT with a placebo in adults with asymptomatic HIV who had CD4 cell counts of fewer than 500 per cubic millimeter. The results were published in Volberding et al. (1990). For this example, we use only the control group data from ACTG019. The binary primary endpoint is death or development of AIDS or AIDS-related complex (ARC). We consider four of the measured covariates used, CD4 cell count (x01) (cell count per cubic millimetre of serum), age (x02), treatment (x03) and race (x04). The covariates CD4 cell count and age are continuous, while the others are binary. The ACTG036 study was also a placebo-controlled clinical trial comparing AZT with a placebo in asymptomatic patients with hereditary coagulation disorders and HIV infection. The results were published in Merigen et al (1991). The endpoint and covariates used are the same as those in the ACTG019 trial. Table 4 summarizes the endpoint and covariates for the two studies.

First, we standardize age for ease of interpretation and take the log of CD4 cell count count.

```
data(actg019)
data(actg036)
Y0 <- actg019$outcome
X0 <- actg019[,-1]
X0$age_std <- scale(X0$age)
X0$T4_log <- log(X0$T4count)
```

	ACTG019 (control group)	ACTG036
No. of patients	404	183
AZT treatment, n (%)	NA	89 (48.6)
CD4 cell count, mean (SD)	332.5 (109.3)	297.7 (130.5)
Age, y; mean (SD)	34.5 (7.7)	30.4(11.2)
White race, n (%)	377 (93.3)	166 (90.7)
Death or ARC, n (%)	36 (8.9)	11 (6.0)

Table 4: Summary of the ACTG019 trial (control group) and the ACTG036 trial data. The binary primary endpoint is death or development of ARC. The sample size of the ACTG019 trial is much larger than the ACTG036 trial. The covariate and endpoint summaries are comparable for the two datasets.

	$a_0 = 0$		$a_0 = 0.5$		$a_0 = 1$		$a_0 \sim$ beta(1,1)	
Intercept	9.14	[3.83; 16.34]	4.89	[1.24; 8.27]	3.95	[0.94; 6.98]	4.39	[1.41; 7.54]
AZT	−0.15	[−1.80; 1.42]	−0.95	[−2.16; 0.25]	−1.00	[−2.12; 0.14]	−0.96	[−2.14; 0.08]
Age (standardized)	0.32	[−0.42; 1.04]	0.36	[−0.01; 0.74]	0.38	[0.11; 0.68]	0.38	[0.06; 0.67]
Race	0.36	[−2.35; 3.23]	0.72	[−1.10; 2.75]	0.93	[−0.83; 3.05]	0.73	[−0.86; 2.44]
log(CD4)	−2.42	[−3.61; −1.35]	−1.48	[−2.04; −0.89]	−1.32	[−1.78; −0.84]	−1.37	[−1.91; −0.86]

Table 5: Posterior mean and 95% credible interval for β incorporating historical data for the four priors, a_0 fixed at 0, 0.5, and 1 and a_0 modeled as random with a beta$(1, 1)$ prior. There is evidence suggesting a negative association between AZT and death but the evidence is not substantial by common criteria (e.g., posterior probability > 0.95).

```
X0 <- as.matrix(X0[,c("age_std","race","T4_log")])

Y <- actg036$outcome
X <- actg036[,-1]
X$age_std <- scale(X$age)
X$T4_log <- log(X$T4count)
X <- as.matrix(X[,c("treat","age_std","race","T4_log")])
```

Suppose we are interested in analyzing the relationship between the outcome and the covariates after incorporating historical information. The code below demonstrates the analysis based on a power prior with a_0 fixed at 0.5 and using only the ACTG019 study data as prior information.

```
set.seed(1)
historical <- list(list(y0=Y0, x0=X0, a0=0.5))
result <- glm.fixed.a0(data.type="Bernoulli", data.link="Logistic", y=Y, x=X,
+                       historical=historical, nMC=10000, nBI=250)
colMeans(result$posterior.samples)
[1]  4.8931870 -0.9459501  0.3645510  0.7201122 -1.4784046
```

Table 5 displays the posterior mean and 95% credible interval for β for four different priors, a_0 fixed at 0, 0.5, and 1 and a_0 modeled as random with a beta$(1, 1)$ prior. There is evidence suggesting a negative association between AZT and death but the evidence is not substantial by common criteria (e.g., posterior probability > 0.95).

For this example we consider designing a new clinical trial that is similar to the historical trial, ACTG019. We hope to acquire a range of sample sizes that can achieve powers around 0.8 to test the hypotheses

$$H_0 : \beta_1 \geq 0$$

and

$$H_1 : \beta_1 < 0$$

based on the chosen sampling priors. Here, β_1 represents the treatment effect of AZT. First, we generate the input for `samp.prior.beta`, a matrix of samples for β representing a discrete

approximation of the sampling prior. For β_1, we sample from a truncated normal distribution with mean -0.5, which is our guess of the effect size of AZT. The distribution is truncated to avoid extreme, implausible values for β_1. For the other parameters, the sampling prior is fixed at the posterior mean of the parameter given the historical data, which can be easily obtained using `glm.fixed.a0` with `current=FALSE`. We then combine the sampling prior for β_1 and the other parameters into a matrix, as follows:

```
library(truncnorm)
set.seed(1)
historical.sp <- list(list(y0=Y0, x0=X0, a0=1))
result <- glm.fixed.a0(data.type="Bernoulli", data.link="Logistic",
                       historical=historical.sp,
                       nMC=10000, nBI=250, current.data = FALSE)
beta.sp <- result$posterior.samples
nSP <- 10000
mat.sp <- matrix(rep(colMeans(beta.sp), each=nSP), nrow=nSP)
beta1.sp <- rtruncnorm(nSP, a=-2, b=-0.1, mean=-0.5)
samp.prior.beta <- cbind(mat.sp[,1], beta1.sp, mat.sp[,2:4])
```

Next, we use `power.glm.fixed.a0` with `approximate=TRUE` to obtain a rough estimate of the sample size required to achieve a power of 0.8. The code below experiments with sample sizes 800, 1000 and 1200. We observe that to reach a power of 0.8, the sample size should be approximately 800 when a_0 is fixed at 0.5.

```
set.seed(1)
sample.sizes <- c(800,1000,1200)
historical <- list(list(y0=Y0, x0=X0, a0=0.5))
results <- NULL
for(i in 1:length(sample.sizes)){
    result <- power.glm.fixed.a0(data.type="Bernoulli", data.size=sample.sizes[i],
                                 historical=historical,
                                 samp.prior.beta=samp.prior.beta,
                                 delta=0, gamma=0.95, approximate=TRUE, N=10000)
    results <- c(results, result$`power/type I error`)
}
results
[1] 0.8037 0.8177 0.8391
```

Finally, we calculate the exact power using the normalized power prior with a_0 modeled as random. The `normalizing.constant` function provides the value for `a0.coefficients` of `power.glm.random.a0`. Since there is only one historical dataset, the `grid` is simply a matrix with one column. The code below demonstrates the usage when sample size is 800. We run $25,000$ iterations of the slice sampler for each of the $10,000$ simulated datasets. The corresponding power is 0.7936. Power curves for the four different priors for sample sizes ranging from 750 to 1200 are plotted in Figure 1. The underlying estimated power values are displayed in Table 6 in the Appendix.

```
grid <- matrix(seq(0.05,1,by=0.1))
historical <- list(list(y0=Y0, x0=X0))
a0_coef <- normalizing.constant(grid=grid, historical=historical,
                                data.type="Bernoulli",data.link="Logistic")
result <- power.glm.random.a0(data.type="Bernoulli",data.link="Logistic",
                              data.size=800, historical=historical,
                              samp.prior.beta=samp.prior.beta,
                              a0.coefficients = a0_coef,
                              delta=0, nMC=25000, nBI=250, N=10000)
result$`power/type I error`
[1]  0.7936
```

Discussion

BayesPPD facilitates Bayesian sample size determination by providing a robust suite of functions for power calculation and analysis using the power and normalized power priors for generalized linear models. A major contribution of this package is the ability to handle covariates for Bernoulli, normal, Poisson and exponential outcomes. Despite the use of MCMC algorithms for analysis and design

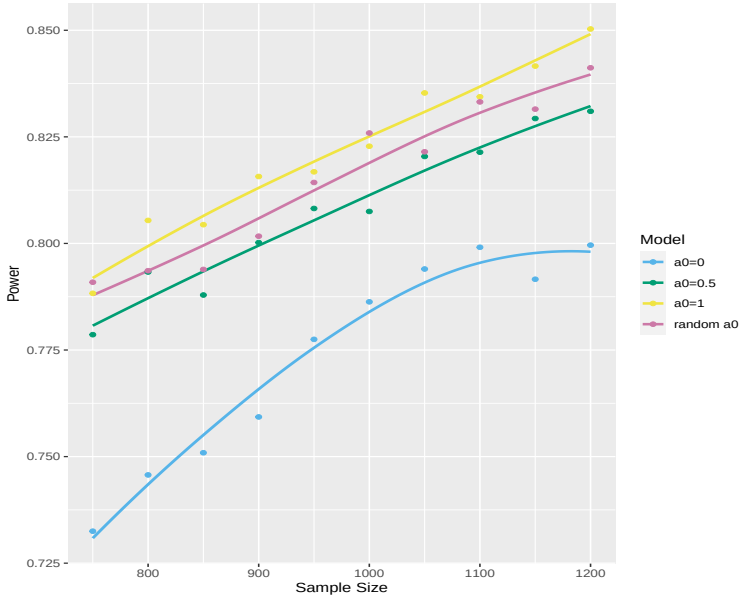

Figure 1: Power curves for a range of sample sizes for the four priors with a_0 fixed at 0 (blue line), 0.5 (green line), 1 (yellow line) and a_0 modeled as random (pink line). The dots represent the power for a particular sample size. LOESS curves have been fitted to the point estimates. We observe that for fixed a_0, the power is higher for higher values of a_0, as more historical information is borrowed. When a_0 is modeled as random, the power curve is higher than the curve with $a_0 = 0.5$ but lower than the curve with $a_0 = 1$.

simulations, **BayesPPD** is computationally efficient, with functions producing results in seconds for many application settings.

A possible extension of the package is the accommodation for longitudinal and time-to-event outcomes. Another potential feature is computing optimal hyperparameters for the beta prior on a_0 to ensure certain characteristics are met, such as the ability to adapt to prior-data conflict or prior-data agreement. The method will be based on ongoing theoretical work by the authors.

Bibliography

S. Balcome, D. Musgrove, T. Haddad, and G. L. Hickey. *bayesDP: Tools for the Bayesian Discount Prior Function*, 2021. URL https://CRAN.R-project.org/package=bayesDP. R package version 1.3.4. [p336]

L. M. Carvalho and J. G. Ibrahim. On the normalized power prior. *Statistics in Medicine*, 40(24): 5251–5275, Jul 2021. doi: 10.1002/sim.9124. [p335]

T. Chandereng, D. Musgrove, T. Haddad, G. Hickey, T. Hanson, and T. Lystig. *bayesCT: Simulation and Analysis of Adaptive Bayesian Clinical Trials*, 2020. URL https://CRAN.R-project.org/package=bayesCT. R package version 0.99.3. [p336]

M.-H. Chen and J. G. Ibrahim. Power prior distributions for regression models. *Statistical Science*, 15(1):46–60, feb 2000. doi: 10.1214/ss/1009212673. [p335, 336, 337]

M.-H. Chen, J. G. Ibrahim, P. Lam, A. Yu, and Y. Zhang. Bayesian design of noninferiority trials for medical devices using historical data. *Biometrics*, 67(3):1163–1170, Sep 2011. doi: 10.1111/j.1541-0420.2011.01561.x. [p335, 338, 343, 344, 345]

Cytel Software Corporation. *East. Software for Design Simulation and Interim Monitoring of Flexible Clinical Trials*. Cambridge, MA, 2014. URL http://www.cytel.com/software-solutions/east/. [p336]

F. De Santis. Using historical data for bayesian sample size determination. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(1):95–113, Jan 2007. doi: 10.1111/j.1467-985X.2006.00438.x. [p335]

Y. Duan, K. Ye, and E. P. Smith. Evaluating water quality using power priors to incorporate historical information. *Environmetrics (London, Ont.)*, 17(1):95–106, feb 2006. doi: 10.1002/env.752. [p335, 337]

D. Eddelbuettel and R. Francois. Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8):1–18, 2011. [p336]

B. Eggleston, D. Wilson, B. McNeil, J. Ibrahim, and D. Catellier. *BayesCTDesign: Two Arm Bayesian Clinical Trial Design with and Without Historical Control Data*, 2019. URL https://CRAN.R-project.org/package=BayesCTDesign. R package version 0.6.0. [p336]

F. Gerber and T. Gsponer. gsbDesign: An R package for evaluating the operating characteristics of a group sequential Bayesian design. *Journal of Statistical Software*, 69(11):1–23, 2016. doi: 10.18637/jss.v069.i11. [p336]

Z. Han, T. Bai, and K. Ye. *NPP: Normalized Power Prior Bayesian Analysis*, 2021. URL https://CRAN.R-project.org/package=NPP. R package version 0.4.0. [p336]

B. P. Hobbs, B. P. Carlin, S. J. Mandrekar, and D. J. Sargent. Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics*, 67 (3):1047–1056, Sep 2011. doi: 10.1111/j.1541-0420.2011.01564.x. [p335]

J. G. Ibrahim, M.-H. Chen, and D. Sinha. On optimality properties of the power prior. *Journal of the American Statistical Association*, 98(461):204–213, mar 2003. doi: 10.1198/016214503388619229. [p335]

J. G. Ibrahim, M.-H. Chen, Y. Gwon, and F. Chen. The power prior: theory and applications. *Statistics in Medicine*, 34(28):3724–3749, dec 2015. doi: 10.1002/sim.6728. [p335, 336, 343]

L. Joseph, C. E. M'Lan, and D. B. Wolfson. Bayesian sample size determination for binomial proportions. *Bayesian Analysis*, 3(2), Jun 2008. doi: 10.1214/08-BA310. [p335]

A. Kopp-Schneider, M. Wiesenfarth, W. Ruth, D. Edelmann, O. Witt, and U. Abel. Monitoring futility and efficacy in phase ii trials with bayesian posterior distributions - a calibration approach. *Biometrical Journal*, to appear, 2018. [p336]

LLC Consultants. *FACTS: Fixed and Adaptive Clinical Trial Simulator*. Austin, TC, 2014. URL http://www.berryconsultants.com/software/. [p336]

C. E. M'Lan, L. Joseph, and D. B. Wolfson. Bayesian sample size determination for case-control studies. *Journal of the American Statistical Association*, 101(474):760–772, Jun 2006. doi: 10.1198/016214505000001023. [p335]

K. Nagashima. *ph2bayes: Bayesian Single-Arm Phase II Designs*, 2018. URL https://CRAN.R-project.org/package=ph2bayes. R package version 0.0.2. [p336]

R. M. Neal. Slice sampling. *Annals of Statistics*, 31(3):705–767, 2003. [p338, 341]

B. Neuenschwander, M. Branson, and D. J. Spiegelhalter. A note on the power prior. *Statistics in Medicine*, 28:3562–3566, 2009. [p337]

B. Neuenschwander, G. Capkun-Niggli, M. Branson, and D. J. Spiegelhalter. Summarizing historical information on controls in clinical trials. *Clinical Trials*, 7(1):5–18, Feb 2010. doi: 10.1177/1740774509356002. [p335]

S. J. Pocock. The combination of randomized and historical controls in clinical trials. *Journal of chronic diseases*, 29(3):175–188, Mar 1976. doi: 10.1016/0021-9681(76)90044-8. [p335]

M. A. Psioda and J. G. Ibrahim. Bayesian design of a survival trial with a cured fraction using historical data. *Statistics in Medicine*, 37(26):3814–3831, Nov 2018. doi: 10.1002/sim.7846. [p336, 339, 343]

M. A. Psioda and J. G. Ibrahim. Bayesian clinical trial design using historical data that inform the treatment effect. *Biostatistics*, 20(3):400–415, Jul 2019. doi: 10.1093/biostatistics/kxy009. [p335]

Y. Qiu, S. Balan, M. Beall, M. Sauder, N. Okazaki, and T. Hahn. *RcppNumerical: 'Rcpp' Integration for Numerical Computing Libraries*, 2019. URL https://CRAN.R-project.org/package=RcppNumerical. R package version 0.4-0. [p341]

E. Rahme and L. Joseph. Exact sample size determination for binomial experiments. *Journal of Statistical Planning and Inference*, 66:83–93, 1998. [p335]

H. Schmidli, S. Gsteiger, S. Roychoudhury, A. O'Hagan, D. Spiegelhalter, and B. Neuenschwander. Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics*, 70(4):1023–1032, Dec 2014. doi: 10.1111/biom.12242. [p335]

Y. Shen, M. A. Psioda, and J. G. Ibrahim. *BayesPPD: Bayesian Power Prior Design*, 2022. URL https://CRAN.R-project.org/package=BayesPPD. R package version 1.1.0. [p335]

R. Simon. Bayesian design and analysis of active control clinical trials. *Biometrics*, 55(2):484–487, Jun 1999. doi: 10.1111/j.0006-341x.1999.00484.x. [p335]

K. Viele, S. Berry, B. Neuenschwander, B. Amzal, F. Chen, N. Enas, B. Hobbs, J. G. Ibrahim, N. Kinnersley, S. Lindborg, and et al. Use of historical control data for assessing treatment effects in clinical trials. *Pharmaceutical Statistics*, 13(1):41–54, Feb 2014. doi: 10.1002/pst.1589. [p335]

F. Wang and A. E. Gelfand. 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, May 2002. doi: 10.1214/ss/1030550861. [p335]

Y.-B. Wang, M.-H. Chen, L. Kuo, and P. O. Lewis. A new monte carlo method for estimating marginal likelihoods. *Bayesian Analysis*, 13(2):311–333, 2018. [p335, 338, 341]

G. Wassmer and R. Eisebitt. *ADDPLAN: Adaptive Designs – Plans and Analyses*. Reston, VA, 2005. URL http://www.addplan.com/. [p336]

S. Weber. *RBesT: R Bayesian Evidence Synthesis Tools*, 2021. URL https://CRAN.R-project.org/package=RBesT. R package version 1.6-2. [p336]

1 Additional tables

Sample size	$a_0 = 0$	$a_0 = 0.5$	$a_0 = 1$	Random a_0
750	0.732	0.779	0.788	0.791
800	0.746	0.793	0.805	0.794
850	0.751	0.788	0.804	0.794
900	0.759	0.800	0.816	0.802
950	0.778	0.808	0.817	0.814
1000	0.786	0.807	0.823	0.826
1050	0.794	0.820	0.835	0.822
1100	0.799	0.821	0.834	0.833
1150	0.792	0.829	0.842	0.832
1200	0.800	0.831	0.850	0.841

Table 6: Power for the four priors of the AIDS study.

Yueqi Shen
University of North Carolina at Chapel Hill
Department of Biostatistics
United States
ys137@live.unc.edu

Matthew A. Psioda
University of North Carolina at Chapel Hill
Department of Biostatistics
United States
matt_psioda@unc.edu

Joseph G. Ibrahim
University of North Carolina at Chapel Hill
Department of Biostatistics
United States
ibrahim@bios.unc.edu