INSTITUTE FOR DEFENSE ANALYSES

Power Approximations for Generalized Linear Models using the Signal-to-Noise Transformation Method

Thomas H. Johnson
Laura Freeman
Jim Simpson
Colin Anderson

February 2017
Approved for public release.
IDA Document NS D-8351
Log: H 2017-000112

INSTITUTE FOR DEFENSE ANALYSES
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

About this Publication
Statistical power is a useful measure for assessing the adequacy of an experimental design prior to data collection. This paper proposes an approach referred to as the signal-to-noise transformation method (SNRx), to approximate power for effects in a generalized linear model. The contribution of SNRx is that, with a couple assumptions, it generates power approximations for generalized linear model effects using F-tests that are typically used in ANOVA for classical linear models. Additionally, SNRx follows Ohlert and Whitcomb's unified approach for sizing an effect, which allows for intuitive effect size definitions, and consistent estimates of power. This paper details the process for defining an effect size, constructing the coefficients for the test, and calculating power for the family of generalized linear models. The focus is on experimental designs that have multi-level categorical factors. A simulation study is performed, which demonstrates that SNRx power results agree with simulation.

For More Information:
Laura Freeman, Project Leader
lfreeman@ida.org • (703) 845-2084

Robert R. Soule, Director, Operational Evaluation Division
rsoule@ida.org • (703) 845-2482

I N S T I T U T E F O R D E F E N S E A N A L Y S E S

IDA Document NS D-8351

Power Approximations for Generalized Linear Models using the Signal-to-Noise Transformation Method

Thomas H. Johnson
Laura Freeman
Jim Simpson
Colin Anderson

Power Approximations for Generalized Linear Models using the Signal-to-Noise Transformation Method

Thomas H Johnson, Laura Freeman, Jim Simpson, and Colin Anderson

Institute for Defense Analyses

4850 Mark Center Drive

Alexandria, VA 22311-1882

Abstract

Statistical power is a useful measure for assessing the adequacy of an experimental design prior to data collection. In an experimental design, power is the probability of correctly concluding a factor or interaction effect in a model is significant. For a fixed model, power increases with sample size, making it a useful measure for determining the scope of a test prior to data collection. For normally distributed response variables, power calculations are widely available in experimental design software. However, many practical applications result in non-normal responses. Generalized linear models provide many useful analysis methods for non-normal responses. While statistical software routinely includes generalized linear models in analysis packages, power calculations for generalized linear models are not widely available in experimental design modules. This paper proposes a signal-to-noise transformation method (SNRx) that enables generalized linear model power approximations using normal linear model power calculations, making them generally available to all practitioners. This paper details the process for defining an effect size, constructing the coefficients for the test, and calculating power for the family of generalized linear models. A simulation study demonstrates that SNRx power results agree with Monte Carlo simulation.

1. Introduction

Experimental designs are used across a variety of fields to aide in the planning, execution, and analysis of an experiment. In the planning phase we determine the test objectives. These objectives guide the development of

Preprint submitted to Journal Name *May 31, 2017*

the factors, levels, and response variables [1]. In the Department of Defense, recent policy has emphasized the importance of using Design of Experiments in all operational testing[2].

Equally important in the planning phase is the assessment of the experimental design. An assortment of measures are available to assess the goodness of an experiment prior to data collection. Hahn, Meeker, and Feder call these "measures of precision"[3]. These measures include standard error of predicted mean responses, standard error of coefficients, correlations metrics, and optimality criteria values. Measures of precision are affected by many aspects of the plan, including the choice of factors and levels, the assumed model form, the combination of factors settings from run to run, and the total number of runs in the experiment.

Effect power is an important measure of precision and the focus of this paper. Effect power is the probability of concluding that an effect impacts the response variable when it is truly active. In general, the power of an effect increases with sample size, making it a useful measure for determining the scope of a test prior to data collection. Here we focus on a second-order model for designs with multi-level categorical factors. Effects considered include the main effects and two-factor interactions [4].

In many Department of Defense tests the response variables are not normally distributed, thus classical linear modeling approaches do not apply. In these cases, generalized linear models provide a viable alternative. Illustrative applications include modeling probability of detection (logistic regression), number of enemies defeated (Poisson regression), and failure rates (gamma or exponential regression).

Experimental design software that calculates power for classical linear models are widely available. However, when it is known before running the experiment that the response will not be normally distributed, the power calculations should reflect that knowledge. Software that calculates power for experimental designs with generalized linear models often are not widely available in commercial software; such calculations usually require Monte Carlo simulation studies. It is important to take into account the knowledge of the planned analysis when planning the test because different distributions can require dramatically different sample sizes to achieve high effect power.

The goal of this paper is to provide a simple method to obtain power for a generalized linear model by "transforming" the effect size in the power calculation for a classical linear model. Existing software (e.g., JMP, Minitab, and Design Expert) that accommodate classical linear model power calculations allow the user to adjust the signal-to-noise ratio or alter the model coefficients under the alternative hypothesis. SNRx provides a means of setting the signal-to-noise ratio or the coefficients so that the calculation represents the generalized linear model power calculation. The target audience of SNRx is the analyst who has statistical design experience and is comfortable working with popular statistical software, but who is not inclined to calculate power for generalized linear models using custom code and Monte Carlo simulation.

Research on generalized linear model power approximations is abundant in the literature. Methods that apply to a single type of generalized model include work by Whittemore [5], Signorini [6], O'Brien [7], Sheih [8], and Dimedvenko [9]. These contributions focus on conducting a hypothesis test on an individual model parameter.

Power approximation methods that generally apply to the family of generalized linear models include work by Self and Mauritsen [10], Self, Mauritsen, and O'hara [11], and Sheih [12]. These universal methods work within a generalized linear model framework and use the score, likelihood ratio, and Wald test statistic, respectively, and accommodate composite hypothesis tests. Additional work by Newson presents a generalized power calculation approach based on the central-limit theorem applied to influence functions [13]. These approaches are more rigorous than SNRx, and more flexible, but can be difficult to understand and apply.

The remainder of the paper is organized as follows. Section 2 reviews the classical linear and generalized linear model forms. Section 3 defines the effect size for multi-level categorical factors. Section 4 proposes the SNRx method. Section 5 presents three examples. Section 6 presents a simulation study that compares SNRx power with power estimates from simulation. Section 7 provides a discussion.

2. Model Formulation

SNRx provides a simple method to obtain power for a generalized linear model by "transforming" the effect size in the power calculation for a classical linear model. In this section we clarify the distinction between a classical linear model and a generalized linear model.

2.1. Classical Linear Model

In introductory Design of Experiments textbooks, the fundamental statistical model, based on the normal distribution, is referred to as a classical linear model. This model has the form $\boldsymbol{Y} = \boldsymbol{M}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\boldsymbol{I})$. The model matrix $\boldsymbol{M}$ is size $n \times r$, where n is the number of runs, r is the number of coefficients in the model, $\boldsymbol{Y}$ is the response vector of size $n \times 1$, $\boldsymbol{\beta}$ is the coefficient vector of size $r \times 1$, and $\boldsymbol{\epsilon}$ is a vector of errors of size $n \times 1$ that are independent and identically distributed.

The coefficient vector, $\boldsymbol{\beta}$, can be split into two vectors to accommodate inference about a subset of the model. In splitting the coefficient vector, we borrow the nomenclature used by Self and Mauritsen [10]. The vector, $\boldsymbol{\psi}$, is size $p \times 1$ and contains the coefficients to be tested. The other vector, referred to as the nuisance parameters, $\boldsymbol{\lambda}$, is size $(r-p) \times 1$ and contains the remainder of the coefficients.

In a similar fashion, the model matrix, $\boldsymbol{M}$, can be partitioned into the corresponding test matrix, $\boldsymbol{Z}$, and nuisance matrix, $\boldsymbol{X}$, so that the linear predictors of the full model can be written as $\boldsymbol{Z}\boldsymbol{\psi} + \boldsymbol{X}\boldsymbol{\lambda}$. The same partitioning of the coefficient vector and model matrix also holds for the generalized linear model.

2.2. Generalized Linear Model

A generalized linear model is called so because it generalizes the classical linear model. Developed by Nelder and Wedderburn in 1972, generalized linear models include as special cases linear regression, logistic regression, log-linear models for count data, and models for survival data. A generalized linear model is defined in terms of its three components:

1. **Random component**. A generalized linear model has response variables, $Y_1, ..., Y_n$, that share the same distribution from the exponential dispersion family[14], where the vth response of the experiment has an expected value equal to the mean μ_v.
2. **Systematic component**. The unknown coefficients systematically specify the linear predictor η_v such that $\eta_v = \boldsymbol{Z}_v\boldsymbol{\psi} + \boldsymbol{X}_v\boldsymbol{\lambda}$, where $\boldsymbol{Z}_v$ and $\boldsymbol{X}_v$ represent the vth row of the test and nuisance matrix.
3. **The link between the random and systematic components**. The link function $g(\cdot)$ relates the mean and linear predictor in the expression $g(\mu_v) = \eta_v$. The link function is monotonic and invertible.

2.3. Model Inference

We are interested in the hypothesis that tests for the significance of a multi-level categorical factor or interaction between multi-level categorical factors. Specifically, whether the coefficients belonging to a main effect or two-factor interaction effect are equal to zero. Thus, the hypothesis test for an individual effect is

$$\begin{aligned} H_0 &: \boldsymbol{\psi} = \mathbf{0} \\ H_1 &: \boldsymbol{\psi} \neq \mathbf{0} \quad . \end{aligned} \tag{1}$$

The classical and generalized linear models use similar techniques for evaluating these hypothesis tests. A classical linear model uses analysis of variance (ANOVA), which is based on an F statistic. The analogue of an ANOVA for generalized linear models is an analysis of deviance, which is based on a likelihood ratio statistic [15].

Some classical linear model software allows the user to specify the details of a planned experiment, and the software outputs the power associated with the effect tests. More specifically, the user can input the design matrix, choose the model form, set the anticipated coefficients $\boldsymbol{\psi}$ under H_1, and obtain power.

The SNRx method is useful in situations where the practitioner only has access to classical linear model software, but is interested in calculating power for a specific generalized linear model. In this situation, the SNRx method sets $\boldsymbol{\psi}$ under H_1 so that the ANOVA hypothesis test well represents an analysis of deviance for the specific generalized linear model.

3. Sizing an Effect

In this paper we adopt Ohlert and Whitcomb's approach for sizing an effect. Among the steps to calculate power for the hypothesis test that we introduced in the previous section, a practitioner sizes an effect by setting $\boldsymbol{\psi}$ under H_1. Because there are nearly countless ways a practitioner could set $\boldsymbol{\psi}$, a primary advantage of Ohlert and Whitcomb's approach is that it provides a consistent way of setting $\boldsymbol{\psi}$. This results in consistent power estimates that accommodate a fair comparison between competing experimental designs.

The best way to understand Ohlert and Whitcomb's approach is to start with the basic marginal means model. An intuitive way to define an effect size is to do so in terms of units of the response variable. A marginal means model provides this capability. Once an effect size is defined in terms of marginal means, that effect size can be converted into classical linear model vector of coefficients $\boldsymbol{\psi}$ for use in power calculations.

The marginal means model in Equation 2 includes two main effect marginal means (ρ_i and τ_j) with a and b levels, respectively. The overall mean is denoted by μ_0 and $\rho\tau_{ij}$ is an interaction marginal mean.

$$E(y_{ij}) = \mu_{ij} = \mu_0 + \rho_i + \tau_j + \rho\tau_{ij}, \quad i = 1, ..., a \quad , \quad j = 1, ...b \tag{2}$$

The marginal means model has additional conditions that are sometimes referred to as the "sum to zero" constraints. In Hocking's notation [16], the constraints on the main effect marginal means are:

$$\rho_a = -\sum_{i=1}^{a-1} \rho_i \quad , \quad \tau_b = -\sum_{j=1}^{b-1} \tau_j \quad . \tag{3}$$

The constraints on the interaction marginal mean ensure that the the interaction marginal mean sums to zero across any subscript:

$$\begin{aligned} \rho\tau_{ib} &= -\sum_{j=1}^{b-1} \rho\tau_{ij} \quad i = 1, ..., (a-1) \\ \rho\tau_{aj} &= -\sum_{i=1}^{a-1} \rho\tau_{ij} \quad j = 1, ..., (b-1) \\ \rho\tau_{ab} &= \sum_{i=1}^{a-1} \sum_{j=1}^{b-1} \rho\tau_{ij} \end{aligned} \tag{4}$$

The constraints imply that an a-level main effect marginal mean can be sufficiently described by $a-1$ coefficients. For the a-level main effect marginal mean (ρ_i) the corresponding coefficient vector ($\tilde{\boldsymbol{\rho}}$) is of size $(a-1) \times 1$. Let $\boldsymbol{\rho}$ represent the vector of ρ_i main effect marginal means, then the relationship between marginal means $\boldsymbol{\rho}$ and coefficients $\tilde{\boldsymbol{\rho}}$ in matrix form is:

$$\boldsymbol{\rho} = \boldsymbol{\Delta}_a^T \tilde{\boldsymbol{\rho}} \quad , \tag{5}$$

where, in Hocking's notation, the constrast matrix is $\boldsymbol{\Delta}_a = (\boldsymbol{I}_{a-1} | -\boldsymbol{J}_{a-1})$, and $\boldsymbol{I}_{a-1}$ is the identity matrix of size $(a-1) \times (a-1)$, and $\boldsymbol{J}_{a-1}$ is a vector of ones of size $(a-1) \times 1$. When $a = 3$ this relation can be written as "

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} \tilde{\rho}_1 \\ \tilde{\rho}_2 \end{bmatrix}. \tag{6}$$

With this notation, the marginal means model in equation 2 can be written in the classical linear model form as:

$$\boldsymbol{\mu} = \boldsymbol{M}\boldsymbol{\beta} \quad , \tag{7}$$

where,

$$\boldsymbol{M} = (\boldsymbol{J}_a \otimes \boldsymbol{J}_b \mid \boldsymbol{\Delta}_a^T \otimes \boldsymbol{J}_b \mid \boldsymbol{J}_a \otimes \boldsymbol{\Delta}_b^T \mid \boldsymbol{\Delta}_a^T \otimes \boldsymbol{\Delta}_b^T) \quad , \tag{8}$$

and,

$$\boldsymbol{\beta}^T = \begin{bmatrix} \lambda_{int} & \tilde{\boldsymbol{\rho}}^T & \tilde{\boldsymbol{\tau}}^T & \widetilde{\boldsymbol{\rho\tau}}^T \end{bmatrix}, \tag{9}$$

and λ_{int} is the intercept coefficient.

As an example, when $a = 3$ and $b = 2$ the marginal means model in equation 7 can be written as equation 10. In equation 10, the mean (μ_0) is said to be marginal to the main effects (ρ_i or τ_j) and the interaction ($\rho\tau_{ij}$), and the main effects are marginal to the interaction [17].

$$\begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{21} \\ \mu_{22} \\ \mu_{31} \\ \mu_{32} \end{bmatrix} = \begin{bmatrix} \mu_0 + \rho_1 + \tau_1 + \rho\tau_{11} \\ \mu_0 + \rho_1 + \tau_2 + \rho\tau_{12} \\ \mu_0 + \rho_2 + \tau_1 + \rho\tau_{21} \\ \mu_0 + \rho_2 + \tau_2 + \rho\tau_{22} \\ \mu_0 + \rho_3 + \tau_1 + \rho\tau_{31} \\ \mu_0 + \rho_3 + \tau_2 + \rho\tau_{32} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & -1 & -1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & -1 & 0 & -1 \\ 1 & -1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_{int} \\ \tilde{\rho}_1 \\ \tilde{\rho}_2 \\ \tilde{\tau}_1 \\ \widetilde{\rho\tau}_{11} \\ \widetilde{\rho\tau}_{21} \end{bmatrix}. \tag{10}$$

Defining effects in terms of marginal means is intuitive to the experimenter because they are in units of the response variable. A drawback of this approach could stem from confusion about the interpretation of marginal mean parameters. For instance, since ρ_i is marginal to $\rho\tau_{ij}$, ρ_i does not represent the overall average departure from μ_0. It represents the departure from μ_0 after the effects of $\rho\tau_{ij}$ have been removed. Despite this potential confusion, interpreting marginal means is still more intuitive than interpreting coefficients. For more in-depth discussion about the interpretation of marginal means, which are sometimes referred to as "least-squares means" or "estimated population marginal means", see references [18] and [19].

Ohlert and Whitcomb's unified approach defines the size of an effect as the change in the mean response across the design space due to that effect [20]. They define this range in multiples of σ, such that the effect size quantifies the ratio of the signal to the noise, denoted by $\kappa = \delta/\sigma$.

For the purpose of describing the general procedure for converting marginal means to coefficients, let $\boldsymbol{\mu}_\psi$ represent the vector of components of the marginal mean that corresponds to the effect under test (i.e. ρ_i or τ_j or $\rho\tau_{ij}$). That is, for the test on an a-level main effect we have $\boldsymbol{\mu}_\psi = \boldsymbol{\Delta}_a^T \boldsymbol{\psi}$, and for a test on an interaction between an a-level factor and b-level factor we have $\boldsymbol{\mu}_\psi = (\boldsymbol{\Delta}_a^T \otimes \boldsymbol{\Delta}_b^T)\boldsymbol{\psi}$. In this notation κ is the range of $\boldsymbol{\mu}_\psi$. With these relationships we can calculate the coefficients under test using equation 11. Note that the inverse in equation 11 is a generalized inverse.

$$\begin{aligned} \text{Main Effect:} \quad & \boldsymbol{\psi} = \left(\boldsymbol{\Delta}_a^T\right)^{-1} \boldsymbol{\mu}_\psi \\ \text{Two Factor Interaction:} \quad & \boldsymbol{\psi} = \left(\boldsymbol{\Delta}_a^T \otimes \boldsymbol{\Delta}_b^T\right)^{-1} \boldsymbol{\mu}_\psi \end{aligned} \tag{11}$$

The second key part of Ohlert and Whitcomb's unified approach is their principle of conservatism. They state, " report as power for a given size effect the smallest possible power among all those effects with the given size." This statement fixes a consistency issue with the previous definition that defines the effect size as the range of $\boldsymbol{\mu}_{\boldsymbol{\psi}}$. There are many possible configurations of $\boldsymbol{\mu}_{\boldsymbol{\psi}}$ that satisfy the constraints in equations 3 and 4, and have a range of κ. The principle of conservatism requires a search for the configuration that yields the smallest power.

In summary, the following steps are used to size an effect within a classical linear model. First, define κ by choosing values for δ and σ. Note that δ and σ are in units of the response. Next, construct $\boldsymbol{\mu}_\psi$ so that the range of $\boldsymbol{\mu}_\psi$ is κ. Then, use equation 11 to convert the marginal means to model coefficients ($\boldsymbol{\psi}$), which are in turn used to calculate power. Iterate on configurations of $\boldsymbol{\mu}_\psi$ to satisfy the principle of conservatism.

4. Signal-to-Noise Transformation Method

The objective of the signal-to-noise transformation method is to allow one to calculate power for a generalized linear model by using power formulas that are intended for classical linear models. The motivation for this approach is that popular software packages accommodate classical linear model power calculations and also allow the user to manipulate the vector of test coefficients, $\boldsymbol{\psi}$, and nuisance coefficients, $\boldsymbol{\lambda}$. Thus, the signal-to-noise transformation method could be used successfully with information available in standard software packages to approximate power for generalized linear models.

The approach assumes that for each run in the experiment $(v = 1, 2, .., n)$ the linear predictors, η_v, in a generalized linear model can be modeled as the response variable, Y_v, in a classical linear model. That is, $Y_v = \eta_v = \boldsymbol{Z}_v\boldsymbol{\psi} + \boldsymbol{X}_v\boldsymbol{\lambda} + \epsilon_v$, where $\epsilon_v \sim N(0, \check{\sigma}^2)$, and the error term ϵ_v is independent and identically distributed. The variance $\check{\sigma}^2$ is the "transformed" noise, meaning it represents the variance of the linear predictor for the generalized linear model.

Another assumption in this approach is that $\check{\sigma}^2$ is constant and is evaluated at the overall mean across the design space $\bar{\mu}$. For example, an analyst may anticipate a 70 percent average probability of detection across the design space fit with a logistic regression model. The overall mean $\bar{\mu}$ impacts $\check{\sigma}^2$ and, in turn, affects power.

A tenet of generalized linear models is that the variance of Y depends on the mean, μ, and dispersion parameter, ϕ. Since we are assuming a non-zero effect size for $\boldsymbol{\psi}$ under the alternative hypothesis, an implication is that μ is not constant, thus neither is $\check{\sigma}^2$. For this reason, only small effect sizes should be considered. In a later section, we use simulation to assess the consequence of this assumption.

Another assumption is that the hypothesis test is constructed without considering nuisance effects. That is, for the hypothesis test $\boldsymbol{\psi} = \mathbf{0}$, the nuisance coefficients take the form $\boldsymbol{\lambda} = (\lambda_{int} \mid \mathbf{0})^T$. Without this assumption, significant values of $\boldsymbol{\lambda}$ could further invalidate the assumption that $\check{\sigma}^2$ is constant because $\boldsymbol{\lambda}$ impacts μ which in turn affects the variance of Y.

As in the previous section, we define the signal-to-noise ratio as $\kappa = \delta/\sigma$. For SNRx, we must "transform" δ and σ to the linear predictor space.

Since Y is a random variable with $\mathrm{E}(Y) = \mu$, we can use $g(Y)$ as an estimator of $g(\mu)$. From Casella and Berger [21], using the delta method, we can say that approximately

$$\begin{aligned} \mathrm{E}(g(Y)) &\approx g(\mu) \\ \mathrm{Var}(g(Y)) &\approx [g'(\mu)]^2\, \mathrm{Var}(Y) \quad . \end{aligned} \tag{12}$$

We also know that for generalized linear models $\mathrm{Var}(Y) = a(\phi)\mathrm{Var}(\mu)$.

Substituting this into equation 12, taking the square root, and evaluating $g'(\mu)$ and $\mathrm{Var}(\mu)$ at $\bar{\mu}$, we obtain an estimate of the noise:

$$\breve{\sigma} = \sqrt{\mathrm{Var}(g(Y))} = g'(\bar{\mu})\sqrt{a(\phi)\mathrm{Var}(\bar{\mu})} \quad . \tag{13}$$

Now that the noise is "transformed," we turn our attention to the signal. If the upper and lower bound of the signal of interest are $\bar{\mu} + \delta/2$ and $\bar{\mu} - \delta/2$, we can convert this quantity to a value in the linear predictor space as $g(\bar{\mu} + \delta/2)$ and $g(\bar{\mu} - \delta/2)$, respectively, where $g(\cdot)$ is the link function for the generalized linear model of interest.

The signal-to-noise ratio is then described as the ratio of the signal and noise within the linear predictor space, as shown in equation 14. Parameters used in equation 14 for common generalized linear models appears in Table 1.

$$\kappa = \frac{g(\bar{\mu} + \delta/2) - g(\bar{\mu} - \delta/2)}{g'(\bar{\mu})\sqrt{a(\phi)V(\bar{\mu})}} \tag{14}$$

Family	$g(\mu)$	$g'(\mu)$	$V(\mu)$	$a(\phi)$	$\breve{\sigma}$
Normal	μ	1	1	σ^2	σ
Binomial	$\log\left(\frac{\mu}{1-\mu}\right)$	$\frac{1}{\mu(1-\mu)}$	$\mu(1-\mu)$	1	$\sqrt{\frac{1}{\mu(1-\mu)}}$
Gamma	$\frac{1}{\mu}$	$-\frac{1}{\mu^2}$	$-\mu^2$	$-\phi$	$-\phi/\sqrt{\mu^2\phi}$
Poisson	$\log(\mu)$	$\frac{1}{\mu}$	μ	1	$\sqrt{\frac{1}{\mu}}$

Table 1: Parameters used in the calculation of κ

The input parameters to the power calculation are intuitive and thus easy to define. In Equation 14 the inputs $\bar{\mu}$ and δ are both in units of the response variable. The other parameter, $a(\phi)$, is less intuitive, but can be understood by plotting the density of Y, as we show in the gamma example later.

4.1. Calculating Approximate Coefficients from the SNRx

After calculating the signal-to-noise ratio in Equation 14, we proceed as if it were a usual power calculation for an analysis of variance of a classical linear model. To illustrate, assume one calculates power for a main effect for a particular generalized linear model. Suppose the main effect ρ_i has four levels. Assume the interest is in detecting an effect size of δ, and the nominal response is $\bar{\mu}$.

To construct the desired effect size, we let $\boldsymbol{\mu}_\psi$ represent the vector of marginal means for ρ_i. We then set the components of $\boldsymbol{\mu}_\psi$ so that the range is κ, and it satisfies the principle of conservatism. Assume the configuration that satisfies the principle of conservatism has the first level of the main effect ρ_1 equal to $\kappa/2$, the second level ρ_2 equal to $-\kappa/2$, and all remaining levels equal to zero. To convert the marginal means to coefficients, one can use Equation 11:

$$\begin{bmatrix} \tilde{\rho}_1 \\ \tilde{\rho}_2 \\ \tilde{\rho}_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{bmatrix}^{-1} \begin{bmatrix} \kappa/2 \\ -\kappa/2 \\ 0 \\ 0 \end{bmatrix}. \tag{15}$$

4.2. Power Calculations for the Linear Model

To calculate power for the hypothesis test $\boldsymbol{\psi} = \mathbf{0}$, let $\boldsymbol{\psi} = [\tilde{\rho}_1 \quad \tilde{\rho}_2 \quad \tilde{\rho}_3]^T$. Then, following the approach by [22], the test statistic is constructed from the residual sum of squares under the full and restricted models and is defined as

$$f = \frac{n-r}{p} \frac{\boldsymbol{Y}^T(\boldsymbol{H}_0 - \boldsymbol{H})\boldsymbol{Y}}{\boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}}, \tag{16}$$

where the hat matrices are $\boldsymbol{H} = \boldsymbol{M}(\boldsymbol{M}^T\boldsymbol{M})^{-1}\boldsymbol{M}^T$ and $\boldsymbol{H}_0 = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$. Under the null hypothesis the test statistic follows a central F distribution where p is the number of parameters to be tested, r is the number of parameters in the full model, and $n-r$ denominator degrees of freedom. Under the alternative hypothesis the test statistic follows a non-central F distribution with p numerator degrees of freedom and $n-r$ denominator degrees of freedom with non-centrality parameter γ_F given in 17.

$$\gamma_F = (\boldsymbol{Z}\boldsymbol{\psi})^T (\boldsymbol{I} - \boldsymbol{H}_0)(\boldsymbol{Z}\boldsymbol{\psi}). \tag{17}$$

Given the distribution of the test statistic under the null and alternative hypotheses, one calculates power in the usual way. The critical value is

$$f_{crit} = \hat{F}(1-\alpha, p, n-r) \quad , \tag{18}$$

where $\hat{F}$ is the F central quantile function, and α is the level of significance. The power of the test is

$$1 - \tilde{F}(f_{crit}, p, n-r, \gamma_F) \quad , \tag{19}$$

where $\tilde{F}$ is the non-central F distribution function.

5. Examples

This section provides three examples that implement the SNRx method to calculate power. In each example a different generalized linear model is chosen.

The design for each example is the same, which is a full factorial experiment with three factors that is replicated four times (96 total observations). The first, second, and third factors have two, three, and four levels, respectively. The ijkth treatment combination of the linear predictor is

$$\eta_{ijk} = \rho_i + \tau_j + \omega_k + \rho\tau_{ij} + \rho\omega_{ik} + \tau\omega_{jk} \quad , \tag{20}$$

where $i = 1, 2; j = 1, 2, 3; k = 1, 2, 3, 4$. A typical approach reports power for all main effects and two factor interactions, assuming the design supports sufficient error degrees of freedom. For brevity, each example demonstrates a power calculation for a single main effect and a single two factor interaction.

5.1. Logistic Regression Example

In logistic regression the response variable for a single run represents a binary random variable (0 or 1). In this example let 1 and 0 represent a success and failure, respectively. In an experiment with N groups or strata, Y_v represents the number of successes in the vth group out of m_v attempts, where $v = 1, 2, ..., N$. Then, a logistic regression model assumes that $Y_v \sim binom(m_v, \pi_v)$.

A few pieces of information are needed to set up the power calculations. The first is the assumed mean response across the design space, $\bar{\mu}$. For logistic regression, the mean response is bounded between zero and one and represents the average probability of success across the design space. For this example we will assume a nominal 70 percent probability of success or $\bar{\mu} = 0.7$.

The second element is δ. Recall that δ is the change in the mean response that is symmetric about $\bar{\mu}$, as shown in Equation 14. In this example we will assume $\delta = 0.3$ so that the change of interest ranges from 0.55 to 0.85, or 55 to 85 percent probability of success.

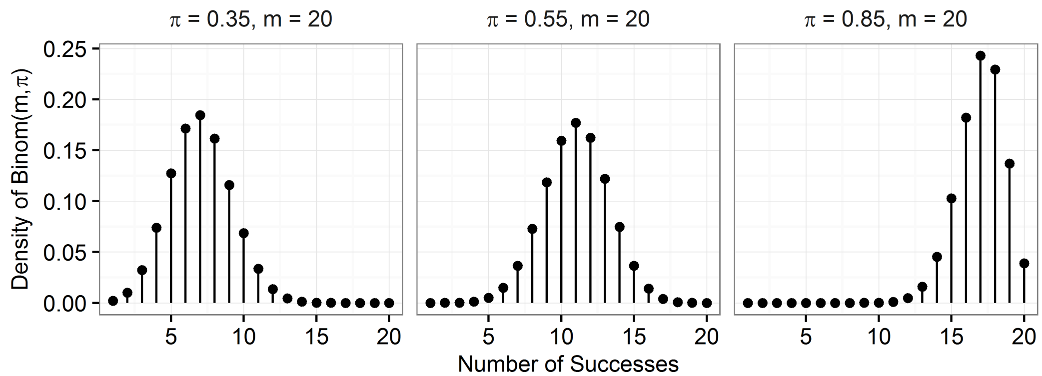

Figure 1: Binomial mass function for different distributional Parameters

The next step is to calculate the signal-to-noise ratio κ. The signal-to-noise ratio can be directly inputted into some software, such as Design Expert, and the corresponding effect power is outputted. In other software, such as JMP, the coefficients anticipated under the alternative hypothesis ($\boldsymbol{\psi}$) must be manually inputted using the approach outlined below. Plugging the assumed values for this example into Equation 14 we get

$$\kappa = \frac{log(\frac{.7+.3/2}{1-(.7+.3/2)}) - log(\frac{.7-.3/2}{1-(.7-.3/2)})}{\frac{1}{.7(1-.7)}\sqrt{(1).7(1-.7)}} = .70 \tag{21}$$

To obtain the approximate coefficients we construct the marginal mean effect so that its range is equal to κ, and then convert the marginal mean effect to coefficients using Equation 11. We demonstrate this process for the three-level main effect τ_j. Using equation 11 the relationship between the effect coefficients and the marginal mean effect is

$$\begin{bmatrix} \tilde{\tau}_1 \\ \tilde{\tau}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix}^{-1} \begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix} \quad . \tag{22}$$

Next, the levels of τ_1, τ_2, and τ_3 are set so that their range is equal to κ. Additionally, Ohlert and Whitcomb's principle of conservatism requires a search of the possible configurations of τ_j to find the pairwise difference between components of τ_j that yields the minimum power. Details about this process can be found in their paper [20]. After conducting the search, we identify a pairwise difference that yields minimum power, and set the first level of τ_j equal to $\kappa/2$ and the second level equal to $-\kappa/2$. Equation 22 now becomes

$$\begin{bmatrix}\tilde{\tau}_1\\ \tilde{\tau}_2\end{bmatrix} = \begin{bmatrix}1 & 0\\ 0 & 1\\ -1 & -1\end{bmatrix}^{-1} \begin{bmatrix}.70/2\\ -.70/2\\ 0\end{bmatrix} \quad . \tag{23}$$

Solving equation 23, we get $\boldsymbol{\psi} = [\tilde{\tau}_1 \quad \tilde{\tau}_2]^T = [.70/2 \quad -.70/2]^T$.

Next, consider the interaction $\rho\omega_{ik}$. Similar to the main effect calculation, we will use Equation 11 to calculate the coefficients from the marginal means. This relationship is written as

$$\begin{bmatrix}\tilde{\rho\omega}_{11}\\ \tilde{\rho\omega}_{12}\\ \tilde{\rho\omega}_{13}\end{bmatrix} = \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\\ -1 & -1 & -1\\ -1 & 0 & 0\\ 0 & -1 & 0\\ 0 & 0 & -1\\ 1 & 1 & 1\end{bmatrix}^{-1} \begin{bmatrix}\rho\omega_{11}\\ \rho\omega_{12}\\ \rho\omega_{13}\\ \rho\omega_{14}\\ \rho\omega_{21}\\ \rho\omega_{22}\\ \rho\omega_{23}\\ \rho\omega_{24}\end{bmatrix} \quad . \tag{24}$$

Next, we set the levels of $\rho\omega_{ik}$ so that the range is equal to κ. Note the values that are chosen for $\rho\omega_{ik}$ must satisfy the constraints shown in Equation 4. Additionally, the principle of conservatism requires a search of the possible configurations of $\rho\omega_{ik}$ for that which yields the minimum power. After conducting this search, we arrive at the configuration in Equation 25.

$$\begin{bmatrix}\tilde{\rho\omega}_{11}\\ \tilde{\rho\omega}_{12}\\ \tilde{\rho\omega}_{13}\end{bmatrix} = \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\\ -1 & -1 & -1\\ -1 & 0 & 0\\ 0 & -1 & 0\\ 0 & 0 & -1\\ 1 & 1 & 1\end{bmatrix}^{-1} \begin{bmatrix}.70/2\\ -.70/2\\ 0\\ 0\\ -.70/2\\ .70/2\\ 0\\ 0\end{bmatrix} \quad . \tag{25}$$

Solving equation 25, we get $\boldsymbol{\psi} = [\tilde{\rho\omega}_{11} \quad \tilde{\rho\omega}_{12} \quad \tilde{\rho\omega}_{13}] = [.70/2 \quad -.70/2 \quad 0]^T$.

At this point we have an approximation of the anticipated coefficients that can be inputed into software. The rest of this example demonstrates the power calculation manually. As mentioned before, the design is a full factorial that is replicated four times so the model matrix $\boldsymbol{M}$ is size 96×18. The first column of $\boldsymbol{M}$ corresponds to the intercept, columns 2 through 7 correspond to the main effects, and columns 8 through 18 correspond to the two factor interactions. The coefficient vector $\boldsymbol{\beta}$ is size 18×1.The power calculation requires that we divide the model matrix into the test matrix $\boldsymbol{Z}$ and the nuisance matrix $\boldsymbol{X}$.

For the test on the main effect τ_j the test matrix $\boldsymbol{Z}$ is size 96×2, and the test coefficient vector that we calculated before ($\boldsymbol{\psi}$) is size 2×1. The nuisance matrix is 96×16. We calculate the hat matrix $\boldsymbol{H}_0$, and use $\boldsymbol{H}_0$, $\boldsymbol{Z}$, and $\boldsymbol{\psi}$ in Equation 17 to find the the non-centrality parameter $\gamma_F = 7.91$. By setting the significance $\alpha = 0.05$, we find the critical F value $f_{crit} = 3.11$, and then power equals 0.69. An adequate test design requires 290 samples to provide 80 percent power.

For the test on the interaction $\rho\omega_{ik}$ the test matrix $\boldsymbol{Z}$ is size 96×3, and the test coefficient vector $\boldsymbol{\psi}$ is size 3×1. The nuisance matrix $\boldsymbol{X}$ is size 96×15. Next, we calculate the hat matrix $\boldsymbol{H}_0$, then use $\boldsymbol{H}_0$, $\boldsymbol{Z}$, and $\boldsymbol{\psi}$ in Equation 17 to find the the non-centrality parameter $\gamma_F = 5.92$. By setting the significance $\alpha = 0.05$, we find the critical F value $f_{crit} = 2.72$, and then power equals 0.49. An adequate test design requires 440 samples to provide 80 percent power.

Figure 2 shows power as a function of δ, $\bar{\mu}$, and sample size for four of the model effects. We assume a fixed significance level α equal to 0.05. The horizontal axis shows the sample size for the experimental design, which was generated using a D-optimal algorithm. As we expect, power increases with an increase of δ or sample size. Power is less for $\bar{\mu} = 0.5$, compared to $\bar{\mu} = 0.7$, because the variance in the response variable is greatest at $\bar{\mu} = 0.5$. Lastly, among the four effects that are plotted as different colors in Figure 2, we see that power decreases as the number of degrees of freedom in the effect increases.

5.2. Gamma Generalized Linear Model Example

In an n run experiment, a gamma generalized linear model assumes that the response variables are independent gamma observations $Y_1, Y_2, ..., Y_n$, where $E(Y_v) = \mu_v$. Y_v is assumed to follow a gamma distribution, which is expressed as $Y_v \sim G(1/\phi, \mu_v\phi)$, where $1/\phi$ is the shape parameter and $\mu_v\phi$ is the scale parameter.

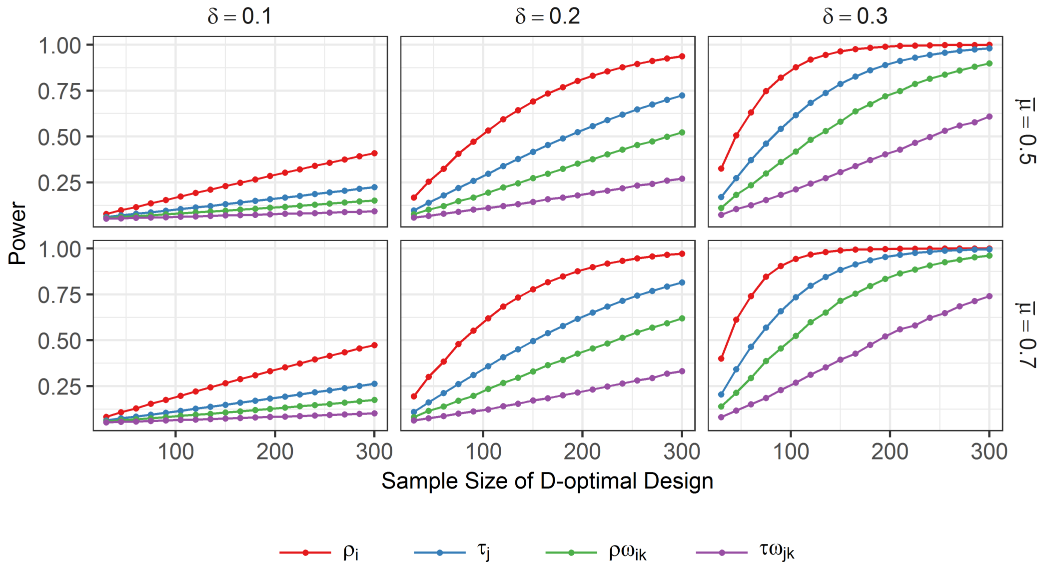

Figure 2: Power Trend for Logistic Regression Model Example

To understand the influence of ϕ, we can plot the densities of $G(1/\phi, \mu_v\phi)$ for different values of μ_v and ϕ, as shown in Figure 3. As ϕ becomes larger, the kurtosis of the density decreases and the variability of Y_v increases. A special case of the gamma distribution is $\phi = 1$, which is the exponential distribution. In planning an experiment, it is always best to estimate ϕ from historical data.

For this example, we are interested in testing a projectile weapon, and the response variable is miss distance. We assume the dispersion parameter ϕ equals .1, and the overall mean response $\bar{\mu}$ equals 20 feet. We are interested in detecting a change in the mean response δ equal to 6 feet.

Using the formula in Table 1 with equation 14, we calculate the signal-to-noise ratio as

$$\kappa = \frac{(20+6/2)^{-1} - (20-6/2)^{-1}}{-.1/\sqrt{(20^2)(.1)}} = .97 \quad . \tag{26}$$

Similar to the logistic regression example, we construct the marginal mean effect so that its range is equal to κ. We then convert that marginal mean effect to coefficients using Equation 11. For the power calculation on the main effect we find $\boldsymbol{\psi} = [\tilde{\tau}_1 \quad \tilde{\tau}_2] = [.48 \quad -.48]$. Then, using the same steps as the logistic regression example, and using the experimental design outlined

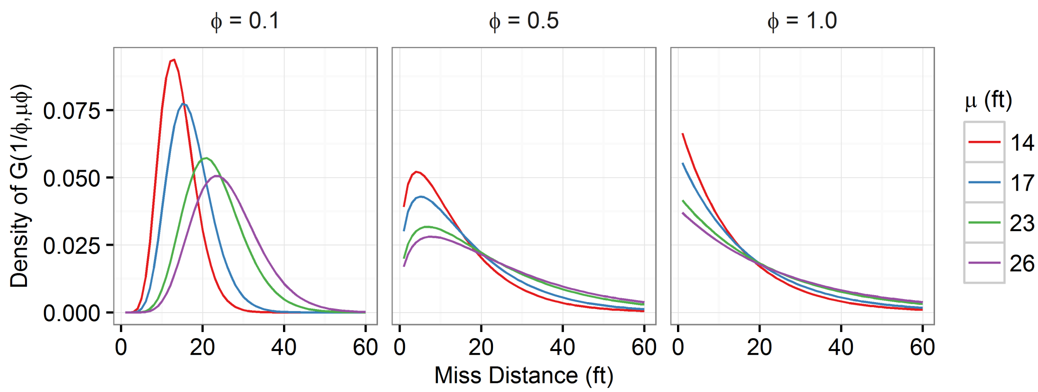

Figure 3: Densities for various shape and scale parameters for the gamma distribution

at the beginning of this section, $\alpha = .05$, $\gamma_F = 15.07$, $f_{crit} = 3.11$, and power is equal to 0.94.

For the power calculation on the interaction we get $\boldsymbol{\psi} = [\widetilde{\rho\omega}_{11} \quad \widetilde{\rho\omega}_{12} \quad \widetilde{\rho\omega}_{13}] = [.48 \quad -.48 \quad 0]$. Then, using the same steps as before, $\alpha = .05$, $\gamma_F = 11.30$, $f_{crit} = 2.72$, and power is equal to 0.79.

Figure 4 shows power as a function of δ, $\bar{\mu}$, ϕ, and sample size for four of the model effects. We assume a fixed significance level α equal to 0.05. The horizontal axis shows the sample size for the experimental design, which was generated using a D-optimal algorithm. As in the previous example, power increases with an increase of δ or sample size. Power is less for $\bar{\mu} = 30$, compared to $\bar{\mu} = 20$, because the noise in the response variable increases as the overall mean increases. Lastly, among the four effects that are plotted as different colors in Figure 4, we see that power decreases as the number of degrees of freedom in the effect increases.

5.3. Poisson Generalized Linear Model Example

Let $Y_1, ..., Y_N$ be independent random variables with Y_v denoting the number of events observed from the vth treatment combination, where $Y_v \sim Poisson(\mu_v)$. In this example we let Y_v represent the number of aircraft failures during the vth mission, and there are N total missions. We assume that the length of each mission is constant. Figure 5 shows the distribution of Y_v for different values of μ_v.

Assume that we anticipate an overall mean response of 2.5 failures ($\bar{\mu} = 2.5$), and that we would like to detect an effect size of 1.5 failures ($\delta = 1.5$).

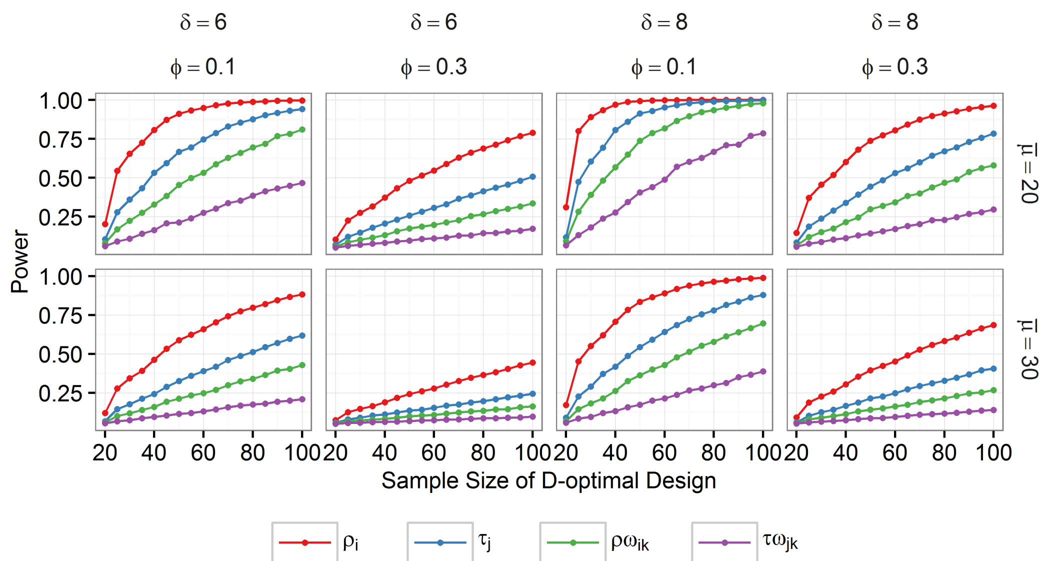

Figure 4: Power Trends for the Gamma Regression Example

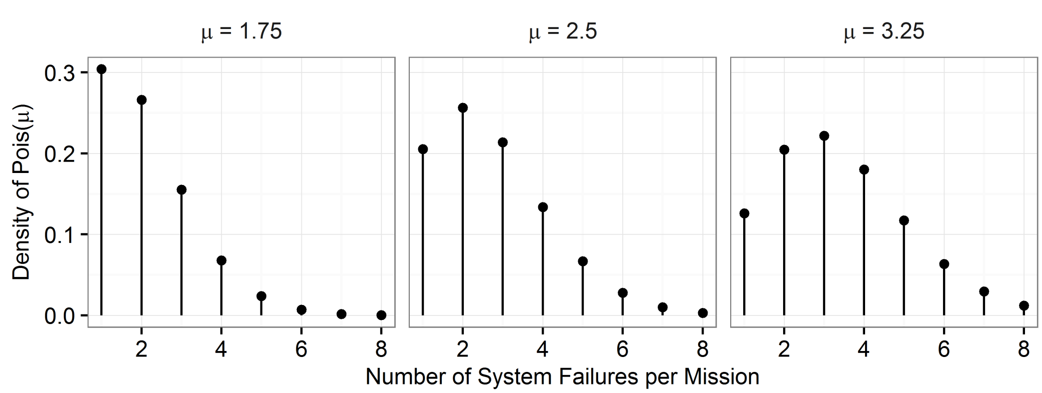

Figure 5: Poisson densities for various values of the mean parameter

The first step is to calculate the signal-to-noise ratio. Using the formula for the Poisson distribution in Table 1 with equation 14, we get

$$\kappa = \frac{log(2.5 + 1.5/2) - log(2.5 - 1.5/2)}{\sqrt{1/2.5}} = .98 \quad . \tag{27}$$

Similar to the logistic regression example, we construct the marginal mean effect so that its range is equal to κ. We then convert that marginal mean effect to coefficients using Equation 11.

For the main effect power calculation we get $\boldsymbol{\psi} = [\tilde{\tau}_1 \quad \tilde{\tau}_2] = [.49 \quad -.49]$. Then, with $\alpha = .05$, $\gamma_F = 15.32$, and $f_{crit} = 3.11$, power is equal to 0.94.

For the power calculation on the interaction we get $\boldsymbol{\psi} = [\tilde{\rho\omega}_{11} \quad \tilde{\rho\omega}_{12} \quad \tilde{\rho\omega}_{13}] = [.49 \quad -.49 \quad 0]$. Then, with $\alpha = .05$, $\gamma_F = 11.50$, $f_{crit} = 2.72$, power is equal to 0.80.

Figure 6 shows power as a function of δ, $\bar{\mu}$, and sample size for four of the model effects. We assume a fixed significance level α equal to 0.05. The horizontal axis shows the sample size for the experimental design, which was generated using a D-optimal algorithm. As we expect, power increases with an increase of δ or sample size. Power is less for $\bar{\mu} = 3.5$, compared to $\bar{\mu} = 2.5$, because the noise in the response variable increases with $\bar{\mu}$. Lastly, among the four effects that are plotted as different colors in Figure 6, we see that power decreases as the number of degrees of freedom in the effect increases.

6. Simulation Study

6.1. Simulation Setup

Model	$\bar{\mu}$ = small	$\bar{\mu}$ = large	δ = small	δ = large	ϕ
Binomial	0.5	0.7	0.1	0.2	1
Poisson	7	11	1	2	1
Gamma	15	20	3	5	0.2, 0.3

Table 2: Simulation parameters.

This section compares power from SNRx to power calculated from Monte Carlo simulation based on a likelihood ratio statistic. We investigate three generalized linear models for various values of $\bar{\mu}$, δ, and ϕ. Each model

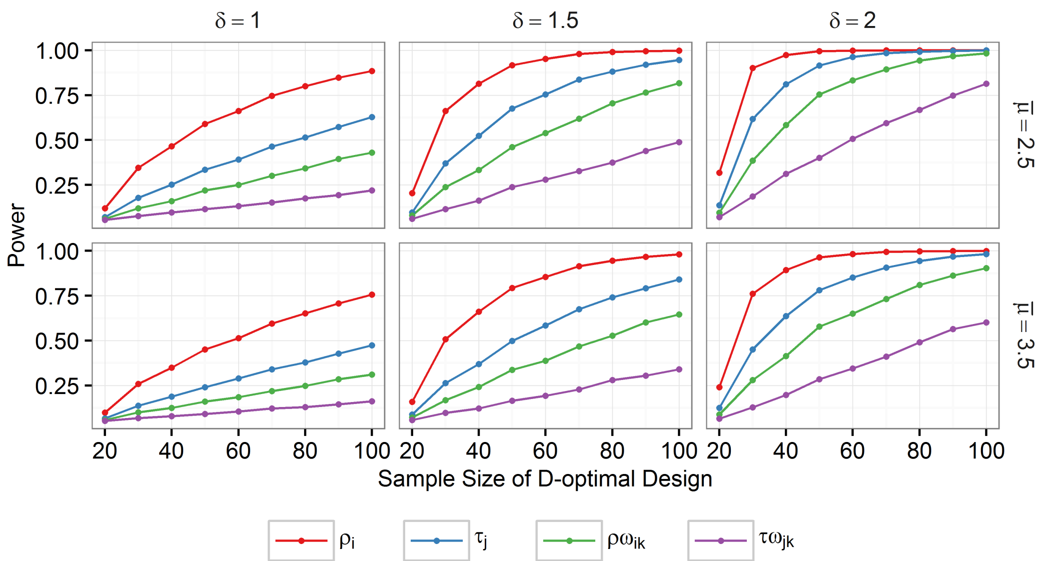

Figure 6: Power Trends for the Poisson Regression Example

uses the canonical link, and has the same model effects as described in the examples section (equation 20). We consider two different replicated factorial experiments with a small and large sample size. For logistic regression, we investigate larger sample sizes than the Poisson or gamma models, so that we can achieve adequate power for reasonable effect sizes. The simulation assumes a Type I error rate (α) equal 0.05. A summary of the simulation input parameters appears in Table 2.

A proper comparison between SNRx and simulation requires a comparable hypothesis test. Recall that the SNRx hypothesis test is conducted on the coefficients of the classical linear model. The simulation approach is based on the likelihood ratio statistic from an analysis of deviance, so it is conducted on the coefficients of the generalized linear model. The hypothesis test on the coefficients of the generalized linear model is $\boldsymbol{\psi} = \mathbf{0}$, and the model takes the form

$$g(\mu_v) = \boldsymbol{X}_v \boldsymbol{\lambda} + \boldsymbol{Z}_v \boldsymbol{\psi} \quad , \tag{28}$$

where the mean response of the ith run in the design is $\mu_v = E(Y_v)$, $\boldsymbol{X}_v$ is the vth row of the nuissance matrix, $\boldsymbol{Z}_v$ is the vth row of the test matrix, and $g(\cdot)$ is the link function. Thus, to compare power between SNRx and simulation we need to construct $\boldsymbol{\psi}$ and $\boldsymbol{\lambda}$ for the generalized linear model.

A comparable hypothesis test uses the same inputs as the SNRx approach. The nominal mean response across the design space is $\bar{\mu}$, and the change in the mean response is δ. This leads to a similar approach as before. We first define the effect size in terms of the response and then convert that effect size to coefficients. A difference here is that the effect size is first converted to values in the linear predictor space (Equation 20), and then it is converted to coefficients ($\boldsymbol{\psi}$ and $\boldsymbol{\lambda}$). The procedure is as follows.

We first solve for the average response in the linear predictor space using the relationship $\eta_0 = g(\bar{\mu})$. Then, we numerically solve equation 29 for δ_η, where η_0 and δ are known. In this equation δ_η represents the range of the effect within the linear predictor space.

$$g^{-1}\left(\eta_0 + \delta_\eta/2\right) - g^{-1}\left(\eta_0 - \delta_\eta/2\right) = \delta \tag{29}$$

Now that we have have calculated the effect size in terms of the linear predictor, we can express that effect size in terms of the marginal means model from equation 20. Then, we use equation 11 to calculate the coefficients ($\boldsymbol{\psi}$ and $\boldsymbol{\lambda}$).

For example, to test the main effect τ_j using a logistic regression model with $\delta = 0.2$ and $\bar{\mu} = 0.7$, equation 11 can be written as

$$\begin{bmatrix} \tilde{\tau}_1 \\ \tilde{\tau}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix}^{-1} \begin{bmatrix} \delta_\eta/2 \\ -\delta_\eta/2 \\ 0 \end{bmatrix} \quad . \tag{30}$$

Inserting the solution of equation 29 into equation 30 and solving equation 30, gives $\lambda_{int} = .847$ and $\boldsymbol{\psi} = [.481 \quad -.481]$. For the two factor interaction we get $\lambda_{int} = .847$ and $\boldsymbol{\psi} = [.481 \quad -.481 \quad 0]$. Note that the nuisance coefficients take the form $\boldsymbol{\lambda} = (\lambda_{int} \mid \mathbf{0})^T$. This procedure of generating $\boldsymbol{\psi}$ and $\boldsymbol{\lambda}$ is carried out for each variation of simulation inputs (each combination of model, δ, $\bar{\mu}$, ϕ shown in Table 2).

Monte Carlo simulation uses the usual likelihood ratio test that is commonly calculated in software that produces analysis of deviance tables for generalized linear models. The simulation calculation steps are described below.

1. Construct $\boldsymbol{\psi}$ and $\boldsymbol{\lambda}$, which represent the assumed true values of the model coefficients under the alternative hypothesis.

2. Draw y_v from the probability distribution or mass function of the form $f(y \mid \boldsymbol{\psi}, \boldsymbol{\lambda}, \phi)$ for each run in the experiment ($v = 1, 2, ..., N$). For instance, for logistic regression, draw y_v from Binom$(1, \mu_v)$ for each run in the design, where $\mu_v = g^{-1}(\boldsymbol{X}_v\boldsymbol{\lambda} + \boldsymbol{Z}_v\boldsymbol{\psi})$.
3. Fit the full model, of the form $\boldsymbol{X}\boldsymbol{\lambda} + \boldsymbol{Z}\boldsymbol{\psi}$, to y_v to obtain estimates of $\hat{\boldsymbol{\lambda}}$ and $\hat{\boldsymbol{\psi}}$.
4. Fit the restricted or null model, of the form $\boldsymbol{X}\boldsymbol{\lambda}$, to y_i to obtain an estimate of $\hat{\boldsymbol{\lambda}}_0$. Set $\boldsymbol{\psi}_0$ equal to zeros.
5. Calculate the likelihood ratio statistic, $T = 2[l(\hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\lambda}}) - l(\boldsymbol{\psi}_0, \hat{\boldsymbol{\lambda}}_0)]$, where $l(\cdot)$ is the log-likelihood.
6. Iterate steps 2 through 5 to obtain a distribution of likelihood ratio statistics. This is the alternative distribution of the test statistic. We used 10,000 iterations.
7. Calculate the critical value from the null distribution of the test statistic as $T_{crit} = \hat{\chi}^2(1 - \alpha, p)$.
8. Calculate power as the proportion of iterations that result in $T > T_{crit}$.

6.2. Simulation Results

The results of the simulation study in Table 3 verify that SNRx successfully generates accurate power estimates. Power from SNRx agrees with the simulated values that are based on the likelihood ratio statistic. As we previously mentioned, a potential source of inaccuracy for SNRx comes from the assumption that $\breve{\sigma}^2$ is constant. More specifically, in SNRx we define the effect size by changes in μ. We know that for generalized linear models Var(μ) depends on μ, thus Var(μ) is not constant. But in SNRx we assume it is anyway, and evaluate it at $\bar{\mu}$ with the hope that small values of δ lead to a negligible inaccuracy of power. For the conditions simulated, this seems to be true.

7. Discussion

Through a series of reasonable assumptions, SNRx enables generalized linear model power approximations using an F-test. The main assumption is that $\breve{\sigma}^2$ is constant. The second assumption is that their are no nuisance effects. Given these assumptions, and for small effect sizes, the SNRx method generates power estimates that closely resemble that of a more typical likelihood ratio statistic, simulation-based approach.

			$\bar{\mu}$ = small δ = small		$\bar{\mu}$ = small δ = large		$\bar{\mu}$ = large δ = small		$\bar{\mu}$ = large δ = large	
Model	N	Effect	SNRx	Sim	SNRx	Sim	SNRx	Sim	SNRx	Sim
Binom	240	τ_j	0.19	0.21	0.62	0.65	0.22	0.25	0.72	0.72
$\phi = 1$		$\rho\omega_{ik}$	0.13	0.15	0.43	0.47	0.15	0.17	0.51	0.53
	480	τ_j	0.34	0.36	0.91	0.91	0.40	0.42	0.96	0.95
		$\rho\omega_{ik}$	0.22	0.25	0.75	0.76	0.26	0.28	0.84	0.83
Poisson	72	τ_j	0.20	0.19	0.64	0.63	0.14	0.14	0.45	0.43
$\phi = 1$		$\rho\omega_{ik}$	0.14	0.13	0.45	0.43	0.10	0.10	0.29	0.28
	144	τ_j	0.37	0.36	0.92	0.92	0.24	0.24	0.76	0.75
		$\rho\omega_{ik}$	0.24	0.23	0.77	0.77	0.16	0.16	0.56	0.55
Gamma	72	τ_j	0.26	0.26	0.60	0.63	0.16	0.16	0.37	0.38
$\phi = .2$		$\rho\omega_{ik}$	0.17	0.17	0.42	0.43	0.12	0.11	0.25	0.25
	144	τ_j	0.47	0.49	0.90	0.92	0.29	0.29	0.68	0.69
		$\rho\omega_{ik}$	0.32	0.32	0.73	0.77	0.20	0.19	0.48	0.49
Gamma	72	τ_j	0.18	0.18	0.43	0.46	0.13	0.12	0.27	0.27
$\phi = .3$		$\rho\omega_{ik}$	0.12	0.13	0.28	0.30	0.09	0.09	0.17	0.18
	144	τ_j	0.33	0.34	0.73	0.78	0.21	0.20	0.49	0.51
		$\rho\omega_{ik}$	0.22	0.22	0.54	0.58	0.14	0.14	0.33	0.34

Table 3: Each cell lists the power for the corresponding $\bar{\mu}$, δ, experimental design, ϕ, and power calculation method. See Table 2 for the specific values of $\bar{\mu}$ and δ for each model.

The SNRx approach is useful for analysts that have access to software that calculates power for classical linear models using an F-test, but don't have programming knowledge or access to software that calculates power for generalized linear models using simulation or other approximation methods. Moreover, it provides a quick methodology for comparing multiple designs of varying sizes.

In this paper we size the effect following the unified approach by Ohlert and Whitcomb. There are two benefits of this approach. The first benefit is related to the first key element of their approach, which states that the effect size is the range in the response due to that effect. This leads to an effect size that is defined in units of the response, which is a very intuitive way of defining an effect size. Also, the inputs to this effect size calculation are few and simple: the mean response across the design space $\bar{\mu}$, the change in the mean due to an effect δ, and the dispersion parameter.

The second benefit comes from the principle of conservatism. This part of the unified effect size definition ensures that the estimate of power is consistent. This means that for an experimental design with a set effect size and model form, two separate organizations will calculate the same value of power. This is beneficial in the world of defense testing where there is much bargaining over experimental design changes. The consistent power estimate simplifies the bargaining process.

Despite all these benefits, we also have a few words of caution. The unified approach to sizing an effect is useful when a new system is tested in a new environment and past test data is unavailable to inform the precise anticipated effect size or model coefficients. If past test data is available, it is more prudent to use that data to define effect sizes for individual effects within the model.

As another word of caution, large effect sizes could lead to inaccurate SNRx power estimates. A large effect size invalidates the assumption that $\breve{\sigma}^2$ is constant. This assumption is further invalidated if large nuisance effects are anticipated. At the same time, we recognize that large effect sizes are easier to detect, often resulting in adequate power. The SNRx method generates accurate power estimates as long as small effect sizes are assumed.

Future work could investigate alternative unified effect size definitions for generalized linear models. Ohlert and Whitcomb's unified approach works well for classical linear model power calculations, making it suitable for the SNRx method, but it is not tailored for generalized linear models. A unified effect size definition for a generalized linear model would account for nuisance effects (i.e. not assume nuisance effects are equal to zero as we do in SNRx), and would accommodate large effect sizes. This alternative unified approach would work entirely within a generalized linear model framework, and could leverage the likelihood ratio statistic-based power approximation approach by Self and Mauritsen.

The authors have developed an R Shiny application that evaluates power using both SNRx and Monte Carlo simulation. The application and source code are available upon request.

8. References

[1] L. J. Freeman, A. G. Ryan, J. L. Kensler, R. M. Dickinson, G. G. Vining, A tutorial on the planning of experiments, Quality Engineering 25 (2013) 315–332.

[2] R. T. Johnson, G. T. Hutto, J. R. Simpson, D. C. Montgomery, Designed experiments for the defense community, Quality Engineering 24 (2012) 60–79.

[3] G. J. Hahn, W. Meeker, P. I. Feder, The evaluation and comparison of experimental designs for fitting regression relationships, Journal of Quality Technology 8 (1976) 140–157.

[4] D. C. Montgomery, Design and analysis of experiments, John Wiley & Sons, 2008.

[5] A. S. Whittemore, Sample size for logistic regression with small response probability, Journal of the American Statistical Association 76 (1981) 27–32.

[6] D. F. Signorini, Sample size for poisson regression, Biometrika 78 (1991) 446–450.

[7] R. G. OBrien, Using the sas system to perform power analyses for log-linear models, in: Proc. 11th Annual SAS Users Group Conference, p. 778784.

[8] G. Shieh, Sample size calculations for logistic and poisson regression models, Biometrika 88 (2001) 1193–1199.

[9] E. Demidenko, Sample size determination for logistic regression revisited, Statistics in medicine 26 (2007) 3385–3397.

[10] S. G. Self, R. H. Mauritsen, Power/sample size calculations for generalized linear models, Biometrics (1988) 79–86.

[11] S. G. Self, R. H. Mauritsen, J. Ohara, Power calculations for likelihood ratio tests in generalized linear models, Biometrics (1992) 31–39.

[12] G. Shieh, On power and sample size calculations for wald tests in generalized linear models, Journal of statistical planning and inference 128 (2005) 43–59.

[13] R. Newson, Generalized power calculations for generalized linear models and more (2004).

[14] J. K. Lindsey, Applying generalized linear models, Springer Science & Business Media, 1997.

[15] J. A. Nelder, R. W. M. Wedderburn, Generalized linear models, Journal of the Royal Statistical Society. Series A (General) 135 (1972) 370–384.

[16] R. R. Hocking, Methods and applications of linear models: regression and the analysis of variance, John Wiley & Sons, 2013.

[17] P. McCullagh, J. A. Nelder, Generalized linear models, volume 37, CRC press, 1989.

[18] J. H. Goodnight, W. R. Harvey, Least squares means in the fixed effects general model, SAS Inst., 1978.

[19] S. R. Searle, F. M. Speed, G. A. Milliken, Population marginal means in the linear model: An alternative to least squares means, The American Statistician 34 (1980) 216–221.

[20] G. W. Oehlert, P. Whitcomb, Sizing fixed effects for computing power in experimental designs, Quality and Reliability Engineering International 17 (2001) 291–306.

[21] G. Casella, R. L. Berger, Statistical inference, volume 2, Duxbury Pacific Grove, CA, 2002.

[22] A. C. Rencher, G. B. Schaalje, Linear models in statistics, John Wiley & Sons, 2008.