Title
Not just normal: Exploring power with Shiny apps

Journal
Technology Innovations in Statistics Education, 13(1)

Authors
Stratton, Christian
Green, Jennifer L
Hoegh, Andrew

Publication Date
2021

DOI
10.5070/T513146468

Not just normal: Exploring power with Shiny apps

Christian Stratton[1], Jennifer L. Green[2], and Andrew Hoegh[1]

[1]Department of Mathematical Sciences, Montana State University

[2]Program in Mathematics Education, Department of Statistics and Probability, Michigan State University

Abstract

Statistical power is an important topic taught in most graduate-level and undergraduate-level mathematical statistics courses, but it is often difficult to understand conceptually. Visualizing the power curve, sampling distributions, and how they interact can help students more easily conceptualize power, but the creation of such visuals can be difficult and time-consuming. Interactive web applications provide a way for students to dynamically visualize power, and many web applications for understanding power exist. Most existing applications assume samples are drawn from a Normal population, concern only the sample mean, and/or were created for introductory classes. We developed a new web application suitable for undergraduate-level and graduate-level mathematical statistics courses that allows users to visualize the complex relationships underlying power for multiple different estimators and population distributions (available at `https://shiny.stt.msu.edu/jg/powerapp/`); source code is provided for instructors who wish to modify the application. Our experience implementing this application across two different semesters is also discussed, and example activities are provided in the appendix.

Keywords: mathematical statistics, statistics education, statistical power, R shiny applications

1. INTRODUCTION

Statistical power is a fundamental component of most graduate-level and undergraduate-level mathematical statistics courses, and yet it is frequently cited as one of the most difficult topics to teach (Aberson, Berger, Healy, and Romero 2002). This difficulty is driven by the conceptual richness of the topic. To fully understand power, students must understand the relationship between the sampling distributions of the test statistic across both the null and alternative parameter spaces and how they are used to compute power, as well as the factors that influence those distributions. Guiding students towards a conceptual understanding of these relationships can be challenging, as the relationships are difficult to visualize using traditional classroom methods (Aberson et al. 2002).

In conventional upper-level mathematical statistics courses, such as curricula based on Wackerly, Mendenhall III, and Scheaffer (2008) or Casella and Berger (2002), power is typically explored through its definition and derivations of power functions. Casella and Berger (2002) define power by considering a hypothesis test of $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_0^c$ and defining

$\mathcal{R}$ as the rejection region for that test. They then define the power function of this test as $\beta(\theta) = P(\boldsymbol{X} \in \mathcal{R}|\theta)$. In words, the power function represents the probability of rejecting the null hypothesis for a given value of θ. Casella and Berger (2002) follow this definition with derivations of the power function for samples from the binomial distribution and the normal distribution, respectively. This method of presenting and exploring statistical power - largely through derivations - is mirrored in many of the traditional curricula based on this text.

From the definition and derivation of a hypothesis test's power function, students are expected to understand a range of learning objectives. Potential learning objectives for power include but are not limited to:

1) understanding how the sample size, null value, alternative hypothesis, significance level, test statistic, and population distribution each affect the power function;

2) understanding the relationship between the power function and the sampling distribution of the test statistic under both the null hypothesis and the true value of θ; and

3) understanding how to calculate the power function of a hypothesis test using a variety of strategies.

Students' conceptual understanding of complex topics such as power can be enhanced by the use of graphical and visualization techniques (delMas, Garfield, and Chance 1999; Chance, Ben-Zvi, Garfield, and Medina 2007; Bobek and Tversky 2016). The *Guidelines for Assessment and Instruction in Statistics Education* (GAISE) College Report promotes using technology to visually represent complex introductory statistics concepts (American Statistical Association 2016). The *2014 Curriculum Guidelines for Undergraduate Programs in Statistical Science* (American Statistical Association 2014) extend these suggestions to undergraduate statistics courses in general, placing renewed emphasis on creating undergraduate statistics curricula that allow students to develop a conceptual understanding of the material and think critically. Wild, Pfannkuch, Regan, and Horton (2011) describe how technology can be used to "change the landscape of statistics education," (p. 248) due to its ability to allow students to conceptualize statistical concepts through visualization. Green and Blankenship (2015) describe various tools that can be used to foster active learning and conceptual understanding of the concepts discussed in a traditional mathematical statistics curriculum; among these tools is the use of technology and visualizations.

To create visualizations that develop a conceptual understanding of power, students must first derive multiple power functions - for differing population distributions, statistics, and alternative hypotheses - and then plot the functions using statistical software. Although there is value to each of these steps, the time commitment and difficulty of the task may frustrate students and distract from understanding the learning objectives. Alternatively, instructors may create graphics either during class in real-time or in advance. The former case requires substantial coding experience on the part of the instructor and demands class time, and the latter case can be restrictive in that students cannot organically explore factors that affect the power function as the visuals were created by the instructor in advance.

Alternatively, instructors may use web-based applications when teaching power that allow students to make interactive discoveries using point-and-click graphical interfaces. Using such applications lets students interactively explore graphical representations of power without having to first create those visualizations. Many graphical tools for understanding statistical power exist (Aberson et al. 2002; Anderson-Cook and Dorai-Raj 2003; Rossman and Chance 2004; Post 2016), but most assume samples are drawn from a normal population and do not provide a visualization of the power curve. To offer other options, instructors may either adapt a currently existing application, which requires access to source code and literacy in the source coding language, or develop one of their own, which requires considerable time and knowledge of coding languages (Doi, Gail, Wong, Alcaraz, and Peter 2016).

The goal of this work is to create a web application that allows students to explore factors that affect the power function and sampling distribution of a test statistic. Additionally, we wanted students to be able to visualize and interactively explore how the sampling distributions of the test statistic over the null and alternative parameter spaces are used to calculate power. Unlike existing options, the application permits many different test statistics and distributions beyond the normal distribution. In this article, we discuss how the layout and features of the application are designed to help promote students' conceptual understanding of power through the use of visualizations. Then we describe how this web application can be used to address the learning objectives provided above and share our experience using it in undergraduate-level and graduate-level mathematical statistics courses. Finally, we discuss the value of this web application and provide opportunities for future work.

2. APPLICATION LAYOUT

The R Shiny power application (Figure 1) serves two primary goals. The first goal is to allow students to visualize the relationship between the power function and various factors that influence it. In particular, students can explore the effect of the population distribution, statistic of interest, null value, alternative hypothesis, significance level, and sample size on the power function in real time without needing to derive or plot that function. The second goal is for students to develop a conceptual understanding of how power is derived by visualizing the relationship between the power function and the sampling distribution under both the null hypothesis and the true value of θ. By addressing these two goals, this web tool provides an efficient way to support students' conceptual understanding of power and sampling distributions through visualizations without spending class time creating them.

The web application features three different families of population distributions: the exponential(θ) family, the normal(θ, σ^2) family, and the uniform(0, θ) family of distributions. For the exponential and normal families, four statistics are available: the sum of n random variables, the sample minimum, the sample maximum, and the sample mean. For the uniform family, the sum of the random variables, the sample minimum, and the sample maximum are available. For each combination of population distribution and statistic, there are three alternative hypotheses available: a not equal to alternative, a greater than alternative, and a less than alternative. The user also has the option to specify the sample size, significance level, and null value. Based on these options, the power curve is plotted in the main plotting window. Beneath that, the sampling distribution of the statistic under both the null hypothesis and true value of θ is plotted once the respective point on the power curve is clicked.

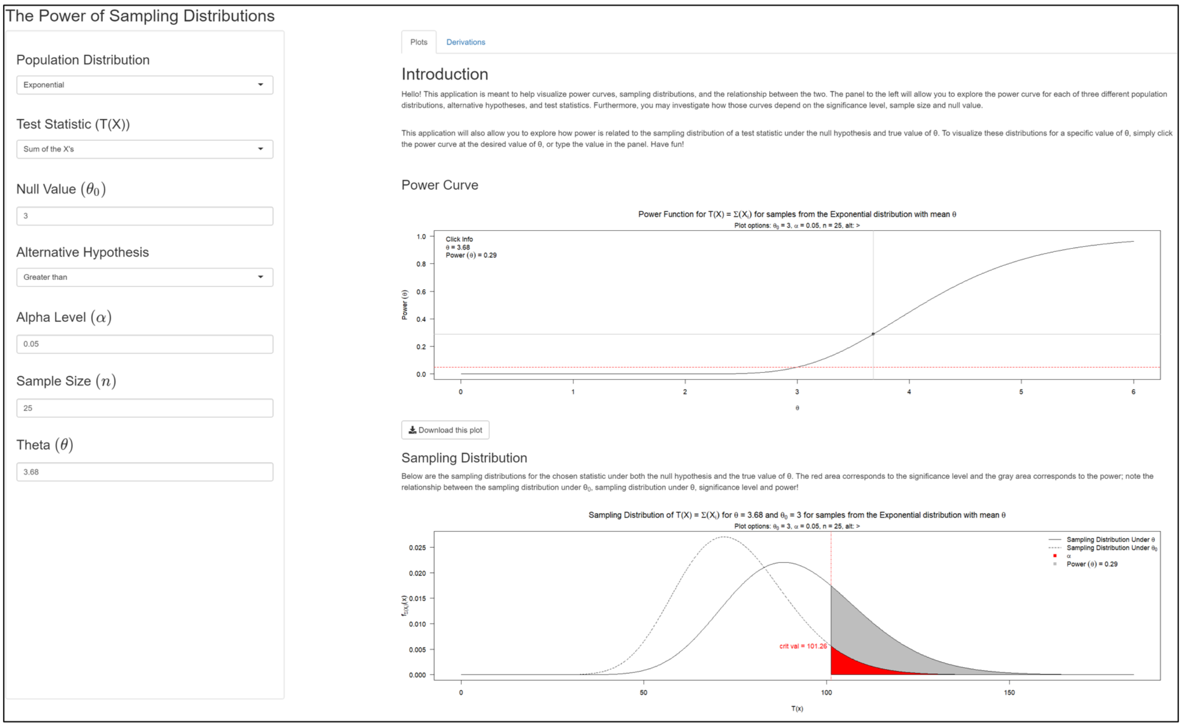


Figure 1: Interface for the web application. In the side panel, users may specify the population distribution, test statistic, null value, alternative hypothesis, significance level, and sample size. Based on these selections, the power curve is created in the main plotting window in real time.

The application takes advantage of R Shiny's reactive programming. Reactive programming allows the application to tailor output based on user input in real time. This is particularly useful when exploring power as it allows users to investigate the relationship between the power curve and the sampling distribution of the statistic in an interactive way. In this web application, the user is able to visualize the sampling distribution of the statistic for a particular value of θ by either specifying the value in the "Theta(θ)" box on the side panel of the application or clicking on a point on the graph of the power function (Figure 1). Note that this plot is not visible until θ is specified using one of these two options. Additionally, both plots respond in real time to any changes in the options provided on the side panel. This allows users to quickly assess the impact of any of those changes on both the power function and the sampling distributions, thereby helping them understand the relationship between the options on the sidebar, the power function, and the sampling distributions.

Many of the features of this application are compartmentalized to allow users to focus on specific topics rather than other potentially distracting features. For example, the conditionality of the sampling distribution plot allows users to explore relationships between the options on the sidebar and the power function before considering sampling distributions. In addition, this web application presents distribution-specific options. For example, specification of the population standard deviation, σ, is only available when the normal distribution is selected. A host of other options appear when both the uniform distribution and sum of random variables are selected, allowing the user to explore the numerical instability present in the distribution function associated with that statistic when sampling from the uniform distribution (see Section 3.3 for discussion). By consciously hiding certain options and tabs when they are not needed, the user is able to focus on learning without added distractions.

Additional features, such as download options for each of the created plots and tabs to work through the derivations of each power curve, make the application more user friendly. The download option supports easy sharing and collaborative explorations within a classroom. For example, in our experience students were asked to investigate a particular relationship (e.g., the relationship between sample size and the power function) and upload images supporting their findings to a shared Google document which was then used for class discussion. If desired, the derivations tab allows users to work through the derivations for each of the power functions provided in the plots tab (Figure 2). Although the focus of this application is on developing understanding through visualizations, derivations are included to help users build connections between the graphical and mathematical representations of power.

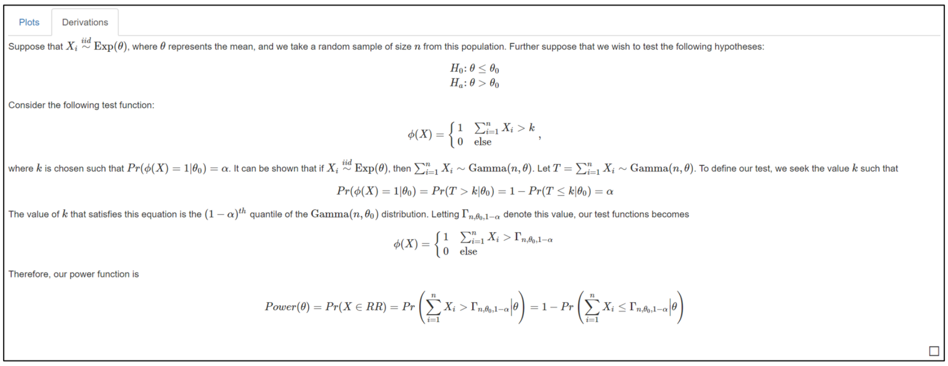


Figure 2: Power function derivations for a hypothesis test based on the sum of random variables for samples drawn from an exponential population assuming a greater than alternative hypothesis. The population distribution, test statistic, and alternative hypothesis are selected using a sidebar panel (not pictured).

3. APPLICATION EXPLORATIONS

An advantage of this web application is that it allows for creativity and flexibility in how the instructor teaches power. Through tabular panels and conditional plots, this application lets instructors choose which aspects of power to focus on, while still allowing students to explore at their own pace. This section describes how the application may be used to explore the learning objectives defined in Section 1. Section 3.1 focuses on the power curve; Section 3.2 focuses on how power curves arise by examining sampling distributions of test statistics under different conditions; and finally, Section 3.3 explores alternatives to traditional methods of determining power.

3.1. Exploring power curves

Statistical power is a complex concept that is influenced by several factors, including the population distribution, the test statistic, the hypothesis, the significance level, and the number of observations. This web application allows students to visualize relationships between the power curve and these factors without having to derive the power function or write the code required to plot that function. In this section, we illustrate how the web application may be used to draw connections between the power function and choice of test statistic. Other comparisons are omitted for sake of brevity, though instructors are encouraged to be creative in their use of this application.

In traditional mathematical statistics courses based on textbooks such as Casella and Berger (2002), students often use power curves to compare test functions based on different test statistics. In this web application, students can make such comparisons by selecting different options in the side panel. For example, Figure 3 provides a set of plots comparing each of the available statistics for the uniform and normal population distributions (assuming a null value of 3, significance level of 0.05, not equal to alternative hypothesis, and sample size of 25).

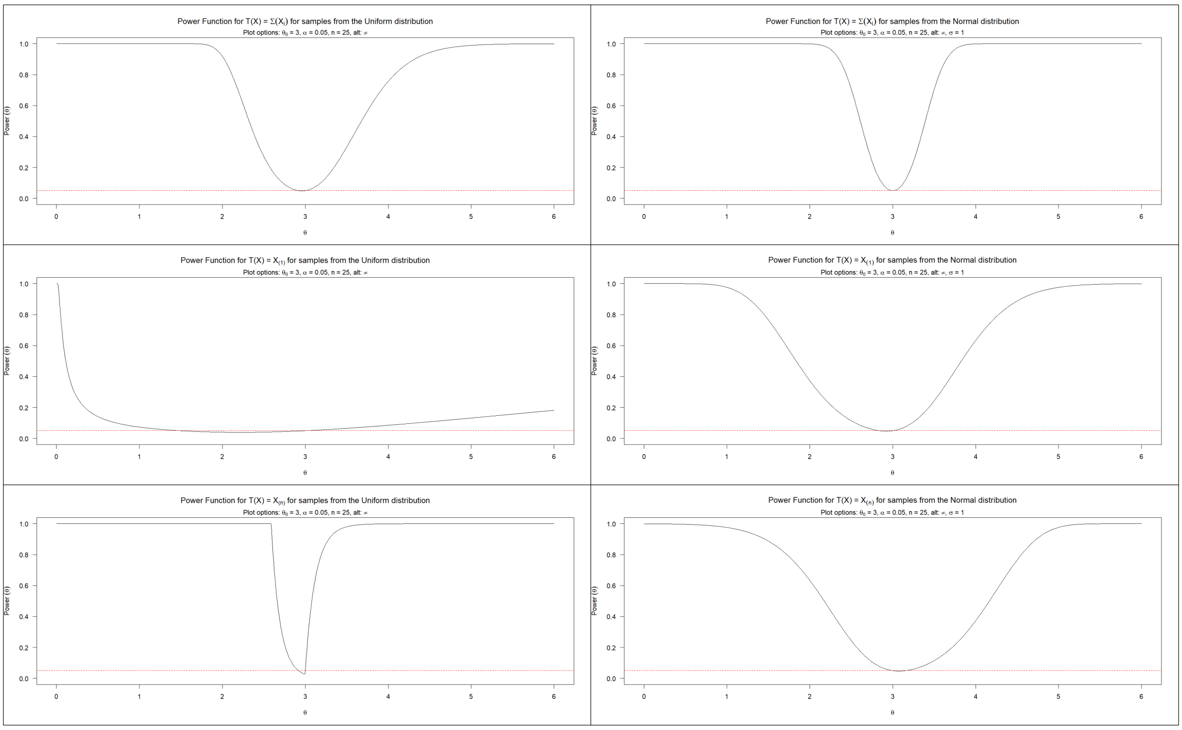


Figure 3: Comparison of power functions for varying statistics and population distributions. The left and right columns correspond to samples from the uniform distribution and the normal distribution, respectively. The top, middle, and bottom rows correspond to test functions based on the sum of random variables, sample minimum, and sample maximum, respectively.

From Figure 3, students can identify which power function is closest to ideal for each population distribution. This exploration helps students understand that the test statistic that provides the highest power in the alternative space depends upon the population distribution of interest. By choosing this particular combination of statistics and population distributions, students may be asked why the sum of random variables performs best for the normal distribution and why the sample maximum performs best for the uniform distribution. This avenue of exploration can lead to discussions of sufficiency in the context of power thereby helping students draw connections between power and previously explored topics.

Students can explore additional relationships by varying other options on the sidebar panel. In the next section, we examine how this application can also be used to explore sampling distributions of test statistics and understand the relationship between these sampling distributions and the power curve.

3.2. Exploring sampling distributions

One of the conceptually challenging aspects of learning power is understanding why the power function is a function of θ and how that function relates to the sampling distribution of the statistic under the null hypothesis. This relationship can be explored graphically, but doing so requires both deriving the sampling distribution of a statistic and plotting that sampling distribution with statistical software. Although working through the derivations and code can be a valuable experience for students, the task can also be time-consuming and difficult, often limiting the opportunities students have to consider a variety of examples. This web application allows users to view the sampling distribution of the chosen statistic under the null hypothesized value of θ and a user-specified value of θ for many different statistics and population distributions without having to work through any derivations.

With this web application, students may click multiple points along the power curve and describe the resulting plot under the sampling distribution heading. Through this process, students recognize that there is an implied sampling distribution of the statistic under both the null hypothesis and the true value of θ for each point along the power curve (Figure 4). Because these plots update in real time, students can click many points along the power curve to see how the sampling distribution of the statistic changes as θ, and subsequently the power, changes. Students are also able to visualize how the sampling distribution under the null hypothesis is static and does not depend on the chosen value of θ. This exploration is valuable as it makes the dependence of the power function on θ graphically explicit.

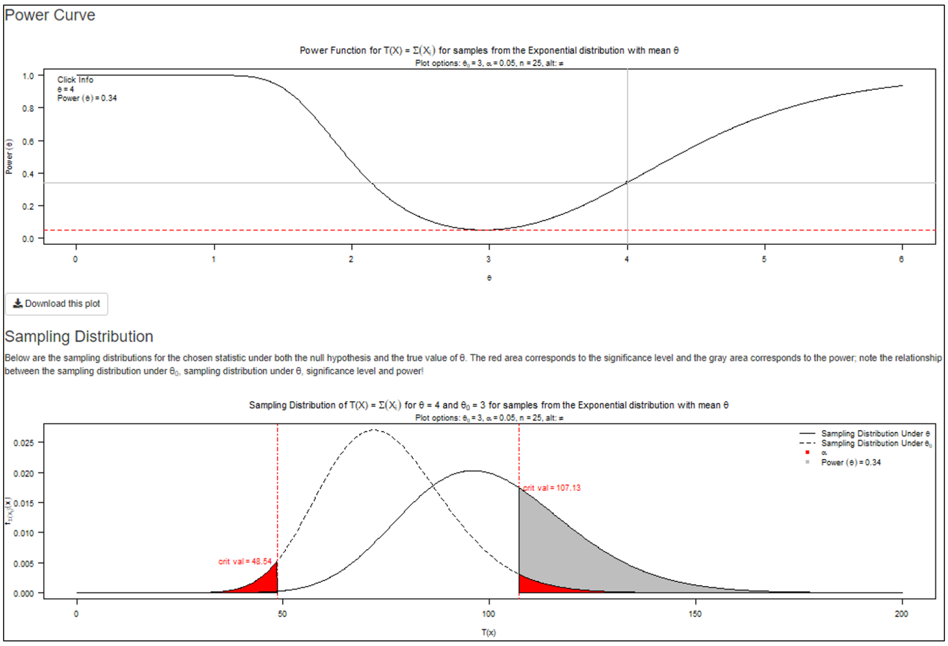


Figure 4: Power function and associated sampling distributions for a hypothesis based on the sum of random variables for a sample of size 25 from the exponential distribution assuming a not equal to alternative hypothesis, significance level of 0.05, and null value of 3.

These explorations can also help students understand how power is determined from the sampling distribution under the null hypothesis and the sampling distribution under the true value of θ. In Figure 4, the power is represented by the larger gray area and the significance level is represented by the smaller red area. From this graphic, students can visually assess how the significance level is used to determine the critical value, which is then used to calculate the power. Developing this conceptual understanding of how power is determined may help inform the derivation process for students and develop an intuition for how one might simulate power, which is further discussed in Section 3.3.

3.3. Exploring Normal approximation and simulation

Section 3.1 and Section 3.2 focused on combinations of population distributions and test statistics for which the power function is well behaved and readily obtainable. However this is not always the case when investigating power. In some cases, issues may arise that make it difficult to derive the formula or create a visualization for the power function. Learning how to resolve the issues that arise in these cases is valuable for students as it not only prepares them for potential challenges that they may encounter as a practicing statistician, but also improves their ability to think critically and reason quantitatively.

The web application currently contains one example of a power function that presents challenges when approached with traditional techniques. The power function for the sum of random variables from the uniform distribution is numerically unstable (see Appendix A for discussion). Consequently, it can be difficult - or in some cases impossible - to either calculate power or visualize the power function for this combination of population distribution and test statistic. As discussed below, this web application provides two potential approaches that practicing statisticians may explore when presented with a challenge like numerical instability.

Exploring the normal approximation

The Central Limit Theorem is the first option the web application offers students to resolve the numerical instability issue. Through normal approximation, students can approximate the sampling distribution of the sum of uniform random variables and therefore the power function. To implement this solution, students click the “Use normal approximation” check box that appears on the side panel for this combination of population distribution and test statistic.

Upon doing so, the power function and sampling distributions are replaced by those based on a normal approximation (Figure 5). This exploration is valuable for students because it gives them an opportunity to use previously discussed tools to resolve a problem, as a practicing statistician would.

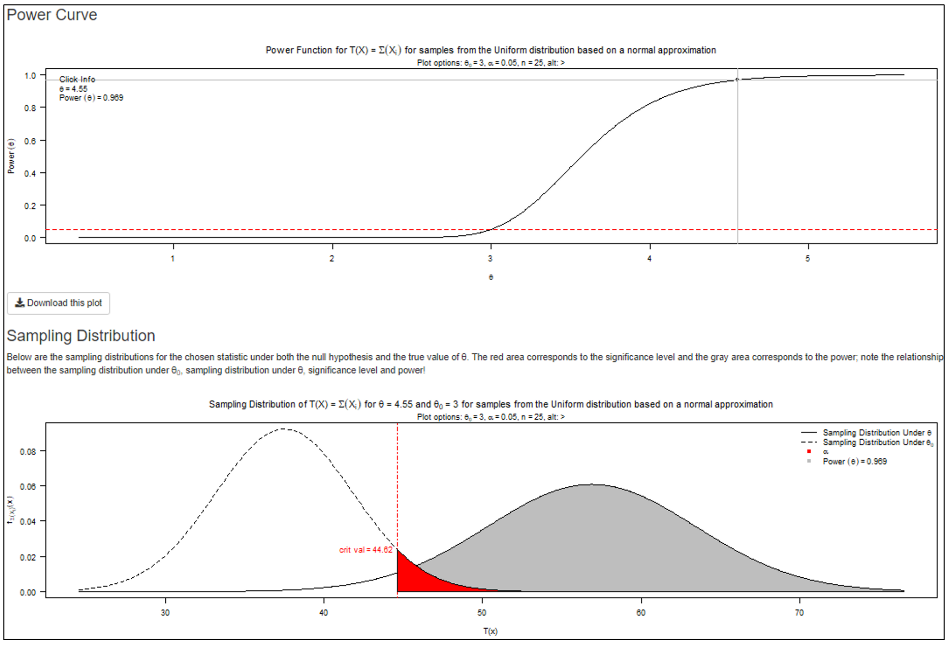


Figure 5: Power function and associated sampling distributions for a test function based on the sum of random variables for samples from the uniform distribution assuming normal approximation through the Central Limit Theorem.

The web application also provides students the option to evaluate how well the Central Limit Theorem approximates the sampling distribution of the sum of uniform random variables through the “Irwin-Hall Normal Approximation” tab. In this tab, students may simulate many samples of a desired size from a uniform$(0, \theta)$ distribution by specifying the number of simulated samples, sample size, and value of θ in the sidebar.

Students are then able to visualize the resulting empirical sampling distribution. Finally, students can assess the quality of the normal approximation by overlaying the approximate normal distribution function obtained through the Central Limit Theorem on the empirical distribution function (Figure 6). The features this tab offers are beneficial because students can assess the quality of the approximation used to produce the power function for this test, similar to how one should when using the Central Limit Theorem in practice. The tab's features also demonstrate how flexible a tool simulation can be when resolving difficult problems in statistics, which is further explored in the next subsection.

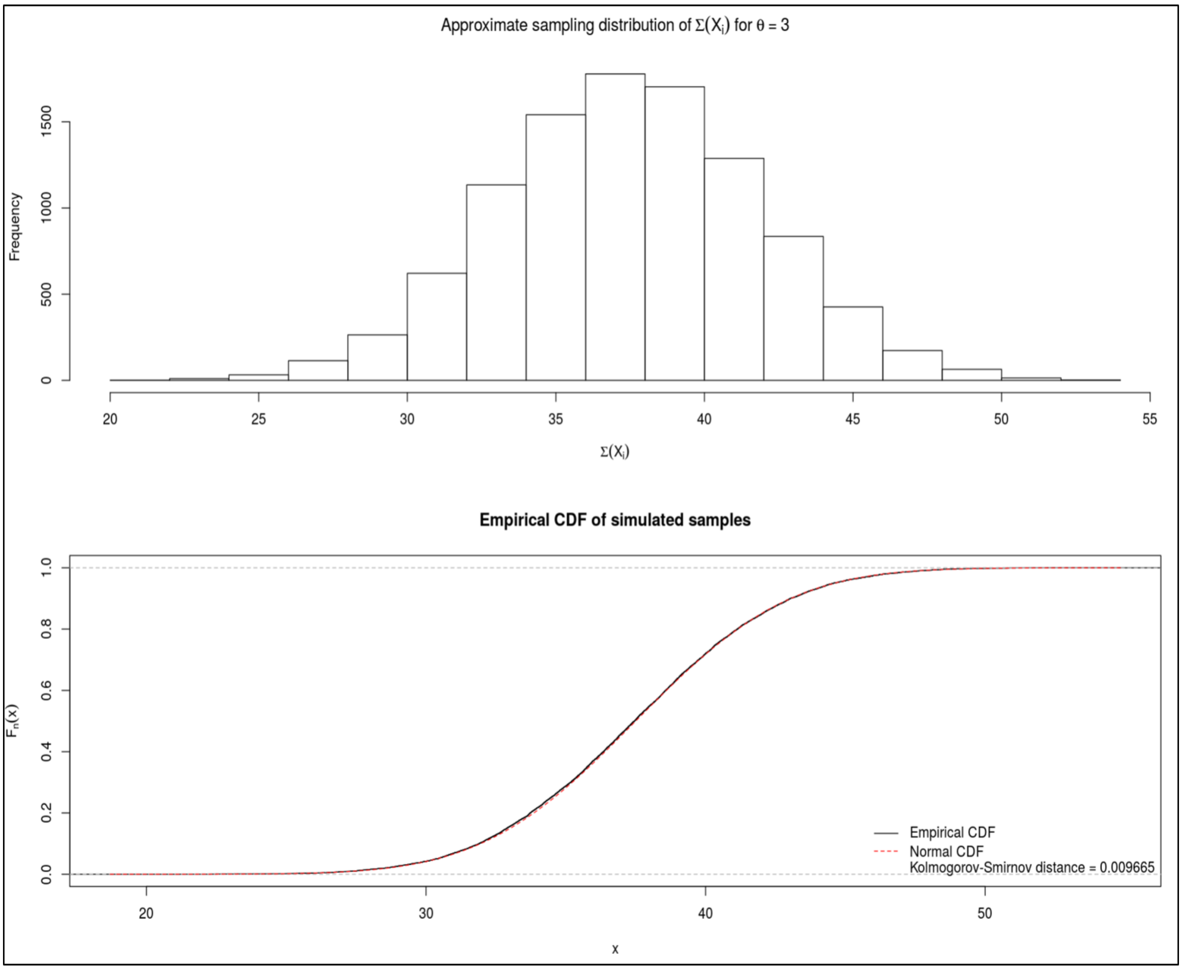


Figure 6: Diagnostic plots for the normal approximation using the Central Limit Theorem. The plot in the top row displays the approximate sampling distribution for the sum of random variables through a histogram and the bottom plot provides the empirical distribution function overlaid by the appropriate normal distribution function.

Exploring simulation

Students can also resolve the numerical instability issue by using simulation to approximate the sampling distribution of the sum of uniform random variables and the power function. The "Simulated Power" tab offers a sidebar with options to specify the number of simulated samples, sample size, null value, significance level (α), alternative hypothesis, and value of θ. Once specified, students may click the "Simulate power!" button to produce both the simulated power function and simulated sampling distributions under the null value and the specified value of θ (Figure 7).

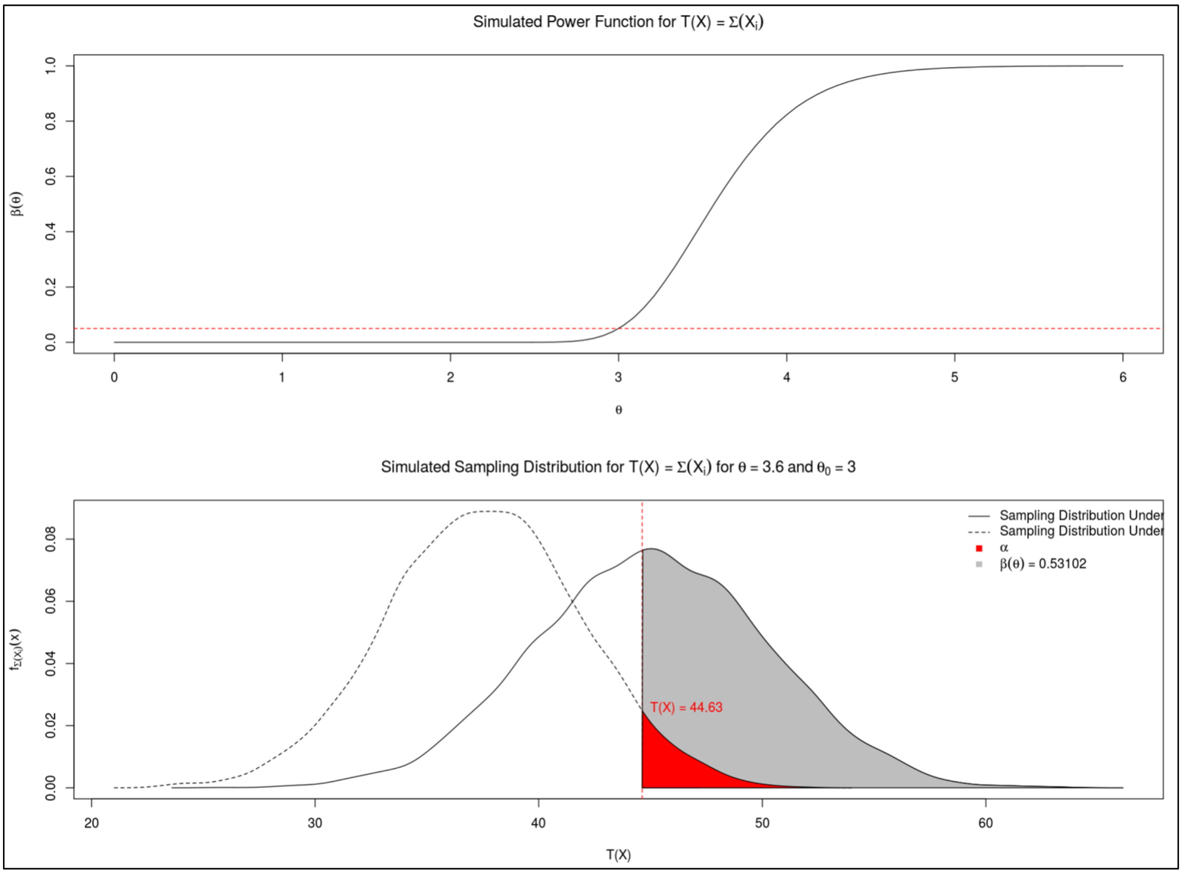


Figure 7: Simulated power function (based on 10000 simulations) and associated sampling distribution for a test based on the sum of random variables when sampling from a uniform population.

To create the simulated power function, this web application first creates a fine sequence of θ values along the horizontal axis. Then for each value of θ, it simulates the specified number of samples, each of size n, from a uniform$(0, \theta)$ distribution and then calculates the sum of simulated values for each of those samples (hereafter called the alternative sums); this process is repeated for the specified null value (hereafter called the null sums). Next, the critical value(s) is(are) determined by calculating the α^{th} percentile of the null sums for a "less than" alternative, the $(1-\alpha)^{th}$ percentile of the null sums for a "greater than" alternative, or the $(\alpha/2)^{th}$ and $(1-\alpha/2)^{th}$ percentiles of the null sums for a "not equal to" alternative. Finally, the simulated power is determined by calculating the proportion of alternative sums that are as or more extreme than the corresponding critical value(s) in the direction(s) of the alternative hypothesis. The simulated power at each value of θ is then plotted against θ to produce the simulated power function.

The simulated sampling distribution plot is created by repeating this process for only the value of θ specified on the sidebar and plotting the density of the resulting null and alternative sums. The shading is done by first determining the appropriate critical values as described above, then shading under each density in the appropriate direction(s) as determined by the alternative hypothesis. The value of this process is two-fold for students. First, students are able to understand through application how powerful of a tool simulation can be for resolving complicated problems in statistics. In many cases, practicing statisticians prefer simulation over analytical methods when conducting power analyses because of its flexibility and ease of implementation; this web application gives students the opportunity to witness this firsthand. Second, students develop a deeper conceptual understanding of the relationship between sampling distributions and power by considering how to simulate a power curve. Doing so requires stepping through the process described above, solidifying the understanding developed in the previous two subsections.

4. APPLICATION IMPLEMENTATION

In this section, we describe the first two authors' experience implementing the web application and encourage others to use this experience as a template for their own implementation. We implemented the web application with a guided activity (see Appendix B) in three classes across two different semesters: (1) in the spring of 2018 with both an undergraduate-level and graduate-level mathematical statistics class and (2) in the spring of 2019 with a graduate-level mathematical statistics class. After each semester, the application was revised based on feedback provided by the students, offering additional population distributions, test statistics, alternative hypotheses, and features to explore (see Table 1). In the following sections, we provide background information about the classes in which this application was implemented, describe the guided activity, and summarize student reflections about the application.

	Version 1 (Spring 2018)	**Version 2 (Spring 2019)**	**Version 3 (Current)**
Distributions	Exponential	Exponential, uniform, normal	Exponential, uniform, normal
Statistics	$\sum_{i=1}^n X_i$, $X_{(1)}$	$\sum_{i=1}^n X_i$, $X_{(1)}$, $X_{(n)}$	$\sum_{i=1}^n X_i$, $X_{(1)}$, $X_{(n)}$, $\bar{X}$
Alternatives	$<, >$	$<, >, \neq$	$<, >, \neq$
Additional features	None	Derivations tab	Derivations tab, CLT approximation tab, simulation tab

Table 1: Table of web application versions. Student feedback was used to revise each version of the application; version 3 is the current version described in Section 3.

4.1. Background context

The application was used in the second semester of both an undergraduate-level and graduate-level two-semester sequence of courses in mathematical statistics. The first semester of each sequence of courses focuses on building the foundation of probability theory, addressing traditional topics such as probability, random variables, named probability distributions, and multiple random variables. The graduate-level course discusses these topics in greater depth, using the Casella and Berger (2002) text while the undergraduate-level course uses the Wackerly et al. (2008) text.

In the second semester, when this web application was implemented, the courses continue to address similar topics such as sampling distributions, point estimation, hypothesis testing, and interval estimation, but the graduate-level course also covers asymptotics. In both courses, students discuss concepts of sufficiency and best unbiased estimation before investigating hypothesis testing and power during the last six weeks of the semester.

Both the undergraduate-level and graduate-level courses include students with diverse mathematics and statistics backgrounds, typically including students majoring in engineering, economics, and physics in addition to those seeking mathematics or statistics degrees. Prerequisites for the undergraduate-level course include a course in multivariate calculus, with a course in mathematical proof also recommended. Prerequisites for the graduate-level course include completion of the full undergraduate sequence (or a similar undergraduate sequence in mathematical statistics), though exceptions are made at the discretion of the instructor.

In all three cases where this application was implemented, students were primed to use the application by walking through a derivation of one of the power functions featured on the web application in the preceding class. In the following two 50 minute-long classes, students were asked to explore various properties of power using the web application through a guided activity, which is discussed in the next section.

4.2. Activity layout

The same guided activity (Appendix B) was used across all three implementations, and was built around an example assuming an exponential population distribution. On the first day of each implementation, students were divided into five small groups and asked to explore one of the five questions provided in Part I of the activity. Students were then asked to summarize their group's exploration and give a short three-minute to five-minute presentation on what they discovered at the end of class. The structure of the second day differed between the undergraduate-level and graduate-level implementations. In the undergraduate-level course, the instructor led an exploration of the relationship between the power function and sampling distribution using the web application at the beginning of class, after which students openly explored other relationships using the application. In the graduate-level implementations, students openly explored the relationship between the power function and sampling distribution for the first half of class and discussed what they discovered for the second half of class. In the Spring 2019 implementation, students also explored additional population distributions which were newly available in that version of the application.

Following all three implementations, students were asked to answer a set of reflection questions (Appendix C). The Spring 2019 cohort was asked an extra set of questions that related to the additional distribution and test statistic options. These responses were used to guide a discussion at the beginning of class on the third day and clarify any lingering questions; the remainder of that third class was used to derive the power function for one of the examples provided on the web application. The students' responses were also used to revise each version of the web application and assess the strengths and opportunities of the web application.

4.3. Student reflections

Students' responses to the reflection questions were reviewed for each class at each implementation and the key points were summarized, paying special attention to comments about how the application helped or hindered their understanding of the material. Then, these summaries were compared across all three implementations to identify commonalities and differences between them. There were four types of comments common to all three implementations: (1) descriptive statements about how the options on the sidebar panel influenced the power curve; (2) comments about how the application allowed students to quickly visualize changes in the power curve; (3) comments about how the application helped students understand the role of the sampling distribution in calculating power; and (4) comments relating power to previously discussed concepts, such as sufficiency. In the following subsections, each of these types of comments are discussed in detail and example student responses are provided.

Descriptive statements

The first type of comments consisted of descriptive statements about how the options on the sidebar panel influenced the power curve. These types of comments did not directly reference how the web application influenced their understanding of power, but instead provided indirect evidence of the web application's effect through descriptive statements. A student in the 2018 undergraduate cohort noted, "The power function [varied] greatly depending on the test statistic used." Another student in the 2018 graduate cohort discovered that as "the sample size increases, the power curve becomes steeper." There were many other similar comments describing fundamental relationships between the options available on the sidebar panel and the power curve. This type of feedback was encouraging, as understanding how the factors on the sidebar panel affect the power curve is one of the learning objectives listed in Section 1.

Visualizing changes in the power curve

Across all three implementations, students also wrote about how the application allows them to quickly visualize changes in the power curve. Many students noted that the application's ability to render the power curve in real time as they changed options on the side panel helped cement their understanding of the effect of each of those factors on the power curve. These comments provided direct feedback about how the application influenced their learning of power, as opposed to the indirect feedback discussed previously. A student in the 2018 undergraduate cohort stated, “It was really helpful to be able to hold some inputs constant, while varying others and see how the power curve was affected.” Another student in that cohort noted that it was helpful to “change a parameter and immediately watch the distribution/power curve change with the parameter.” Many students in the graduate cohorts echoed this sentiment. One student in the 2019 cohort felt that “being able to adjust the things we control as researchers and seeing the instant feedback” was helpful. Overall, many students thought that being able to immediately visualize change in the power function and sampling distributions as they changed inputs on the side panel was beneficial.

The role of the sampling distribution

In addition, students frequently described how the application helped them understand the role of the sampling distribution in calculating power as well as how the sampling distributions are used to determine the critical value(s) of a test. Many students wrote that the web application helped them understand how the sampling distribution of a statistic is related to the power function and how the power function changes across test statistics. For example, a student in the 2018 undergraduate cohort observed, “The sampling distribution of $X_{(1)}$ hardly changes with increased sample size, so power does not really change for $X_{(1)}$ with increased sample size which is vice versa of the $\sum_{i=1}^{n} X_i$ statistic.” Another student in that cohort noted that she “really liked seeing the relationship between the null and observed sampling distributions... It really solidified the ideas behind power and critical values.” Overall, students thought the visuals provided by this web application helped them understand different aspects of deriving a power function, including determining critical values. This understanding was mirrored in the graduate cohort. A student in the 2019 cohort described how she used the application to explore the relationship between sampling distributions and power in the following way:

> One aspect I found very helpful was being able to select various values of θ and seeing how this changed the sampling distribution. Not only did this connect power to the area under the distribution depending on α and θ but it also allowed us to make connections between power and the variance [of the test statistic].

This type of feedback was again encouraging, as it suggested that this web application helped students better understand the relationship between the sampling distribution and the power curve, which is the second learning objective provided in Section 1.

Relating power to sufficiency and other concepts

The final type of comments students frequently made focused on interesting relationships they noticed between the power function and previously discussed concepts. For example, students across all three cohorts noticed that power functions based on sufficient statistics tended to provide higher power across the alternative space than those based on other statistics. Two students in the 2018 undergraduate cohort noted that “it is better to use a sufficient statistic to obtain higher power," and that "when a sufficient statistic is used the power is higher in the alternative space." We again saw similar discussions in the graduate cohorts. One student in the 2018 cohort provided the following reflection:

> I find the relationship between complete, sufficient statistics and power intriguing. Are there ways to prove that maximum power is realized when using a complete, sufficient statistic? Is this always the case? For all alpha? In general, what is the relationship between complete, sufficient statistics, uniform minimum variance unbiased estimators, and power?

The value in these student reflections is two-fold. It suggests that students are using this application to think critically about power and its relationship to previously discussed concepts, and it also lays the foundation for comparing test functions, which is the topic that follows power. In addition to sufficiency, some students sought to explore other relationships. One graduate student in the 2018 cohort described his exploration in the following way:

> I also found exploring the 'extreme' cases to be interesting. For example, what happens when our alternative value of θ is less than θ_0, but we have a greater than alternative hypothesis? Interestingly, we still have power greater than 0 due to random chance.

In this reflection, the student used the web application to understand why there is a positive probability of rejecting the null hypothesis over the null parameter space. This kind of feedback is indicative of the insight students may gain by using this web application to explore their curiosities.

In the reflection questions, students were also asked to provide suggestions about how the web application could be improved. In the Spring 2018 implementation, students in both the undergraduate-level and graduate-level courses noted that it would be useful to consider more population distributions, additional test statistics, and to explore a not equal to alternative hypothesis. These suggestions led to the development of the second version of the web application, which added the normal and uniform distributions, sample maximum as a test statistic, and not equal to alternative hypothesis (Table 1). When using this updated version in the Spring 2019 implementation, students' comments then included distribution-specific observations. For example, one student noted, "Changing theta doesn't affect the variation of the normal distribution whereas it does affect the exponential distribution." This feedback was exciting to see, as it suggested that the new options available in the second version of the web application helped students gain additional insight into what factors affect the power function. Another student from that cohort summarized this effect: "[The additional options] helped me see that power curves don't always look similar and that changing the distribution will drastically change the form of $\beta(\theta)$."

Students were again asked for suggestions to improve the web application in the 2019 implementation. There were few requests for changes, though one student did suggest we add additional distributions to the application, including discrete distributions; this is a potential area of further work. Instead, additional features such as the normal approximation and simulation tabs were added. These features added more flexibility for exploring and building connections across a variety of topics, including simulation, the Central Limit Theorem, and strategies for practicing statisticians when conducting power analyses.

5. STUDENT ASSESSMENT

The web application was designed to help promote students' conceptual understanding of power through the use of visualizations. In this section, we provide examples of final exam questions that were used to assess student learning at both the undergraduate-level and the graduate-level (Appendices D and E, respectively). Additionally, we describe how each question relates to the types of understanding the web application encourages, highlighting the many aspects of power addressed. Collectively, these questions help illustrate the depth and types of understanding we expect students to gain from interacting with the web application.

5.1. Undergraduate-level assessment

The undergraduate-level prompt (Appendix D, meant to be completed in-class) describes a hypothesis test based on the sample maximum, $X_{(n)}$, of a random sample of Poisson random variables with unknown mean μ. The prompt also provides a graph of a power function for that test. Students are then asked to complete three parts. Part *a* asks students to explain why a Poisson distribution is an appropriate model for the random variables described in the prompt; part *b* asks students to deduce the set of hypotheses tested based on the graph of the power curve; and part *c* asks students to sketch a power function for a test based on the sum of random variables, $\sum_{i=1}^{n} X_i$, and describe how that power function relates to the power function for the test based on $X_{(n)}$. For a more thorough description of the prompt, see Green and Blankenship (2015).

Parts *b* and *c* of the question directly relate to the types of understanding the web application encourages. Part *b*, which concerns deducing the hypotheses of interest based on a graph of the power function and the scenario's description, targets students' ability to ascertain characteristics of the power function based on a graphical representation of that function. The web application was specifically designed to encourage students to understand the relationship between a power function and the factors that affect it by visualizing the impact of each of those factors on the power function. In regards to part *b* of this exam question, students using the web application have the opportunity to vary the alternative hypothesis of a test and visualize the effect of those changes on the power function in real time. Part *c* concerns graphically comparing the power function for two different test statistics and justifying the differences between those power functions. The web application encourages this type of thinking by allowing students to select multiple test statistics for each distribution offered; furthermore, the guided activity developed to accompany this web application (discussed in Section 4) encourages students to think deeply about the relationship between a test statistic and the power function.

5.2. Graduate-level assessment

The graduate-level prompt (Appendix E, meant to be completed as a "take-home" exam) also describes a hypothesis test based on the sample maximum, $X_{(n)}$, of a sample of independent Poisson distributed random variables and provides a graph of the power function for that test. However, this prompt has an additional layer of complexity; each random variable, X_i, is distributed Poisson($a_i\theta$), where θ is the unknown mean rate and a_i is a known constant that may differ for each of the n observations. Students are asked to complete seven parts. Parts a and b are identical to the undergraduate-level prompt, except students are also asked to describe what $a_1, \ldots, a_n$ represent in the context of the problem. Parts c and d ask students to derive the power function for the test and explain how the significance level of the test was determined. Part e asks students to sketch the power function of a new test based on the sum of random variables, $\sum_{i=1}^{n} X_i$, similar to part c of the undergraduate-level prompt. Finally, part f requires that students use simulation to obtain an empirical power function, and part g asks students to conduct the hypothesis test based on provided data.

Many of these questions relate to the types of understanding the web application encourages. For example, parts b and e assess the same types of understanding as their analogous questions in the undergraduate-level prompt; the discussion provided in Section 5.1 also applies to these parts. Part c focuses on the derivation of a power function, which is a common learning objective of graduate-level mathematical statistics curricula. Although the web application was primarily created to help students visualize the complex relationships that underlie power without need for derivation, the application also offers students numerous derivations through the "Derivations" tab. These derivations illustrate the process of deriving different power functions, offering students options to explore power functions both visually and theoretically if desired.

For part d, the primary goal is to have students consider how the maximum probability of rejecting the null hypothesis over the null parameter space changes for different critical values when the test statistic ($X_{(n)}$) is discrete instead of continuous. In this case, the size of the test with critical value c is a step function over the different possible values one could choose for c. For the scenario described in the prompt, it is not possible to choose a critical value that would lead to a test function size exactly equal to 0.07. Though the web application currently concerns only continuous distributions, it can be used to foster discussions about how the critical value for a size-alpha test is obtained. The application draws attention to many of the components required to determine the critical value for a size-alpha test, including the distribution from which the critical value is determined and the relationship between the significance level and the critical value. By focusing on these relationships, the web application can be leveraged to lay a foundation for understanding the key concepts needed to explore similar questions within a discrete setting.

Part f, which has students use simulation to obtain an empirical power function for the test function based on $\sum_{i=1}^{n} X_i$, targets students' ability to dissect the process of how each point in a power function arises. The web application can be used to explore this type of thinking in multiple ways. For example, by clicking on a point on a power function under the "Plot" tab, students can visualize the corresponding distributions of the test statistic under both the null hypothesized value of θ and chosen value of θ. The web application provides these distributions for each point a student selects, allowing students to compare the resulting sampling distributions for different values of θ and to consider how each point on the power function is determined. Additionally, when the Uniform population and $\sum_{i=1}^{n} X_i$ test statistic are selected, the web application offers additional tabs showcasing how simulation can be used to create an empirical power function, reinforcing the process of obtaining each point on the power curve.

Finally, part g has students conduct a hypothesis test based on sample data that is given to them, and make a decision based on a p-value. While not explicitly emphasized through the web application, a p-value is obtained by calculating tail probabilities (consistent with the alternative hypothesis) of the sampling distribution of the test statistic under the null hypothesis. These sampling distributions are readily available through the web application at the bottom of the "Plot" tab, allowing students to explore a visual representation of how p-values are calculated. In our experience, using the null sampling distributions in this manner led to engaging discussions about how the critical value for a test and a p-value are obtained, and about what each quantity represents.

Together, these prompts help exemplify the depth and types of understanding the web application is meant to promote for both undergraduate-level and graduate-level students, as well as how these types of understanding can be assessed. In the future, we intend to use similar assessments to help us explore how the web application influences students' understanding of power.

6. CONCLUSIONS

In this article, we described a web application that can be used to visualize many different aspects of power - such as the factors that affect power and the relationship between power and the sampling distributions of the test statistic across the null and alternative parameter spaces - for a variety of different population distributions and test statistics. In doing so, class time typically spent on creating those visualizations can instead be used to allow students to explore the intricate relationships that make power so conceptually rich. This exploration provides students with a more dynamic and authentic experience, as it does not depend on a subset of graphics created by either the student or the instructor *a priori*.

As evidenced by student feedback, this web application encourages students to develop a conceptual understanding of statistical power in which they can recognize the relationship between the power function and the sampling distribution of a statistic and draw connections between power and previously covered topics. We recommend this application be paired with a guided activity such as the one provided in Appendix B; additional questions that can be used to target specific aspects of power are provided in Appendix F. Users can access a running version of the application at https://shiny.stt.msu.edu/jg/powerapp/.

In the future, we intend to conduct further research about how this application influences students' learning of power, as well as how other instructors adapt the application. Source code for this application is provided on our GitHub page (available at https://github.com/StrattonCh/PowerApp/tree/master/PowerApp) so that others may tailor it to their own needs. By showcasing the flexibility and versatility of this web application in teaching power in a modern mathematical statistics curriculum, we hope to encourage others to harness the power of technology when approaching conceptually difficult topics in their own curriculum.

7. ACKNOWLEDGMENTS

We thank all the students who have interacted with the web application and provided their feedback and suggestions to help improve it. We also thank the reviewers and editor for their insightful comments. Finally, we thank Dr. Katharine Banner and Tim Hansen for their feedback and suggestions on improving this paper.

6. REFERENCES

Aberson, C. L., Berger, D. E., Healy, M. R., and Romero, V. L. (2002), "An Interactive Tutorial for Teaching Statistical Power," *Journal of Statistics Education*, 10, 1–7, https://doi.org/10.1080/10691898.2002.11910682.

American Statistical Association (2014), *2014 Curriculum Guidelines for Undergraduate Programs in Statistical Science*, American Statistical Association, https://www.amstat.org/asa/files/pdfs/EDU-guidelines2014-11-15.pdf.

— (2016), *Guidelines for Assessment and Instruction in Statistics Education (GAISE): College report*, American Statistical Association, https://www.amstat.org/asa/files/pdfs/GAISE/GaiseCollege_Full.pdf.

Anderson-Cook, C. M. and Dorai-Raj, S. (2003), "Making the Concepts of Power and Sample Size Relevant and Accessible to Students in Introductory Statistics Courses Using Applets," *Journal of Statistics Education*, 11, 1–12, https://doi.org/10.1080/10691898.2003.11910721.

Bobek, E. and Tversky, B. (2016), "Creating Visual Explanations Improves Learning," *Cognitive Research: Principles and Implications*, 1, 1–14, https://doi.org/10.1186/s41235-016-0031-6.

Casella, G. and Berger, R. L. (2002), *Statistical Inference*, Brooks/Cole Cengage Learning, 2nd ed.

Chance, B., Ben-Zvi, D., Garfield, J., and Medina, E. (2007), "The Role of Technology in Improving Student Learning of Statistics," *Technology Innovations in Statistics Education*, 1, 1–26, https://escholarship.org/uc/item/8sd2t4rr.

delMas, R., Garfield, J., and Chance, B. (1999), "A Model of Classroom Research in Action: Developing Simulation Activities to Improve Students' Statistical Reasoning," *Journal of Statistics Education*, 7, 1–16, http://www.amstat.org/publications/jse/secure/v7n3/delmas.cfm.

Doi, J., Gail, P., Wong, J., Alcaraz, I., and Peter, C. (2016), "Web Application Teaching Tools for Statistics Using R and Shiny," *Technology Innovations in Statistics Education*, 9, 1–33, https://escholarship.org/uc/item/00d4q8cp.

Green, J. L. and Blankenship, E. E. (2015), "Fostering Conceptual Understanding in Mathematical Statistics," *The American Statistician*, 69, 315–325.

Post, J. (2016), *Interactive Math Stat Visualizations Using R Shiny*, poster presented at the Electronic Conference on Teaching Statistics. Available at https://www.causeweb.org/cause/ecots/ecots16/posters/c/8.

Rossman, A. and Chance, B. (2004), *The Rossman/Chance Applet Collection*, Available at http://www.rossmanchance.com/applets/.

Wackerly, D. D., Mendenhall III, W., and Scheaffer, R. L. (2008), *Mathematical Statistics with Applications*, Brooks/Cole Cengage Learning, 7th ed.

Wild, C. J., Pfannkuch, M., Regan, M., and Horton, N. J. (2011), "Towards More Accessible Conceptions of Statistical Inference," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174, 247–295, https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-985X.2010.00678.x.

7. APPENDIX

A. General Irwin-Hall Distribution Derivations and Numerical Instability

The General Irwin-Hall Distribution Derivations

Before discussing the numerical instability that is present in the distribution function of the sum of uniform(0, θ) random variables, we must first derive that function. It can be shown that if $X_i \overset{iid}{\sim}$ Unif$(0,1)$ and we take a random sample of size n from this population, then $Y = \sum_{i=1}^{n} X_i \sim$ Irwin-Hall(n) (Marengo et al. 2017), where Y has the following density and distribution functions:

$$f_Y(y) = \begin{cases} \frac{1}{(n-1)!}\sum_{k=0}^{\lfloor y \rfloor}(-1)^k\binom{n}{k}(y-k)^{n-1},\ y \in \mathbb{R} & 0 < y < n\theta \\ 0 & \text{else} \end{cases}$$

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ \frac{1}{n!}\sum_{k=0}^{\lfloor y \rfloor}(-1)^k\binom{n}{k}(y-k)^{n} & 0 \le y \le n\theta \\ 1 & y > n\theta \end{cases}$$

We desire the distribution of the sum of n independent, uniform$(0,\theta)$ random variables. If $W_i \overset{iid}{\sim}$ Unif$(0,\theta)$, then $\sum_{i=1}^{n} W_i$ has the same distribution as θY, where $Y \sim$ Irwin-Hall(n). Then for $T = \sum_{i=1}^{n} W_i$,

$$Pr(T \le t) = Pr(\theta Y \le t) = Pr\left(Y \le \frac{t}{\theta}\right)$$

Therefore,

$$F_T(t) = F_Y\left(\frac{t}{\theta}\right)$$
$$f_T(t) = \frac{1}{\theta} f_Y\left(\frac{t}{\theta}\right)$$

Numerical Instability

For even modest sample sizes, this distribution function exhibits numerical instability that manifests in the form of a strange oscillating pattern (see Quijano Xacur 2019 for more detail); the instability is exacerbated for larger sample sizes. This instability arises because the density function is a piece-wise polynomial function (spline) of degree $n-1$, where each piece of the spline can be represented by a power series with coefficients that alternate in sign. As a result, the density at any point is numerically calculated by summing an alternating series. For density values close to zero or one, the precision with which the density can be calculated is limited by the precision of the statistical software used to plot the density function. This phenomenon is responsible for the aforementioned oscillating trend that may appear in the distribution function, and therefore in the power function.

The web application provides an opportunity to explore the issue of numerical instability. By clicking on the warning information that appears when the uniform distribution and sum of random variables is selected, students obtain additional plots and information (Figure 8). The error message that appears allows students to zoom in on the numerical instability if the oscillating trend is not obvious in the power function. The message also describes two alternative ways to calculate the power of this test: (1) through normal approximation using the Central Limit Theorem or (2) through simulation (discussed in Section 3.3)

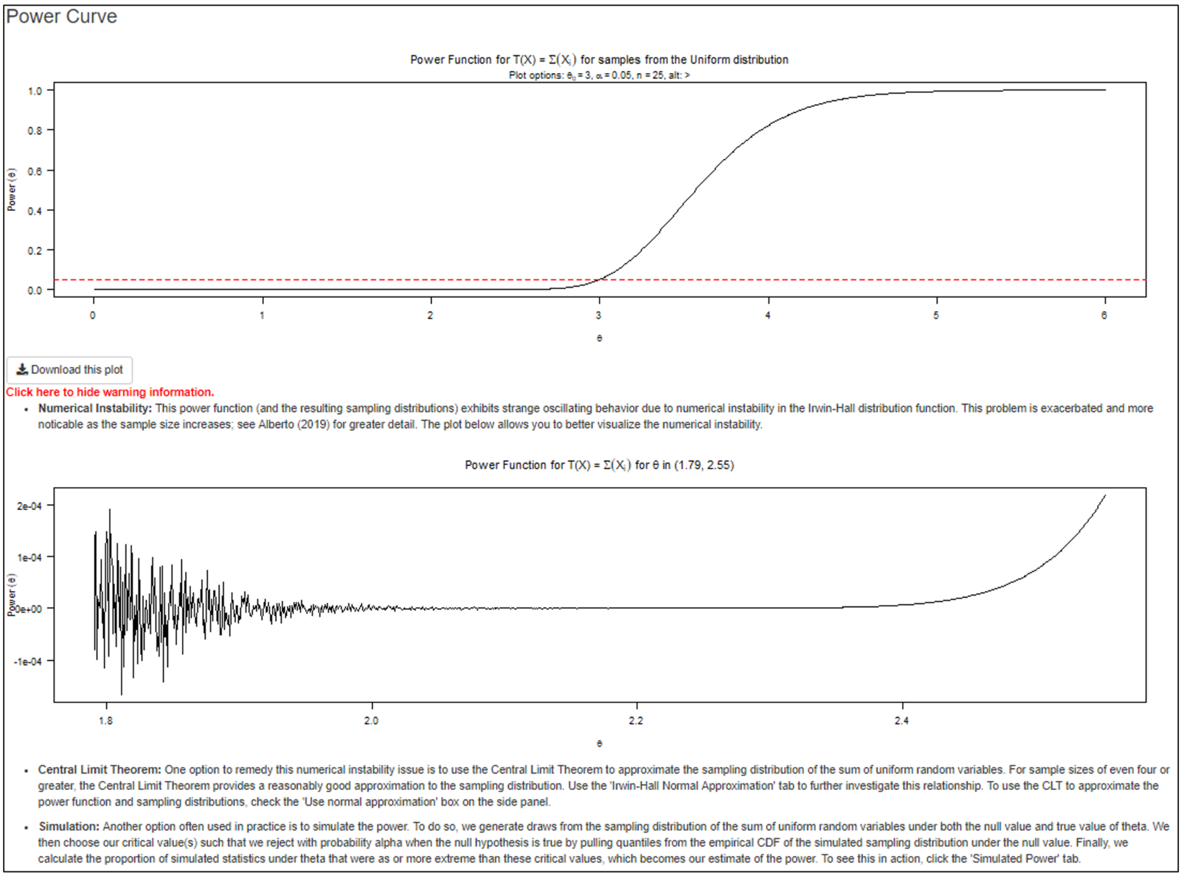


Figure 8: Toggleable error message notifying the user that numerical instability is present. The plot allows the user to "zoom" in on the numerical instability, and the text proposes potential solutions.

REFERENCES

Marengo, J., Farnsworth, D., and Stefanic, L. (2017). A geometric derivation of the irwin-hall distribution. *International Journal of Mathematics and Mathematical Sciences*, 2017:1–6. https://doi.org/10.1155/2017/3571419.

Quijano Xacur, O. A. (2019). *unifed: The Unifed Distribution.* R package version 1.1.1.

B. Power Application Activity

Objective: In this activity, each group is going to explore power curves and how these curves compare for different test statistics, null values, alternative hypotheses, alpha levels, and sample sizes. We will also explore how these functions arise, relating them to the sampling distribution of the test statistic.

Context: Annual loss (in hundreds of dollars) is assumed to be an Exponential random variable with mean θ. In a small metropolitan area, the average annual loss (in hundreds of dollars) due to theft has been assumed to be θ_0. Suppose the police chief is interested in testing hypotheses about the mean annual loss. A random sample of n annual losses due to theft is taken, and two hypothesis tests are proposed, one based on the test statistic, $X_{(1)}$, and one based on the test statistic, $\sum_{i=1}^{n} X_i$.

Directions: Use the power applet (https://christianstratton.shinyapps.io/PowerApp/) to explore the following questions. We will discuss questions 1-5 as a class and then dig deeper into our explorations with questions 6-7.

Part I: Varying Test Statistics, Null Values, Alternative Hypotheses, Alpha Levels, and Sample Sizes

---

1) Holding all other values constant, how does the power curve for a test based on the test statistic, $X_{(1)}$, compare to the power curve for a test based on the test statistic, $\sum_{i=1}^{n} X_i$? Why do you think this is the case?

2) Holding all other values constant, what happens to the power curve as the null value changes? Why?

3) Holding all other values constant, what happens to the power curve if the alternative hypothesis changes? Why?

4) Holding all other values constant, what happens to the power curve as the alpha level changes? Why?

5) Holding all other values constant, what happens to the power curve as sample size changes? Why?

Part II: Connecting Power Curves to Sampling Distributions

---

6) What is the power curve representing over the parameter space, Θ? How is each point on the power curve obtained? To explore this, let's consider the following questions:

Scenario 1: Fix $\theta_0 = 3, \alpha = 0.05$, and $n = 25$ to test $H_1 : \theta > \theta_0$ using the following test function: $\phi_S(\boldsymbol{X}) = \begin{cases} 1 & \sum_{i=1}^{n} X_i > k \\ 0 & \text{else} \end{cases}$.

- Click on the point on the power curve corresponding to $\theta = 3$.
    - Where is each sampling distribution centered? Why?
    - What is k for this situation? How is it determined?
    - How does the power at $\theta = 3$ relate to the significance level? Why?

- What happens to the power when $\theta > 3$? Click on at least one more point on the power curve corresponding to a value of θ greater than 3. For each point, reflect on the following questions:
  - How does the sampling distribution used to calculate power at this value of θ change?
  - What happens to the sampling distribution under θ_0?
  - What is k? How does this relate to the value of k found earlier? Why?
  - How is power calculated for this value of θ? How does it compare to the significance level? Why?
- What happens to the power when $\theta < 3$? Click on at least one more point on the power curve corresponding to a value of θ less than 3. For each point, reflect on the following questions:
  - How does the sampling distribution used to calculate power at this value of θ change?
  - What happens to the sampling distribution under θ_0?
  - What is k? How does this relate to the value of k found earlier? Why?
  - How is power calculated for this value of θ? How does it compare to the significance level? Why?
- Why is power increasing as the distance between θ and θ_0 increases over the alternative parameter space? Why is power decreasing as the distance between θ and θ_0 increases over the null parameter space?

<u>Other Scenarios</u>: Consider the questions for Scenario 1 for different scenarios of interest. What do you notice? How is each point on the corresponding power curve obtained? How does this differ across different scenarios?

7) Relating back to your explorations earlier in questions 1-5, how do different values of the test statistic, null value, alternative hypothesis, alpha level, and/or sample size change the power curves? How does this relate to the sampling distribution of the test statistic?

C. Reflection Questions

The following reflection questions were asked of all students that used this web application in class. Note that only the students from the Spring 2018 implementation - who had access to a version of the application that featured additional population distributions and statistics - were asked question 3.

1) What did you discover or notice as you explored how different values of the test statistic, null value, alternative hypothesis, alpha level, and/or sample size change the power curves, and how this all relates to the sampling distributions of the test statistics?

2) How did the applet for the exponential distribution help or hinder your understanding of power?
   - What aspects of the applet were helpful? Why?
   - What aspects weren't as helpful and/or could be improved? Why?

3) How did looking at other population distributions, statistics, and alternative hypotheses help or hinder your understanding about power?
   - What aspects of this were helpful? Why?
   - What aspects weren't as helpful and/or could be improved? How?

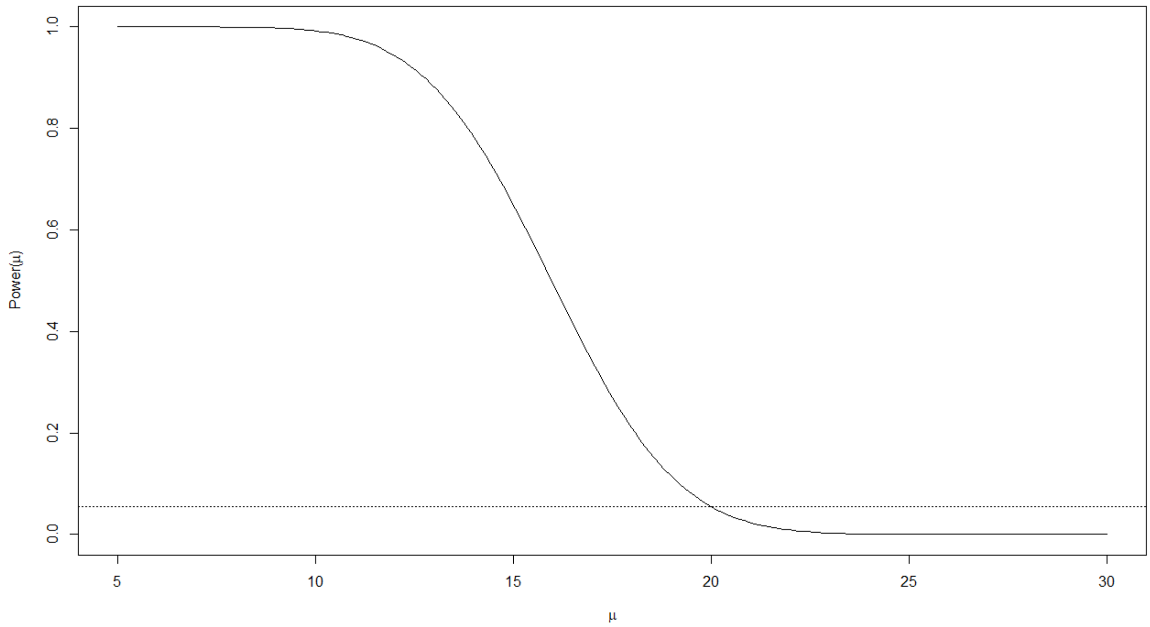
4) What remaining questions or concerns do you have about power?

D. Undergraduate-level Final Exam Question

A vice president in charge of sales for a large company claims that the average number of sales contacts per week has decreased from what it was 10 years ago. To check this claim, 5 salespeople are selected at random, and the number of contacts made by each is recorded for a single randomly selected week. To test her claim, the vice president is presented with the following hypothesis test function:

$$\zeta_m(X_1, ..., X_n) = \begin{cases} 1 & X_{(n)} < c \\ 0 & else \end{cases}$$

The following plot of the power function for ζ_m was produced, and a horizontal line was drawn at the significance level, $\alpha = 0.0546$.



a) Explain why a Poisson distribution with mean μ would be a reasonable model for each of the random variables $X_1, \ldots, X_n$. Your explanation should also include **in words** what $n, X_1, \ldots, X_n, \mu$ and $x_1, \ldots, x_n$ are in this situation, and, if applicable, give the numerical values of each.

b) Based on the power curve provided in the plot above, what are the hypotheses the vice president wants to test? Be sure to include any appropriate numerical values and use words to describe any symbols used.

c) The vice president was presented with another hypothesis test function:

$$\zeta_s(X_1, ..., X_n) = \begin{cases} 1 & \sum_{i=1}^{n} X_i < k \\ 0 & else \end{cases}$$

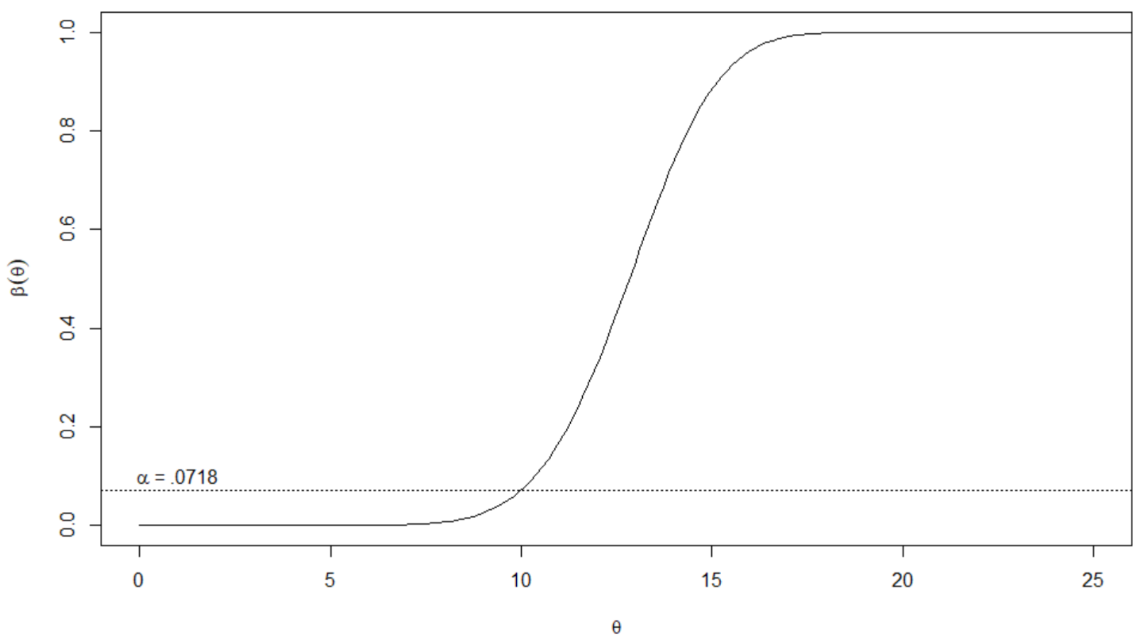
- **On the plot provided above**, sketch what you think the power function of this hypothesis test would look like in relation to the power function of ζ_m. Assume approximately the same significance level is used. (*Note:* The power function must be drawn clearly on the plot provided above in order to receive credit.)
- **In words**, explain thoroughly **why** the power function for ζ_s should look similar to what you sketched. Justify your choice with appropriate statistical evidence and support. A complete explanation should include rationale for the relative positions chosen for the power functions over the **entire** parameter space.

E. Graduate-level Final Exam Question

An ecologist suspects the mean rate of diseased pine trees per one acre, θ, is underestimated. To check this claim, the ecologist randomly selects 23 different areas: 5 areas each of size 0.25 acres, 10 areas each of size 1 acre, and 8 each of size 0.75 acres. The number of diseased pine trees in each randomly selected area is recorded and assumed to be independent of one another. To test the proposed claim, the ecologist is presented with the following hypothesis test function:

$$\phi_m(\boldsymbol{X}) = \begin{cases} 1 & X_{(n)} > c \\ 0 & else \end{cases}$$

The following plot of the power function for ϕ_m was produced, and a horizontal line was drawn at the significance level, $\alpha = 0.07183788$.



a) Explain why a Poisson($a_i\theta$) distribution would be a reasonable model for the random variables $X_1, \ldots, X_n$. Your explanation should also include **in words** what n, $X_1, \ldots, X_n$, $a_1, \ldots, a_n$, θ, and $x_1, \ldots, x_n$ are in this situation, and, if applicable, give the numerical values of each.

b) Based on the power function provided in the plot, what are the hypotheses the ecologist wants to test? Be sure to include an appropriate numerical values and use words to describe any symbols used.

c) Derive the power function for ϕ_M that is plotted. Find the cut-off, c. Show and explain all work.

d) Explain why $\alpha = 0.07183788$ instead of the more typical 0.07. (Hint: How does $\sup_{\theta \in \Theta_0} \beta(\theta)$ change as the value of the cut-off, c, changes?)

e) The ecologist is presented with another hypothesis test function:

$$\phi_S(\boldsymbol{X}) = \begin{cases} 1 & \sum_{i=1}^{n} X_i > 191 \\ 0 & else \end{cases}$$

- **On the plot provided above**, sketch what you think the power function of this hypothesis test would look like in relation to the power function of ϕ_M. Assume approximately the same significance level is used.
- **In words**, explain thoroughly **why** the power function for ϕ_S should look similar to what you sketched. A complete explanation should also include rationale for the relative positions chosen for the power functions over the **entire** parameter space.

f) Use simulation to obtain an empirical power function for the test in part e). Briefly explain how you obtained this empirical power function. (Note: Simulate values from the population distribution(s), not the sampling distribution of the test statistic.)

g) For the 23 areas that were randomly selected, the ecologist observed that a total of 197 pine trees were diseased, and the maximum number of diseased pine trees in an area in the sample was 16 pine trees. Is there evidence to suggest that the mean rate of diseased pine trees per one acre is underestimated? Explain why or why not, using an appropriate p-value to justify your answer. Explain and interpret all evidence within the context of the scenario.

F. Additional Activity Questions

We provide the following questions as additional resources to explore particular aspects of power in more detail.

Varying the population distribution

Choose the sample maximum as the test statistic and greater than as the alternative hypothesis. Holding all other values constant, what happens to the power curve as the population distribution changes between exponential, normal, and uniform? Why?

Which power function is closest to an ideal power curve? Why?

Varying the test statistic

Choose the uniform distribution as the population distribution and not equal to as the alternative hypothesis. Holding all other values constant, compare the power functions for each of the available test statistics. Which statistic provides the highest power across the alternative space?

Switch the population distribution to the normal distribution and repeat the previous exercise. Explain why your answer to each of these questions differs.

Varying the alternative hypothesis

Choose the exponential distribution as the population distribution and sum of the X's as the statistic. Holding all other values constant, what happens to the power curve as the alternative hypothesis changes between greater than, less than, and not equal to? Why?

Why is the power function not symmetric about the null value?

Choose the uniform distribution as the population distribution, sample maximum as the statistic, not equal to as the alternative hypothesis, and set the sample size to one. Is this test unbiased? Why or why not?

Varying the significance level

Choose any combination of population distribution and statistic. Holding all other values constant, what happens to the power function as the significance level increases? Why?

Explain how the significance level affects the power curve in terms of the type I error rate. What trade-offs are made in terms of the probability of making a type II error when increasing the significance level? (Hint: recall that the power of a test is equal to one minus the probability of a type II error)

Varying the null value

Choose any combination of population distribution and statistic. Holding all other values constant, what happens to the power function as you change the null value? Why does this happen?

Explain the relationship between the power function, null value, and significance level.

Varying n

Choose the normal distribution as the population distribution, sum of the X's as the test statistic, and not equal to as the alternative hypothesis. Holding all other values constant, what happens to the power function as the sample size increases? Why?

Change the statistic to the sample minimum. What happens to the power function as the sample size increases?

Why do you think the effect of the sample size on the power curve is lesser for the sample minimum than for the sum of the X's?

Sampling distributions

Choose the exponential distribution as the population distribution, sum of the X's as the test statistic, greater than as the alternative hypothesis, a sample size of 25, and set the null value and theta to be three and four respectively. Describe where each of the distributions plotted under the Sampling Distribution heading is centered and what each distribution represents. Hint: recall that the distribution of the sum of n *exponential(*θ*) random variables is gamma(*n, θ*).*

Describe what the red and grey shaded areas represent. How is the red area used to determine the grey area?

Which of the factors on the sidebar panel that affect each of these sampling distributions are controlled by the researcher and which are unknown? Based on your answer, how would you recommend a researcher increases the power of their hypothesis test?