Power Approximations for Reliability Test Designs

Tom Johnson
Rebecca Medlin
Laura Freeman

2018 Fall Technical Conference

Institute for Defense Analyses

Three Federally Funded Research & Development Centers

Systems and Analyses Center (Alexandria, VA)

Science and Technology Policy Institute

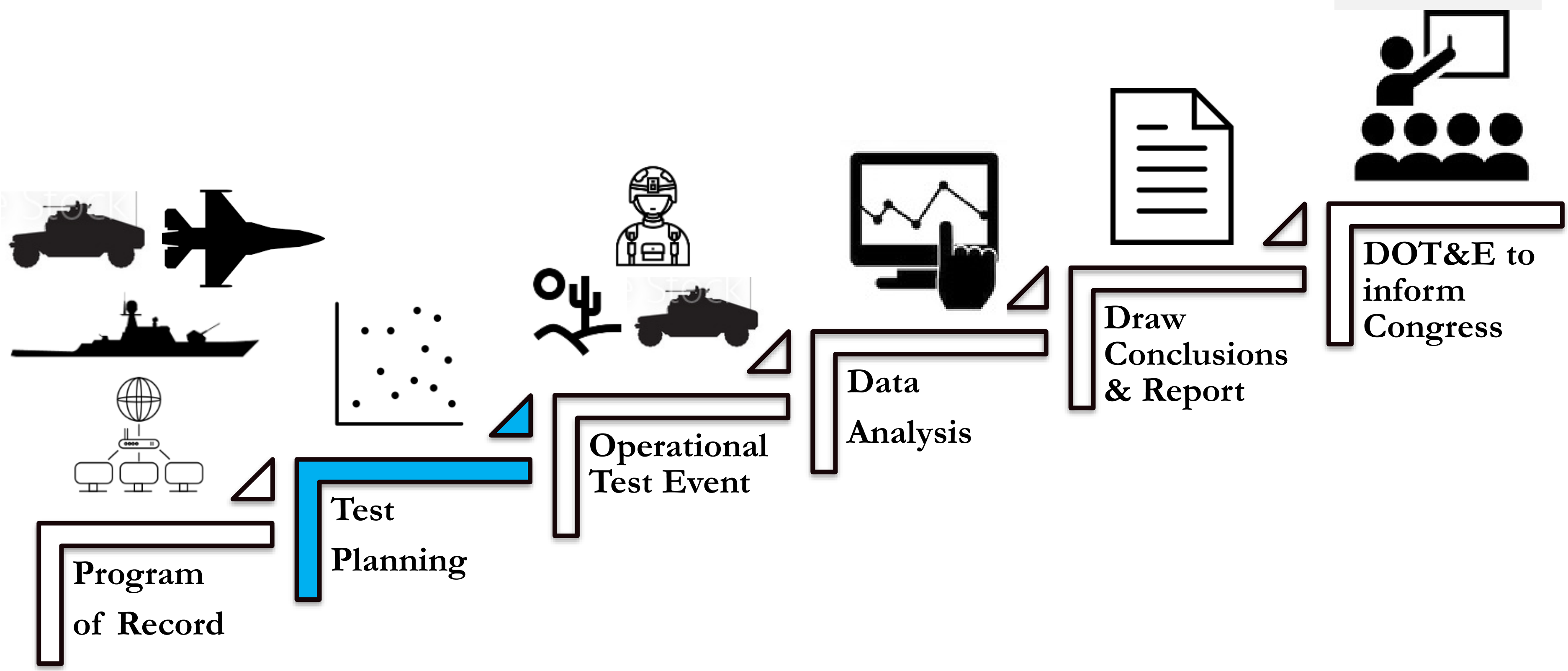
Center for Communications and Computing

- Cost Analysis and Research
- Information Technology and System
- Intelligence Analyses
- Joint Advanced Warfighting
- **Operational Evaluation Division**
- Science and Technology
- Strategy, Forces and Resources
- Science and Technology

OED provides technical analytical support to the Department of Defense's Director, Operational Test & Evaluation and its mission to conduct independent assessments of the military services' test and evaluation of new weapons systems.

- Help design good tests of military systems in realistic operational environments.
- Evaluate the operational effectiveness, suitability, and survivability of those systems from an objective, disinterested, and factual perspective.

Operational Test and Evaluation



Program of Record

Test Planning

Operational Test Event

Data Analysis

Draw Conclusions & Report

DOT&E to inform Congress

Synopsis

Reliability experiments determine which factors drive system/product reliability.

Reliability data tend to follow distinctly non-normal distributions and include censored observations.

Our experimental designs should accommodate the skewed nature of the response and allow for censored observations.

Monte Carlo simulations are frequently used to evaluate the design properties (e.g., power) for reliability experiments.

Simulation can be inefficient to compare multiple experiments of various sizes.

We have developed a closed form approximation to calculate the power of a reliability experiment.

Illustrative Example

Planning a reliability experiment for **N = 240** "products."

A **2 x 2 x 3** full factorial experiment **(design)**

Expose until failure or up to **10 days (right censoring time)**

Identify seven days as the crucial juncture in the product's lifetime; we anticipate 80% of the products will fail by this time **(nominal failure rate)**.

Detect 10% change in probability of failure due to an exposure factor **(effect size)**.

Lower probability of failure $p_1 = .75$ and upper probability of failure $p_2 = .85$

Failure times follow a **lognormal** distribution with fixed scale parameter, $\boldsymbol{\sigma = 2}$

Effect sizes, in terms of the location parameters $\mu_1$and μ_2 are found by:

$$p1 = F\left(t_p, \mu_{p_1}\right) \text{ and } p_2 = F(t_p, \mu_{p_2}):$$

$$\mu_1 = .6 \text{ and } \mu_2 = -.13$$

How do we evaluate our design?
How do we calculate the **power of test** for our nine model coefficients?

Discussion of power is common in classical experimental design evaluation

Discussions of power **are not** prevalent in reliability research.

Meeker (1977; 1992; 1992; 1994; 1995; 1998; 2006).

- Precision around a Quantile Estimate or Hazard Functions
- Good for quality control applications…

So why not just use Monte Carlo?

It is a flexible and accurate approach.

It could quickly become computationally inefficient.

A closed form approximation is computationally efficient.

The Failure-Time Regression Model

$$y_i = \log(T_i) = m_i^T \beta + \sigma \epsilon_i$$

$i = 1,2, \dots k$ design points

Model the location $\mu_i = m_i^T \beta$; fix the scale σ

Lognormal Model: $\mathrm{T}_i \sim LogN(\mu_i, \sigma)$ and $\epsilon_i \sim N(0,1)$

Weibull Model: $\mathrm{T}_i \sim Weibull(\mu_i, \sigma)$ and $\epsilon_i \sim SEV(0,1)$

Censoring: Fixed, Type I Right Censoring Scheme

$$\delta_{ij} = 1 \; if \; T_{ij} < t_c \, ; \delta_{ij} = 0 \; otherwise$$

Maximum likelihood estimation

$$l_{n\sigma} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \log \left[f_{T_{ij}, \mu_i} \right] \delta_{ij} + \log[1 - F_{t_c, \mu_i}](1 - \delta_{ij})$$

Testing for Significance: Likelihood Ratio Test

Coefficients under test Nuisance coefficients

$$\beta = [\psi, \lambda] \text{ and } M = [X, Z]$$

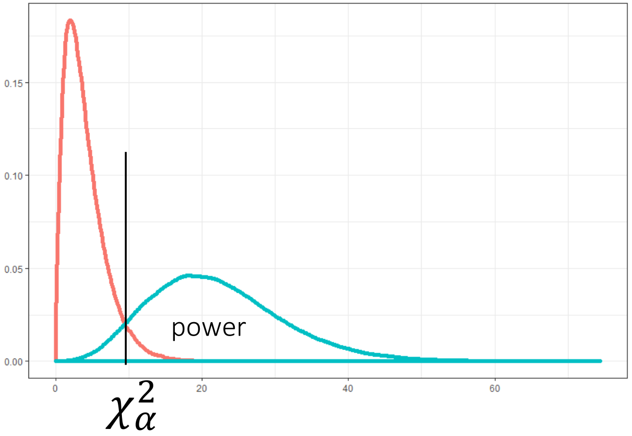
$$\mu_i = x_i^T \psi + z_i^T \lambda$$

$$H_0 \colon \psi = \psi_0$$

$$H_A \colon \psi \neq \psi_0$$

Likelihood Ratio Statistic: $2[l_{n\sigma}(\hat{\psi}, \hat{\lambda}) - l_{n\sigma}(\psi_0, \hat{\lambda}_0)]$

The power of a test is:



Central χ^2 and non-central $\tilde{X}^2$

$$Pr\left(t > \mathrm{X}_{\alpha}^{2}\right) = 1 - \tilde{X}^{2}(\mathrm{X}_{\alpha}^{2}, p, \gamma)$$

t is the non-central chi-square random variable.

X_{α}^{2} is the upper α percentage point of the central chi-square distribution

$\tilde{X}^2$ is the non-central chi-square distribution with p degrees of freedom (number of coefficients under test) and non-centrality parameter γ

Self, Mauritsen, and O'Hara (1992)

Describe a power approximation approach for the generalized linear model (glm) framework.

Based on a non-central chi-square approximation to the distribution of the likelihood ratio statistic.

$$LRS \sim \tilde{\mathrm{X}}^2_{p,\boldsymbol{\gamma}}$$

Technique accommodates any model within the exponential family of distributions that can be arranged into the glm canonical form:

$$\mathrm{f}(\mathrm{y_i}) = \exp\left(\frac{\mathrm{y_i}\theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i, \phi)\right)$$

Examples: logistic, Poisson, and gamma regression models.

Failure-time regression models share many qualities of the generalized linear model, **but they cannot be arranged into the canonical form**.

If we can solve for the non-centrality parameter γ, then we can estimate power.

We know the expected value of a noncentral chi-square random variable is $p + \gamma$.

If we equate the expected value of the LRS to the expected value of the noncentral chi-square random variable, we can solve for γ.

Great Expectations [and Expansions]

Coefficients under test

Nuisance coefficients

$$E_{\psi,\lambda}\{2[l_{n\sigma}(\hat{\psi},\hat{\lambda}) - l_{n\sigma}(\psi_0,\hat{\lambda}_0)]\} =$$

$$E_{\psi,\lambda}\{2[l_{n\sigma}(\hat{\psi},\hat{\lambda}) - l_{n\sigma}(\psi,\lambda\ )]\}$$

$$- E_{\psi,\lambda}\{2[l_{n\sigma}(\psi_0,\hat{\lambda}_0) - l_{n\sigma}(\psi_0,\lambda_0^*\ )]\}$$

$$+ E_{\psi,\lambda}\{2[l_{n\sigma}(\psi,\lambda\ ) - l_{n\sigma}(\psi_0,\lambda_0^*\ )]\}$$

$$p + \gamma = A - B + C$$

$E_{\psi,\lambda}\{.\}$ taken with respect to the true parameters ψ and λ

λ_0^* is the limiting value of the null coefficients

$$p + \gamma = A - B + C$$

The first term (A): A = $\boldsymbol{p} + \boldsymbol{q}$

- Where p is the number of coefficients under test and q is the number of nuisance coefficients (i.e., coefficients not under test)

The second term (B): take limits and use Taylor series expansion…

- Closed form solution for GLM within the canonical form (Self et al., 1992).
- No closed form solution for failure-time regression models with censoring. A numerical solution is possible, but would undermine our objective!

The third term (C): a closed form solution exists!

- Closed form solution for GLM (shown in Self et al., 1992)
- Closed form solution for Failure Time regression models with fixed, Type 1, right-censored data (our work)

All you need is C

$$\boldsymbol{\gamma} = \boldsymbol{A} - \boldsymbol{B} + \boldsymbol{C} - \boldsymbol{p}$$

$$\approx$$

$$\boldsymbol{\gamma} = \boldsymbol{C}$$

*“**the dominance of the [C] term** in the calculation of the non-centrality parameter,”* - Self et al. (1992)

*“In our experience, **the term** [$\boldsymbol{A} - \boldsymbol{B} - \boldsymbol{p}$] **is usually very close to zero.**”* - Self et al. (1992), O’Brien and Shieh (1998), Shieh (2000), Brown et al. (1999)

Solving for C

$$C = E_{\psi,\lambda}\{\log\left[f_{T_{ij},\mu_i}\right]\delta_{ij} + \log[1 - F_{t_c,\mu_i}]\left(1 - \delta_{ij}\right) - \log\left[f_{T_{ij},\mu_i^*}\right]\delta_{ij} + \log\left[1 - F_{t_c,\mu_i^*}\right]\left(1 - \delta_{ij}\right)\}$$

Lognormal Solution

$$C = \sum_{i=1}^{k} n_i \left\{ \frac{2 f_{t_c,-\mu_i}(\mu_i^* - \mu_i)}{t_c^{-1-\frac{2\mu_i}{\sigma^2}}} + \frac{F_{t_c,\mu_i}(\mu_i - \mu_i^*)^2}{\sigma^2} + \left(2 - 2F_{t_c,\mu_i}\right)\log\left(\frac{2 - 2F_{t_c,\mu_i}}{2 - 2F_{t_c,\mu_i^*}}\right)\right\}$$

Weibull Solution

$$C = \sum_{i=1}^{k} n_i \left\{ -\frac{2}{\sigma} e^{-\frac{\mu_i^*}{\sigma}} F_{t_c,\mu_i}\left(e^{-\frac{\mu_i}{\sigma}}\sigma + e^{\frac{\mu_i^*}{\sigma}}(\mu_i - \mu_i^* + \sigma)\right)\right\}$$

Equation Notes:

$f_{t,\mu} = f(t,\mu,\sigma)$; $\mathrm{F}_{t,\mu} = F(t,\mu,\sigma)$

λ_0^* is the limiting value of the null coefficients and is found by fitting a lognormal regression model to the alternative data. The alternative data are the failure times that represent the perfect fit to the alternative coefficients: $T_{ij}^* = e^{\mu_i}$. Use standard failure-time model fitting software to fit the reduced model to T_{ij}^*. The fitted coefficients are equal to λ_0^* and $\mu_i^* = x_i^T\lambda_0^* + z_i^T\psi_0$.

Returning to the Example

N: 204
Design: 2 x 2 x 3
Censor time: t_c: 10
Nominal failure rate at 7 days: $\bar{p}$ = 0.8
Effect size: Δ= .1
$p_1 = .75$ and $p_2 = .85$
Lognormal distribution with $\sigma = 2$
$p_1 = F_{t_p,\mu_{p_1}}$ and $p_2 = F_{t_p,\mu_{p_1}}$
μ_1=.6 and μ_2=-.13

Main Effects Model:

$$\beta_{int} + \beta_{F1} + \beta_{F2} + \beta_{F3_1} + \beta_{F3_2}$$

Sum-to-zero contrast, effect size of alternative coefficients:

$$\beta^T = [.23, .36, .36, .36, .00]$$

Illustrate power for F_1

$$H_o: \psi = 0$$
$$H_A: \psi \neq 0$$

$$\lambda^T = [\beta_{int}, \beta_{F2}, \beta_{F3}], \qquad \psi^T = \beta_{F1}$$

Calculate C

$$\mu_i = x_i^T \lambda + z_i^T \psi$$
$$\mu_i^* = x_i^T \lambda_0^*$$

$$\gamma = C = 7.59$$

Power

$$Pr(t > X^2_{\alpha=0.05})$$
$$= 1 - \tilde{X}^2\{X^2_{\alpha=0.05}, p = 1, \gamma = 7.59\}$$
$$= 1 - \tilde{X}^2\{3.84, p = 1, \gamma = 7.59\}$$
$$= 0.7869$$

Does the simplifying assumption work?

Model	Design	N	$\bar{p}$	Δ_p	σ
Lognormal	2^2	64	0.5	.06	.5
Weibull	2^6	128	0.8	.12	1
		256			3

Conditions varied using a 40-run D-optimal experimental design

$t_c = 10; t_p = 7$

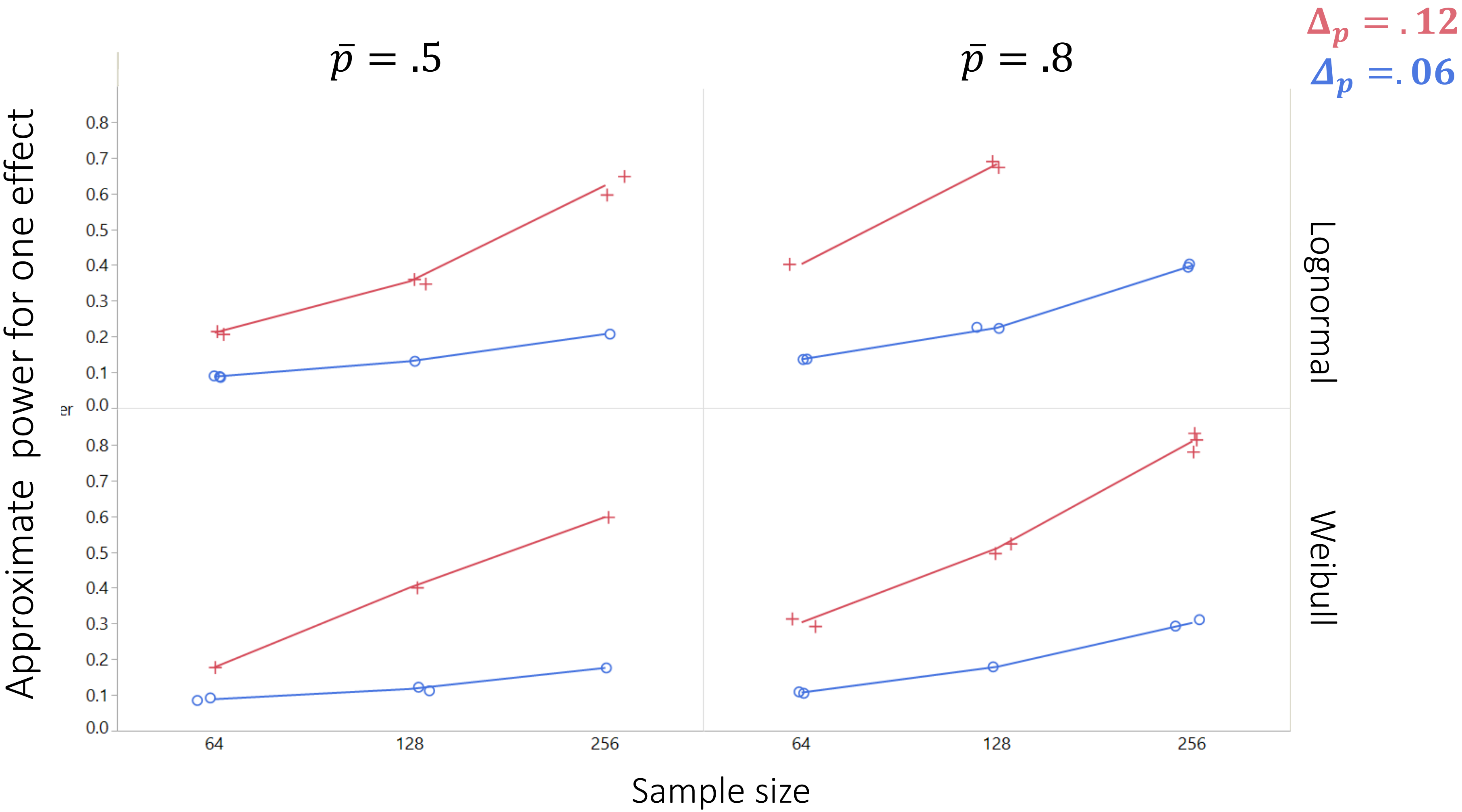
Power Approximation (C)	Monte Carlo Power*

Questions of Interest:

- Approximate Power: do we get a reasonable solution?
- |Approximate Power – Monte Carlo Power|: how accurate is that solution?
  - Is calculating C good enough?

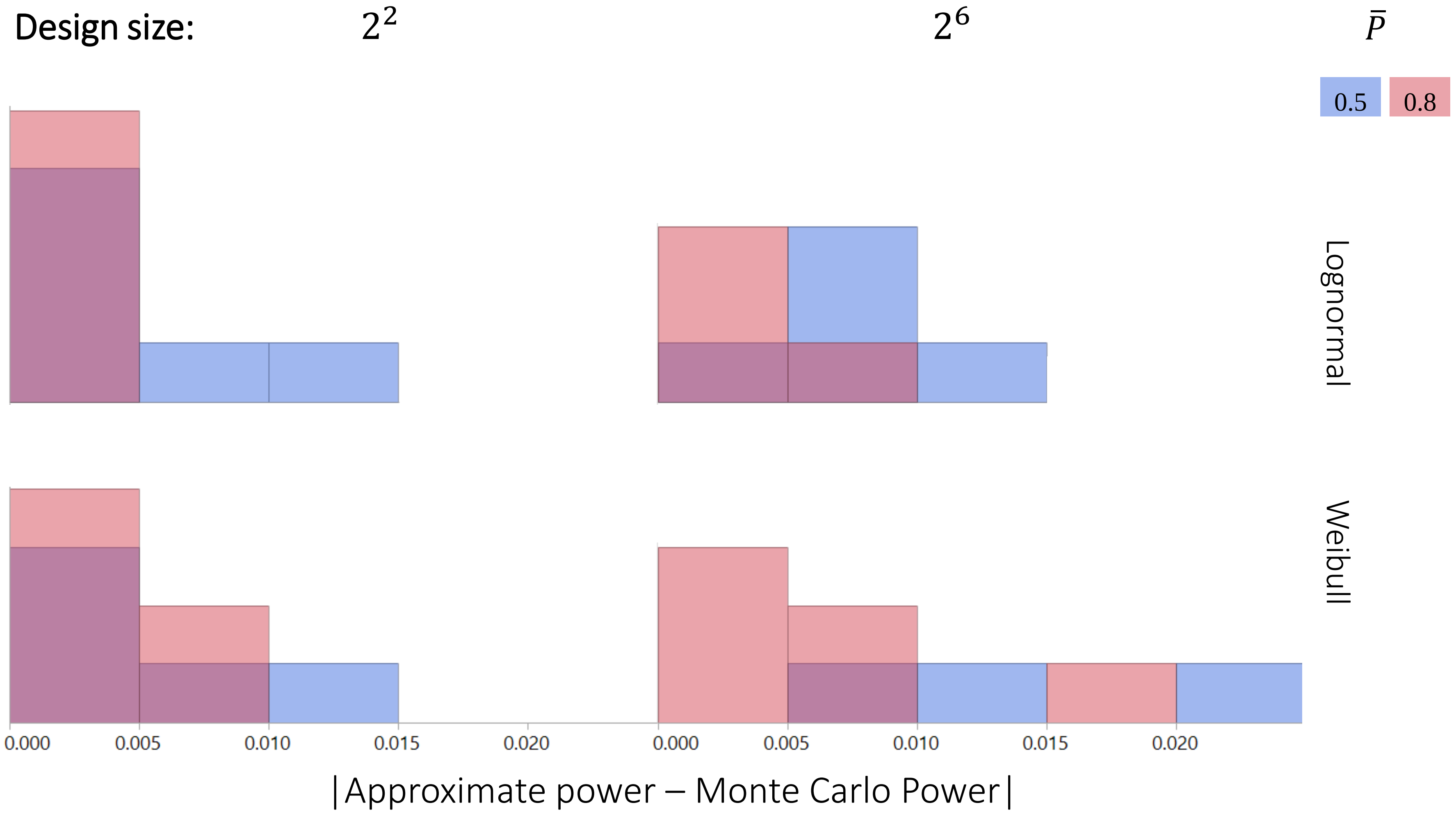
*Monte Carlo Power is based on 10,000 iterations

Approximate power results behave as expected



Differences from Monte Carlo Power are small

Most influential conditions: $\bar{\boldsymbol{p}}$,Design, Model



Is the method computationally efficient?

We explore accuracy/timeliness trade-off:

- Study based on example construct.
- Power calculated for one coefficient, β_{F_1}
- ‘True’ power, Monte Carlo Simulation with 5 million iterations: 0.78820

Monte Carlo: Stochastic

Iteration cases considered: 100, 500, 1,000, 5,000, 10,000, 15,000, 20,000, 25,000

*Replicate each iteration case 20 times to estimate an average iteration power and corresponding accuracy, RMSE, and an average iteration time.

Power Approximation: Deterministic

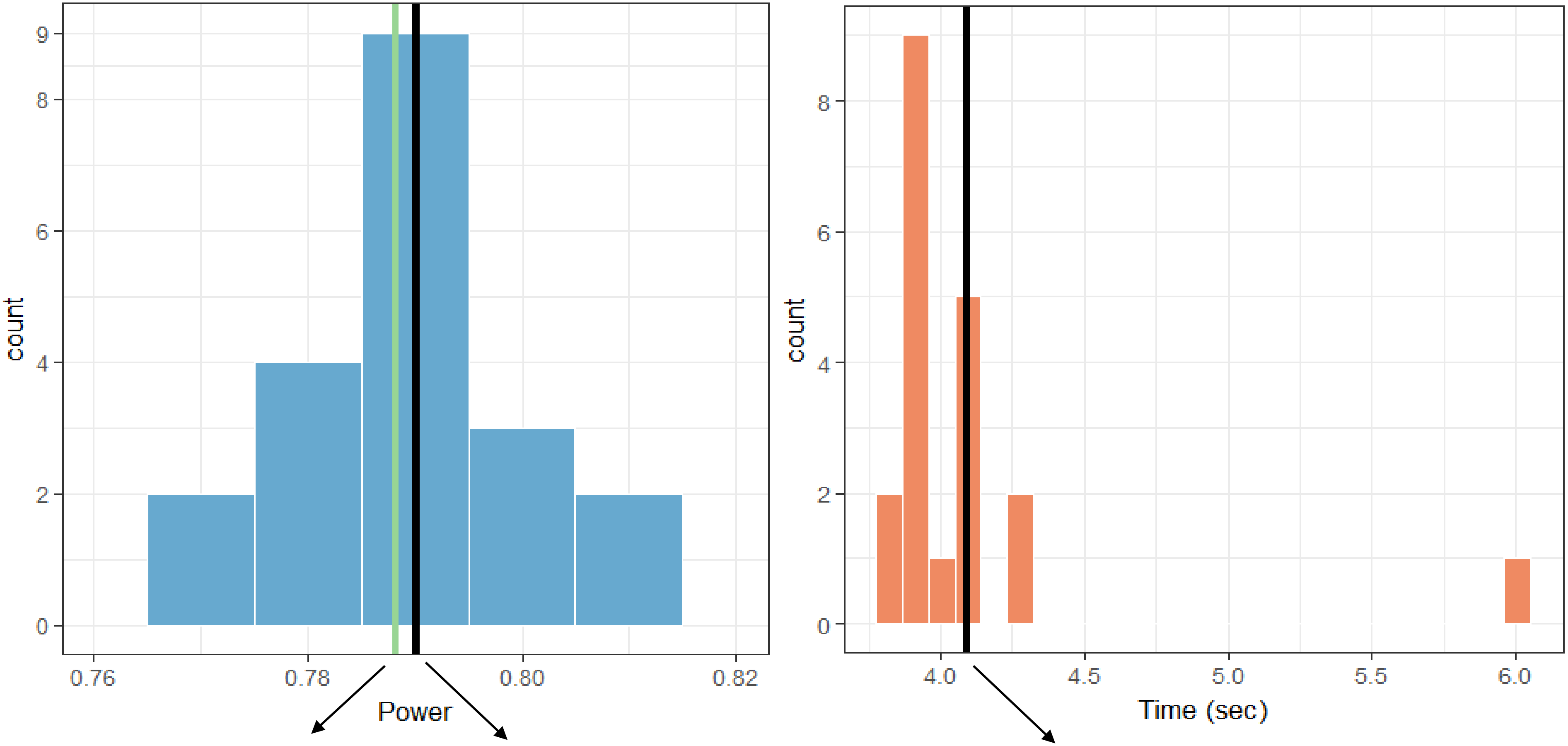
Power: 0.78688

Accuracy: 0.00132

Time: 0.25 seconds

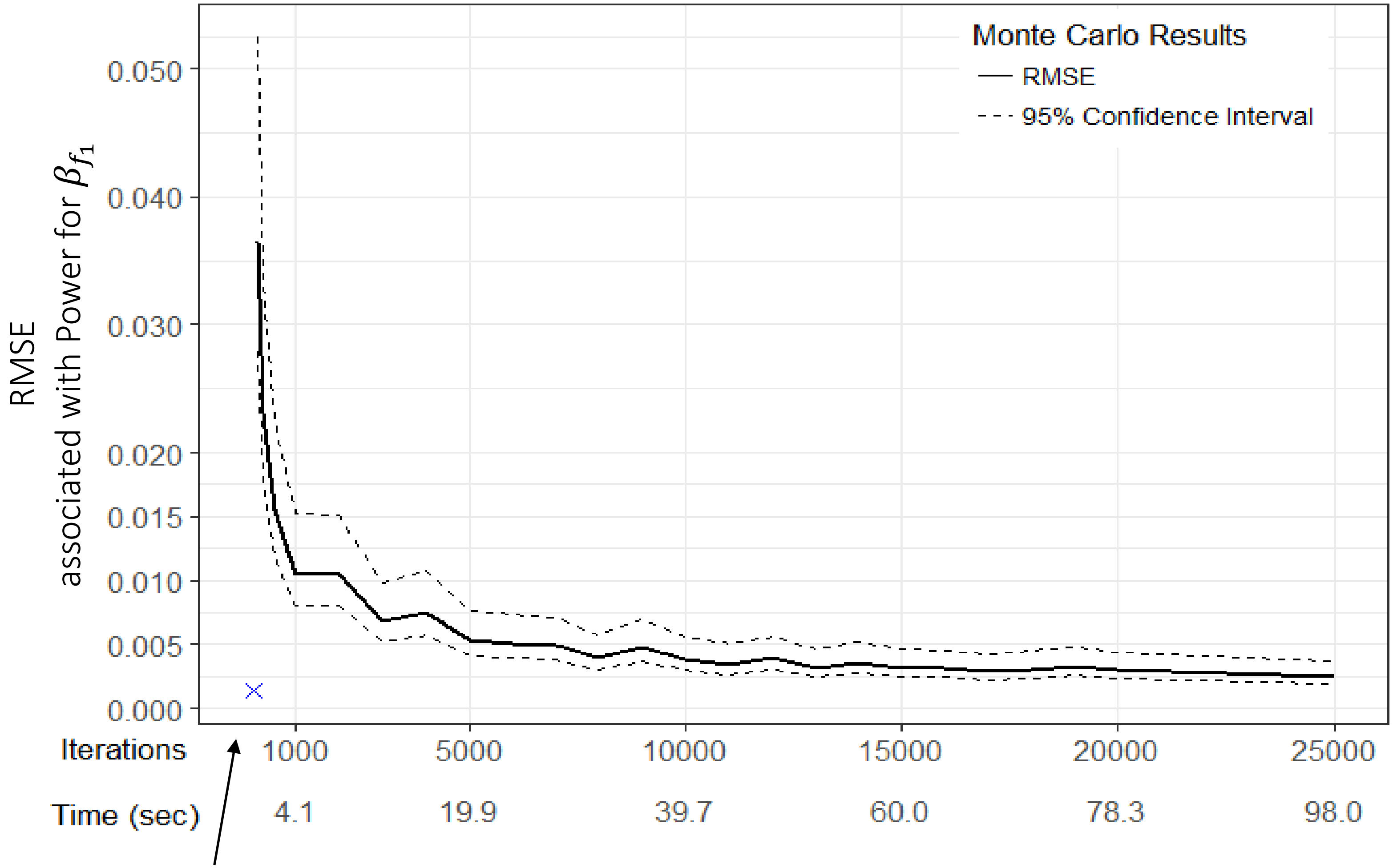
Example Monte Carlo Output

Monte Carlo output for the 1,000 iteration case, with 20 replicates



`true` power = 0.78820 1,000 iteration mean power = 0.79 1,000 iteration mean time = 4.10 sec

The benefit of our approximation method relies on its computational efficiency.



Power Approximation Method

Absolute Error = .001

Time (sec) = 0.25

Computational time can quickly compound!

A main effects + two-factor interaction model has nine coefficients.

→ We need nine power computations.

Typically, confidence and power are fixed and we solve for sample size.

→ This requires iterative numerical techniques.

→ Suppose for example, our problem requires five iterations.

The 1,000 iteration case now takes:[1]

Monte Carlo: 4 seconds x 9 coefficients x 5 iterations to optimize

= **3 minutes**

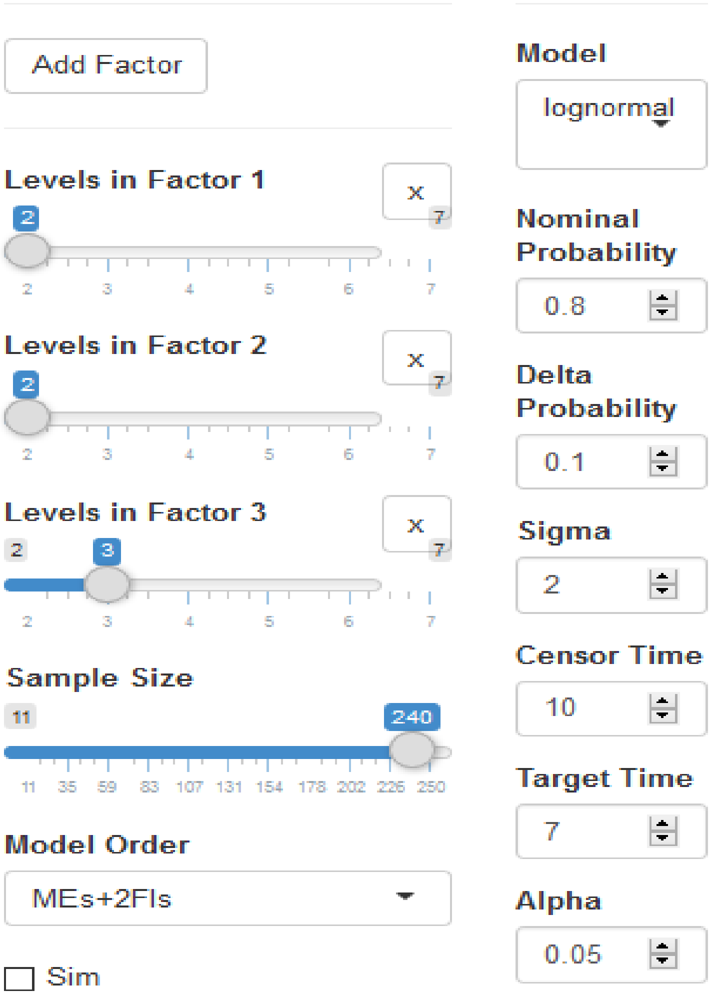
Power Approx.: 0.25 seconds x 9 coefficients x 5 iterations to optimize

= **11 seconds**

---

[1] Using R Software with a Processor: Intel® Core™ i7-7600U CPU @2.80 Ghz (4 CPUs), ~2.9 GHz

We've made it easy to implement!
https://test-science.shinyapps.io/survpow/



Levels in Factor 1: 2

Levels in Factor 2: 2

Levels in Factor 3: 2 – 3

Sample Size: 11 – 240

Model Order: MEs+2FIs

☐ Sim

Model: lognormal

Nominal Probability: 0.8

Delta Probability: 0.1

Sigma: 2

Censor Time: 10

Target Time: 7

Alpha: 0.05

Calculate Power

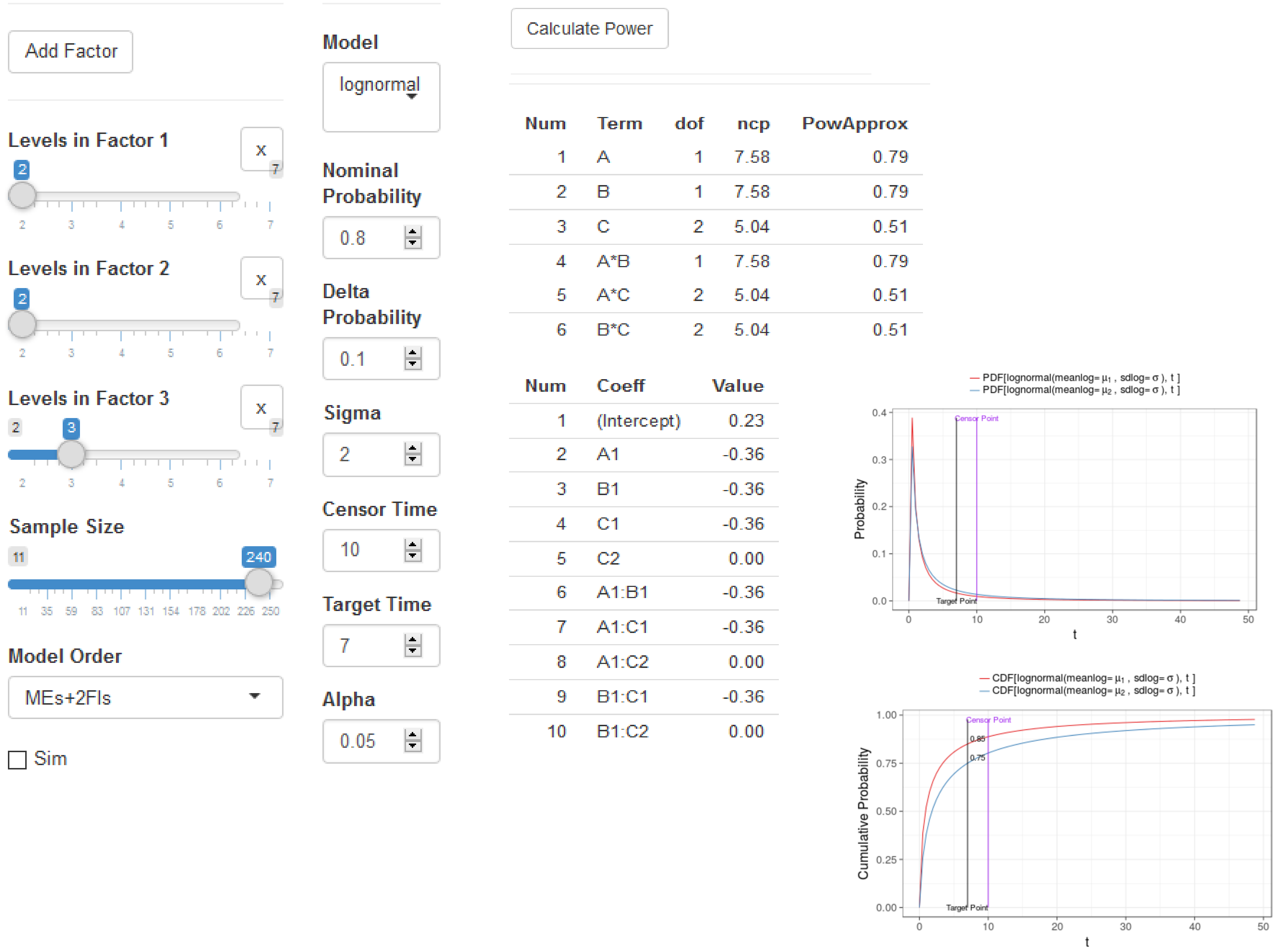
- **Model** specifies the expected distribution of failure times
- **Nominal Probability** is the nominal probability of failure at the Target Time.
- **Delta Probability** is the effect size. That is,

$$p_1 = p_{\text{nom}} + \frac{p_{\text{delta}}}{2}$$

$$p_2 = p_{\text{nom}} + \frac{p_{\text{delta}}}{2}$$

and the Location parameters μ_1 and μ_2 are calculated using p_1 and p_2

- **Sigma** is the scale parameter (constant)
- Failure times are not observed after the **Censor Time** (Fixed, Type 1, Right Censoring)
- **Alpha** is the Type I error rate.
- **Sample Size** adjusts the number of runs in the D-optimal experiment.
- The **simulate checkbox** allows the user to use Monte Carlo to calculate power for the likelihood ratio test.



Add Factor

Levels in Factor 1 x

2

Levels in Factor 2 x

2

Levels in Factor 3 x

2 3

Sample Size

11 240

Model Order

MEs+2FIs

☐ Sim

Model

lognormal

Nominal Probability

0.8

Delta Probability

0.1

Sigma

2

Censor Time

10

Target Time

7

Alpha

0.05

Calculate Power

Num	Term	dof	ncp	PowApprox
1	A	1	7.58	0.79
2	B	1	7.58	0.79
3	C	2	5.04	0.51
4	A*B	1	7.58	0.79
5	A*C	2	5.04	0.51
6	B*C	2	5.04	0.51

Num	Coeff	Value
1	(Intercept)	0.23
2	A1	-0.36
3	B1	-0.36
4	C1	-0.36
5	C2	0.00
6	A1:B1	-0.36
7	A1:C1	-0.36
8	A1:C2	0.00
9	B1:C1	-0.36
10	B1:C2	0.00