

Monte Carlo Integration & R Package Implementation for Rare Events

MAHE DDS 6106: Statistical Research Methodology | Exemplar Reading Guide

· **Classification:** Instructional Statistics Thesis Teaching Model (Simulation & R Package)

.. Reading Time: 20-25 mins | · Difficulty: · Advanced | · Audience: M.Sc. Data Science & Biostatistics Students

? Why Selected: Features mathematical derivations, Monte Carlo simulation experiments, and open-source R package code.

STRUCTURAL COLOR LEGEND & ANNOTATION GUIDE

-  Title & Abstract: High-impact summary, study rationale, & indexing keywords
-  Introduction & Literature: Research problem, background literature, & study hypothesis
-  Methods & Study Design: Study design, participants, sample size, SAP, & ethics approval
-  Results & Data Analysis: Empirical findings, statistical tables, p-values, & figures
-  Discussion & Conclusion: Interpretation, comparison with literature, limitations, & conclusion

.. THESIS STRUCTURAL ROADMAP

- Front Matter & R Package Code Repository Link
- Chapter 1: Introduction & Computational Problem
- Chapter 2: Review of Importance Sampling & Variance Reduction
- Chapter 3: Mathematical Formulation of Novel Estimator
- Chapter 4: Monte Carlo Simulation Experiments
- Chapter 5: Application to Financial Risk & Genomic Benchmarks
- Chapter 6: R Package Architecture & Code
- Chapter 7: Discussion & Future Optimization

· MILESTONE ICONS TO LOCATE

- **Mathematical Derivation**
Unbiased estimator & variance formula
- **Simulation Chapter**
Monte Carlo experiment evaluating MSE & SE
- **R Package**
Exported package functions, S3 classes, & unit tests
- **Real Data Test**
Financial loss tail-risk benchmark application
- **Future Code**
C++ Rcpp integration for 100x speedup

· HOW A THESIS BECOMES A JOURNAL PAPER (IMRaD)

Data science theses generate TWO outputs: an empirical journal paper (IMRaD) and an open-source R package hosted on CRAN/GitHub! Chapter 6 forms the package vignette.

Full Monograph (150-200 Pages)

Contains literature review, raw tables, complete proofs, & code

· Select One Primary Question

Extract single core empirical finding or novel theoretical contribution

· Condense to 8-10 Pages (IMRaD)

Introduction (Ch1) -> Methods (Ch3) -> Results (Ch4) -> Discussion (Ch5)

· Submit to Peer-Reviewed Journal

Peer review, revision, and open-access journal publication

-> Turn page to inspect the annotated exemplar!

MONTE CARLO INTEGRATION & R PACKAGE IMPLEMENTATION FOR R

A Dissertation Presented for the Degree of Thesis Blueprint 5: Statistics & Data Science Archetype

Department of Applied Statistics and Data Science
Prasanna School of Public Health, MAHE, Manipal

DECLARATION & CERTIFICATE

I hereby declare that this dissertation is my own original work and has not been submitted for any other degree.

ABSTRACT

ABSTRACT SUMMARY:

This dissertation presents a comprehensive investigation into monte carlo integration & r package implementation for rare events. We detail theoretical foundations, execute simulation studies, and validate results on real health data.

CHAPTER 1: INTRODUCTION

COMPUTATIONAL PROBLEM:

Standard Monte Carlo integration requires impractically large sample sizes ($N > 10^7$) to estimate rare tail event probabilities $P(X > x) < 10^{-5}$.

THESIS OBJECTIVES:

1. Formulate an adaptive importance sampling estimator.
2. Execute a Monte Carlo simulation study comparing variance reduction.
3. Release an open-source R package.

CHAPTER 3: MATHEMATICAL METHODOLOGY

ESTIMATOR DERIVATION:

Let $I = E_f[h(X)]$. We propose proposal density $g^*(x) = |h(x)| f(x) / I$. The importance sampling estimator $I_{\text{hat}} = (1/N) \sum (h(X_i) f(X_i) / g(X_i))$ is un-biased with minimal variance.

CHAPTER 4: MONTE CARLO SIMULATION STUDY

EXPERIMENT SETUP:

Compared Raw Monte Carlo, Antithetic Variates, and Proposed Adaptive Importance Sampler across $N = 10^3$ to 10^6 draws.

RESULTS:

Proposed sampler achieved a 94.2% variance reduction and cut standard error by 4.1x.

CHAPTER 6: R PACKAGE IMPLEMENTATION

PACKAGE STRUCTURE:

The algorithm is packaged in R package ``rareeventr``.

- Function ``importance_sim()``: Main sampler interface.
- Unit tests: 100% test coverage using ``testthat``.
- Vignettes: Worked reproducible case studies.

INTERACTIVE STUDENT CHECKLIST: CAN YOU LOCATE...?

MAHE DDS 6106: Statistical Research Methodology | Active Self-Assessment

Before closing this paper, test your ability to locate these 10 core research components:

· 1. Research Problem & Gap ·

Where do the authors explicitly state what previous studies failed to solve?

· 2. Research Question & Objectives ·

Where is the primary study aim or formal statistical hypothesis stated?

· 3. Study Design & Target Population ·

Is this an RCT, Cohort, Case-Control, Cross-Sectional, or Methodological design?

· 4. Variables & Primary Outcome ·

What is the primary dependent variable / clinical endpoint?

· 5. Sample Size Calculation ·

Did the authors provide a formal power calculation or sample size justification?

· 6. Statistical Analysis Plan (SAP) & Software ·

Which software packages (R, SAS, Python) and tests (ANOVA, Cox, Logistic) were used?

· 7. Primary Empirical Finding & Effect Size ·

Locate the primary result, confidence interval, p-value, or Odds/Hazard Ratio.

· 8. Study Limitations & Confounding ·

Where do the authors acknowledge unmeasured bias, attrition, or design constraints?

· 9. Ethics Approval & IRB Number ·

Locate the institutional ethics committee protocol approval code.

· 10. Funding & Conflict of Interest ·

Who funded the study, and did the authors declare any financial conflicts?

· Completed this checklist? Excellent work mastering scientific paper navigation!