Monte Carlo Simulation Tools for REDD+ Uncertainty Estimates

2024-12-19

Contents

Objective

When preparing for Monte Carlo simulations, it is best practice to start by examining descriptive statistics to characterize the empirical distributions of input variables. This preliminary analysis typically includes statistical tests for normality and visualizations of univariate distributions, such as histograms, kernel density plots, and Q-Q plots. Together, these tools provide critical insights into the shape, spread, symmetry, skewness, and presence of potential outliers in the data. Although this preliminary step may seem minor, it substantially influences uncertainty estimates, which can directly translate into increased financial returns, particularly within forest project landscapes exhibiting non-normal data distributions.

Accurately characterizing data distributions also helps in identifying and addressing biases, thereby ensuring high data quality and increasing confidence in subsequent estimations of biomass and carbon emissions. Selecting appropriate statistical distributions, informed by exploratory analyses, significantly enhances the reliability and precision of Monte Carlo simulations. Consequently, such careful statistical characterizations reduce overall uncertainty in forest biomass and emissions estimates. In turn, this strengthens the credibility of jurisdictional claims made under REDD+ programs and maximizes potential financial returns for Guyana from carbon financing initiatives.

Univariate distribution visualizations additionally provide auditors with useful diagnostic resources, enabling rapid identification and characterization of biases commonly encountered in biomass data. These diagrams help auditors efficiently assess the technical rigor and statistical approaches implemented by the project to monitor and manage uncertainty (ART, 2021: 8). Winrock strongly recommends incorporating distribution analyses early in a project's quantitative planning and throughout its technical standard operating procedures (SOPs). Such early integration represents a low hanging fruit with cost-effective strategy and significant potential in reducing audit findings, lowering uncertainty, and enhancing financial outcomes for Guyana's REDD+ activities. Specifically, early attention to data distributions directly informs appropriate simulation selection from the available options in SimVoi.

To effectively guide practitioners and stakeholders in selecting appropriate statistical distributions for Monte Carlo methods within forestry and REDD+ contexts, the following two tables present findings from a rapid review of relevant literature. The review identified and summarized statistical distributions frequently encountered in forestry, biomass estimation, and emissions analysis, which are disaggregated below between discrete and continuous types and according to their inherent statistical characteristics.

Table 1: Continuous data distributions, and example use cases for Monte Carlo simulations.

Distribution	Statistical Criteria & Use Cases	PDF
Normal (Gaussian)	Symmetric, bell-shaped distribution used for modeling continuous variables: biomass/ha	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$
Lognormal	Right-skewed distribution suitable for variables constrained to positive values (e.g., emission rates).	$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x-\mu)^2}{2\sigma^2}\right)$

Distribution	Statistical Criteria & Use Cases	PDF
Exponential	Models waiting times between independent events, such as forest fire occurrences or logging events.	$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$
Continuous Uniform	Assumes all values in an interval [a, b] are equally likely; useful for random spatial sampling in forests.	$f(x) = \frac{1}{b-a}, \quad a \leq x \leq b$
Chi-Square	Often used in goodness-of-fit tests to evaluate model accuracy in biomass estimation.	$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{\frac{k}{2}-1} e^{-x/2}, \quad x > 0$
t-Distribution	Suitable for small sample sizes with unknown population stdev (e.g., limited forest carbon data).	$f(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)} \left(1 + \frac{x^2}{v}\right)^{-\frac{v+1}{2}}$
Gamma	Models positively skewed data, such as biomass growth rates or carbon accumulation over time.	$f(x) = \frac{x^{k-1} e^{-x/\theta}}{\theta^k \Gamma(k)}$
Weibull	Flexible distribution used in reliability analysis, e.g., modeling tree mortality.	$f(x) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}$

Table 2: Discrete data distributions, and example use cases designed with Monte Carlo simulations.

Distribution	Statistical Criteria & Use Cases	PMF
Bernoulli	Binary outcome probability, e.g., presence/absence of deforestation in an area.	$P(X = x) = p^x (1-p)^{1-x}, \quad x \in \{0, 1\}$
Binomial	Probability of fixed #no. of successes over n Bernoulli trials, e.g., no. of heads in 10 coin flips.	$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n$
Poisson	Models counts of independent events within an interval, e.g., number of wildfire incidents per year.	$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots$
Geometric	Models #no. of trials until the first success, e.g., number of inspections until detecting deforestation.	$P(X = k) = (1-p)^{k-1} p, \quad k = 1, 2, \ldots$
Negative Binomial	Counts #no. failures until r successes occur, treats overdispersed or repeated deforestation detections.	$P(X = k) = \binom{k+r-1}{k} (1-p)^r p^k, \quad k = 0, 1, 2, \ldots$
Discrete Uniform	Assumes outcome in a finite set is equally likely, e.g., random sampling of inventory across a forest.	$P(X = x) = \frac{1}{n}, \quad x = 1, 2, \ldots, n$

Discrete distributions describe forestry monitoring scenarios where data outcomes are countable and finite. Common examples include the number of deforestation events, occurrences of wildfires, or counts of logged trees within a defined monitoring interval. Accurate representation of discrete events using appropriate distributions such as Binomial, Poisson, or Negative Binomial significantly enhances the accuracy of model predictions and uncertainty assessments. For instance, employing a Poisson distribution to model occurrences of illegal logging events can improve the precision of estimated deforestation emissions and reduce uncertainty around compliance risks.

In contrast, continuous distributions capture variables capable of taking any value within a specified range and are particularly relevant in forestry when modeling measurements such as tree heights, carbon stock densities, or biomass values. Continuous distributions like the Normal (Gaussian), Lognormal, Weibull, and Gamma distributions frequently arise in ecological modeling and biomass estimations due to their ability to realistically represent ecological variability and complex environmental factors. For example, using a Lognormal distribution for tree biomass data often provides more reliable estimates, particularly when the dataset is right-skewed due to natural variability in tree growth and forest conditions.

Central to these distributions are two mathematical concepts: Probability Mass Functions (PMFs) for discrete data and Probability Density Functions (PDFs) for continuous data. PMFs allocate specific probabilities to discrete outcomes, essential for accurately simulating events such as species occurrences or forest disturbances. PDFs describe the relative likelihood of continuous data points, enabling the robust estimation of variables like forest carbon content or annual biomass increment.

In Monte Carlo simulations, precise definition and utilization of PMFs and PDFs are crucial. These functions underpin random sampling processes that directly influence the reliability, precision, and credibility of uncertainty estimates. Given that forestry data is known to exhibit non-normal distributions due to inherent ecological heterogeneity that, informed selection and rigorous application of these functions are vital. Accurate modeling of the underlying data distribution enhances biomass and emissions estimates, significantly reduces uncertainty, and bolsters the financial and ecological credibility of REDD+ reporting initiatives (Morgan & Henrion, 1990; IPCC, 2019; ART, 2021).

Practitioners are encouraged to conduct exploratory data analysis early in their project planning stages, integrating statistical tests of normality and visual assessments (histograms, kernel density plots, Q-Q plots). Such preliminary analyses assist in diagnosing data distributions accurately, improving model selection, reducing potential auditor findings, and ultimately enhancing the financial and environmental outcomes of national REDD+ monitoring programs.

Method

Import

```r
# Point this to the correct path where your file is located:
workbook = "./data/art/GuyanaARTWorkbookMC-thru2022-April2024_values.xlsx"
CarbonStocks = readxl::read_excel(workbook, "CarbonStocks") |>
    janitor::clean_names() |>
    mutate(across(where(is.numeric), ~round(.x, 1)))
CarbonStocks_MC = readxl::read_excel(workbook, "CarbonStocks (MC)") |>
    janitor::clean_names() |>
    mutate(across(where(is.numeric), ~round(.x, 1)))

DeforestationEF = readxl::read_excel(workbook, "Deforestation EFs") |>
    janitor::clean_names() |>
    mutate(across(where(is.numeric), ~round(.x, 1)))
DeforestationEF_MC = readxl::read_excel(workbook, "Deforestation EFs (MC)") |>
    janitor::clean_names() |>
    mutate(across(where(is.numeric), ~round(.x, 1)))

DegradationEF = readxl::read_excel(workbook, "Degradation EFs") |>
```