M12 Simulations

Topics in Insurance, Risk, and Finance

Yuyu Chen

Department of Economics, University of Melbourne

2026

Table of contents

Learning outcomes

- Understand why Excel is needed for evaluating cash series.
- Can describe the process of simulation.
- Understand the theoretical basis for simulations and the Monte Carlo estimator.
- Can compute the Monte Carlo estimator and its confidence interval.
- Implement simulations to model rates of return and calculate related risk metrics in Excel.

Why Excel

- We can use Excel (also other programming languages) to simulate random variables, and use these to generate random observations of cash flow series.
- These generations will result in an empirical distribution of the cash flow series. We can then use this distribution to estimate moments, quantiles, ranges, and probabilities.
- Simulation is widely used in the work by actuaries.
- Watching someone else use Excel has limited value, however, and so you must attempt these in your own time.

Table of contents

Generating random numbers: uniform distribution

- A computer cannot generate true random numbers.
- Instead, Excel (and other programs) generate approximations that are called pseudorandom numbers.
- Once the parameter of the simulation is fixed, the same random samples will be generated each time.
- We will refer to pseudorandom numbers as random for brevity.
- There are multiple ways to generate random numbers in Excel.
- To generate random numbers from a uniform distribution on $(0, 1)$, we can use the `RAND()` (or `RANDARRAY`) function, which gives us realizations of independent and identically distributed uniform random variables.
- Generating uniform random numbers is at the core of simulations.

Inverse transformation

Theorem

Let F be a distribution function and F^{-1} be its generalized inverse, i.e., for $p \in (0,1]$

$$F^{-1}(p) = \inf\{x : F(x) \geqslant p\},$$

with the convention that $\inf \emptyset = \infty$.

- *Inverse transformation (quantile transformation): If $U \sim U(0,1)$, then $\mathbb{P}(F^{-1}(U) \leqslant x) = F(x)$.*
- *Probability transformation: If X has a continuous distribution F, then $F(X) \sim U(0,1)$.*

Generating random numbers: inverse transformation

- Therefore, we can use inverse transformation to generate random numbers of a distribution as long as the (generalized) inverse of the distribution is known (i.e., VaR).
- That is, for a random variable X, we first generate a sample of a standard uniform random variable $u_1, \ldots, u_n$. Then, by inverse transformation, a sample of $X \sim F$ is

$$F^{-1}(u_1), \ldots, F^{-1}(u_n).$$

- Note that not all distributions have nice formulas for their generalized inverse (e.g., normal distributions). One may need to rely on z-table for normal distributions.

Generating random numbers: inverse functions in Excel

- The inverse functions of many commonly used distributions are available in Excel.
- The general form of such inverse function is `F.INV(p,...)`: If `p = F.DIST(x,...)`, then `F.INV(p,...) = x`.
- Here, `F.DIST` is the distribution function.

Some inverse functions are summarized below

- `BETA.INV` for Beta distribution
- `LOGNORM.INV` for lognormal distribution
- `NORM.INV` for normal distribution
- `NORM.S.INV` for standard normal distribution

To see more inverse/probability functions Excel, go to Formulas → More Functions → Statistical.

Generating random numbers: example A

Generate random numbers from normal distributions.

- To get a standard normal random number, we take the inverse cumulative normal of a uniform random variable in the range $(0, 1)$. In Excel, `NORMSINV(RAND())` will generally do.
- To get a large sample of random numbers, one can use `NORMSINV(RANDARRAY())`.
- Plot a histogram of the sample.
- To get a general normal random variable $X \sim N(\mu, \sigma^2)$, from a standard normal random variable Z, we simply take

$$X = \mu + \sigma Z.$$

Generating random numbers: example B

Generate random numbers from a uniform distribution F on the interval (a, b) where $b > a$.

- We first find the inverse function of F, which is

$$F^{-1}(p) = a + (b - a)p.$$

- To get a random number from the uniform distribution, we use $a + (b - a)\times$ `RAND()`.
- To get a large sample of random numbers, one can use $a + (b - a)\times$ `RANDARRAY()`.
- Plot a histogram of the sample.

Generating random numbers: example C I

Generate random numbers from the following distribution

$$F(x) = \begin{cases} 1 - \frac{1}{x+1}, \quad 0 \leqslant x < 9, \\ 0.95, \quad 9 \leqslant x < 10, \\ 1, \quad x \geqslant 10. \end{cases}$$

Generating random numbers: example C II

We first find the inverse function

$$F^{-1}(p) = \begin{cases} \frac{1}{1-x} - 1, & 0 < p \leqslant 0.9, \\ 9, & 0.9 < p \leqslant 0.95, \\ 10, & 0.95 < p \leqslant 1. \end{cases}$$

Generating random numbers: example C III

- We first generate a random number from a standard uniform distribution using `RAND` (or a sample using `RANDARRAY`). Store it in A1.
- To write the inverse function, we use the `IF` function: `IF(condition, value1 if the condition is true, value2 if the condition is false)`
- Logic operator: `AND(condition1,condition2)`, `OR(condition1,condition2)`
- The inverse function can be written as `IF(A1<=0.9,1/(1-A1)-1,IF(AND(A1>0.9, A1<=0.95),9,10))`

Monte Carlo estimator

- Given a sample of a random variable X, we are now able to approximate
$$\theta = \mathbb{E}(X).$$
- Let $X_1, \ldots, X_n$ be iid copies of X. The Monte Carlo estimator of θ can be obtained by
$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$
- Note that this estimator also works for functions of random variables, i.e., $f(X)$.
- If $f(x) = x^k$, we can approximate the kth moment.
- If $f(x) = \mathbb{1}_{\{x \in [a,b]\}}$, we can approximate the probability of landing in $[a, b]$.

Theoretical basis of simulation: Why does it work?

- $\hat{\theta}$ is unbiased, i.e., $\mathbb{E}(\hat{\theta}) = \theta$.
- The Law of Large Numbers: for iid random variables $X_1, \ldots, X_n$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}(X_1)$$

  with probability 1. Moreover, for iid random variables $Y_1, \ldots, Y_n$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f(Y_i) \to \mathbb{E}(f(Y_1))$$

  with probability 1.
- Also

$$\mathrm{Var}(\hat{\theta}) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\mathrm{Var}(X).$$

  Hence, if n is large, $\hat{\theta}$ can be very close to θ.

Theoretical basis of simulation: standard error

- By the Central Limit theorem, the distribution of $\sum_{i=1}^{n} X_i/n$ can be approximated by $N(\mathbb{E}(X_1), \mathrm{Var}(X_1)/n)$.
- So the error of the simulation, i.e., $\sum_{i=1}^{n} X_i/n - \mathbb{E}(X_1)$ is normally distributed with mean 0 and variance $\mathrm{Var}(X_1)/n$. The standard deviation $\sqrt{\mathrm{Var}(X_1)/n}$ is called the standard error.
- A problem is that we do not know the variance of X (even the expectation is unknown, otherwise simulation is redundant).
- For this, we can replace $\mathrm{Var}(X)$ by sample variance

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\theta})^2,$$

which is an unbiased estimator of $\mathrm{Var}(X)$.

Theoretical basis of simulation: confidence interval I

- By the Law of Large numbers, $s_n^2 \to \mathrm{Var}(X)$ with probability 1.
- Then with the Central Limit Theorem and Slutsky's Theorem[1], we have the following result.

Theorem

Let X_n be an iid sequence with finite mean and variance with $\mathbb{E}(X_1) = \theta$. Then

$$\frac{\hat{\theta} - \theta}{s_n/\sqrt{n}} \to N(0,1),$$

in distribution.

[1] You do not need to know this.

Theoretical basis of simulation: confidence interval II

By the previous theorem, a $1-\alpha$ confidence interval of θ is

$$\left(\hat{\theta} - z_c \frac{s_n}{\sqrt{n}}, \hat{\theta} + z_c \frac{s_n}{\sqrt{n}}\right)$$

where $\mathbb{P}(-z_c < Z < z_c) = 1 - \alpha$. Here $Z \sim N(0,1)$.

Example: MC estimate I

Suppose that interest rates $i_1, i_2, i_3 \sim U(0, 0.5)$ are independent. An agent simulates 3 paths of i_1, i_2, i_3 using inverse transformation. The random numbers, generated from a standard uniform distribution, for i_1, i_2, i_3 are documented sequentially below:

path 1: $0.04875077, 0.34849265, 0.70849849$
path 2: $0.69201482, 0.33769932, 0.13393407$
path 3: $0.60479379, 0.41474126, 0.14565376$

Compute a Monte Carlo estimate of the mean of $S_3 = \prod_{t=1}^{3}(1 + i_t)$ with a 90% confidence interval.

Example: MC estimate II

The simulated values of $i_1, \ldots, i_3$ are:

path 1: $0.02437538, 0.17424633, 0.35424924$
path 2: $0.34600741, 0.16884966, 0.06696703$
path 3: $0.30239689, 0.20737063, 0.07282688$

Hence S_3 in each path is:

path 1: $S_3 = 1.629$
path 2: $S_3 = 1.679$
path 3: $S_3 = 1.687$

Example: MC estimate III

Therefore, the Monte Carlo estimate of $\mathbb{E}(S_3)$ is

$$\frac{1}{3}(1.629 + 1.679 + 1.687) = 1.665,$$

and $s_n^2 = 0.000988$. The confidence interval is

$$(1.665 - 1.65 * 0.01815, 1.665 + 1.65 * 0.01815).$$

Table of contents

Summary statistics

Some functions:

- mean: `AVERAGE()`
- variance of a sample: `VAR()`
- variance of a population: `VARP()`
- standard deviation of a sample: `STDEV()`
- standard deviation of a population: `STDEVP()`
- percentile: `PERCENTILE()`

Analysis ToolPak

Load the Analysis ToolPak in Excel for Mac

- Click the Tools menu, and then click Excel Add-ins.
- In the Add-Ins available box, select the Analysis ToolPak check box, and then click OK.
- If Analysis ToolPak is not listed in the Add-Ins available box, click Browse to locate it.
- If you get a prompt that the Analysis ToolPak is not currently installed on your computer, click Yes to install it.
- Quit and restart Excel. Now the Data Analysis command is available on the Data tab.

Analysis ToolPak

Data Analysis can complete basic statistical tasks including:

- generate random numbers (we can use seed here)
- summary statistics
- frequency table

Table of contents

Exercise A: S_n I

- Recall that if rates are iid, even if the accumulation factors do not follow a lognormal distribution, one can still use lognormal distribution to approximate S_n or V_n.
- This can be seen by noting that $\log S_n = \sum_{t=1}^{n} \log(1 + i_t)$. By central limit theorem, $\log S_n$ can be approximated by a normal distribution for large n.
- Next, we will assess the performance of using lognormal distribution as an approximation.

Exercise A: S_n II

Suppose that in a varying rate model, $i_1, \ldots, i_{30} \sim U(0.06, 0.10)$. Complete the following tasks:

- Generate a random sample of $(1 + i_t)$ with size 6000×30.
- Compute the first and second moments of S_{30} numerically.
- Plot the mean of S_{30} against the simulation number $k = 1, \ldots, 6000$.
- Compute the first and second moments of S_{30} analytically and compare them with the numerical result.
- Use the analytical results for moments of $\log S_{30}$ as the corresponding parameters of a lognormal distribution. Compare the histogram of the sample of S_{30} and the density of the lognormal distribution. Explain the observation.
- Use `FREQUENCY` to generate frequency table given Bin series.

Compare the simulated results with lognormal distribution.

Exercise A: S_n III

We compute the first and second moments of S_{30} analytically.

Exercise A: S_n IV

Some observations:

- The mean and second moment of the sample are close to the analytical solution.
- As the simulation number increases, the mean of S_{30} becomes more and more stable and it is closer to the true value.
- The simulated distribution of S_{30} is very close to the lognormal distribution. This is due to the central limit theorem: as $\log S_{30}$ is close to a normal distribution, S_{30} is close to a lognormal distribution.

Exercise A: S_n Remarks

Some remarks:

- To fill a column/row, use the fill function (Home $\rightarrow$ Editing $\rightarrow$ Fill $\rightarrow$ Series).
- To select a large range of cells, use the name box in the top left of the workbook (e.g., A1:B2).
- When using the autofill function the columns should be adjacent.
- Although Excel states the lognormal function as `LOGNORM.DIST(x,mean,standard dev)`, the mean and standard deviation in this function refer to the equivalent normal distribution mean and standard deviation.

Exercise B: A_n I

Suppose that $1 + i_1, \ldots, 1 + i_{10} \sim LN(0.09, 0.04^2)$ are independent. Complete the following tasks.

- Calculate the mean and standard deviation of A_{10} using recursive formulae.
- Calculate the mean and standard deviation of A_{10} numerically using 5000 simulations.
- Estimate the 5% and 95% percentiles of A_{10}.
- Estimate $\mathbb{P}(A_{10} < 14.75)$.
- Get summary statistics of A_{10}.

Exercise B: A_n II

The recursive formulae: Since $A_n = (1 + i_n)(1 + A_{n-1})$, we have

$$\mathbb{E}(A_n) = \mathbb{E}((1 + i_n))(1 + \mathbb{E}(A_{n-1})),$$

and

$$\mathbb{E}(A_n^2) = \mathbb{E}((1 + i_n)^2)(1 + 2\mathbb{E}(A_{n-1}) + \mathbb{E}(A_{n-1}^2)).$$

Simulations of the compound Poisson distribution

Complete the following tasks in Excel.

- Generate 500 samples from $Poisson(\lambda)$ where $\lambda = 5, 10, 20, 30$.
- Generate 500 samples from the compound Poisson distributions, for which the individual claim follows the exponential distribution with mean 1.
- Compute the sample means of the above simulations and compare with theoretical values.
- Draw histograms of the samples of the compound Poisson distributions for $\lambda = 5, 10, 20, 30$. What is one key observation here?

Simulations

Assume that $N(t)$ is a Poisson process with parameter 30, the individual claim amount distribution is lognormal with parameters $\mu = 3$ and $\sigma^2 = 1.1$, $c = 1200$, and $u = 1000$. Find $P(U(2) < 0)$ by simulations.

Simulations

In a classical risk process,

- X_i follows the Gamma distribution with parameters $\alpha = 0.5$ and $\beta = 2$;
- average number of claims per unit of time is 1;
- premium loading factor is 0.1.

Find $\psi_4(3.5)$ using simulations.