Monte Carlo Method:
a powerful tool for the Simulation of Pharmaceutical Processes

Before delving into the topic, it is worth clarifying that:

the reference to the "casino" is not at all extravagant!

Before the advent of modern computers, games of chance were the "artificial laboratory" in which the Theory of Probability was developed.

In fact, the most reliable random number generator is exactly roulette.

D. Cooper, B. Grinder, *Risk Management.pdf (moaf.org)*, Financial History (Winter 2009)

Why are we talking about Monte Carlo?

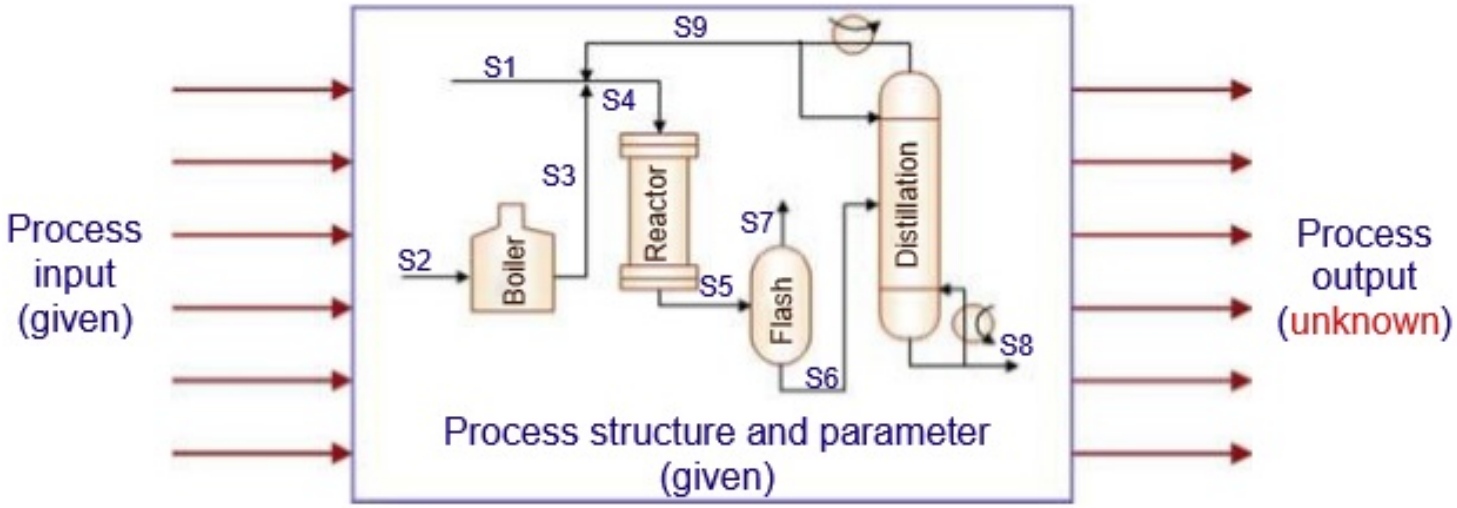
- In pharmaceutical industry, even a tiny mistake can cost a lot—both in terms of money and the quality of the medicine we produce.

- That's why we can use something called ***simulation methods.***

  In this context, *simulation* obviously means *imitation* and *reproduction of phenomena.*

Why are we talking about Monte Carlo?

- The aim for *process simulation* is to predict how a defined process would actually behave under a given set of operating conditions.



- We can think of *simulation methods* as “practical exercises” that help us predict potential problems before they occur.

Monte Carlo simulation

- Our focus is on a specific type of simulation called ***Monte Carlo simulation***.
- **A good Monte Carlo simulation starts with a solid understanding of how the underlying process works**.
- We'll explore how Monte Carlo simulation can help us understand and control the process of producing better pharmaceutical products through simple case studies.

What is Monte Carlo method?

- Imagine you're trying to understand a dartboard by throwing darts at it.
- Where these darts land gives us clues about the shape and characteristics of the dartboard.
- It is obvious that *the more darts we throw, the better we will be able to know our target !*

How does Monte Carlo work?

- Monte Carlo works by taking many *'random samples'*—like our darts—and using these to make educated guesses about the whole situation.

- It's like learning about an entire beach by studying a handful of sand ☺

How does Monte Carlo work?

- Simply put, a Monte Carlo simulation generates random input values and calculates how a given system (*e.g.*, a chemical reaction, a manufacturing process, an industrial operation, *etc.*) responds to those random inputs.

- ***It is just matter of defining a mathematical function (or transfer function) that represents the system and then Monte Carlo simulation can be easily applied.***

A Little History

1930s Enrico Fermi uses random numbers to calculate the properties of neutrons

1940s After the advent of computers, scientists at the Los Alamos National Laboratories (S. Ulam and J. von Neumann) refined Monte Carlo simulation and used it to predict the effects of nuclear explosions.

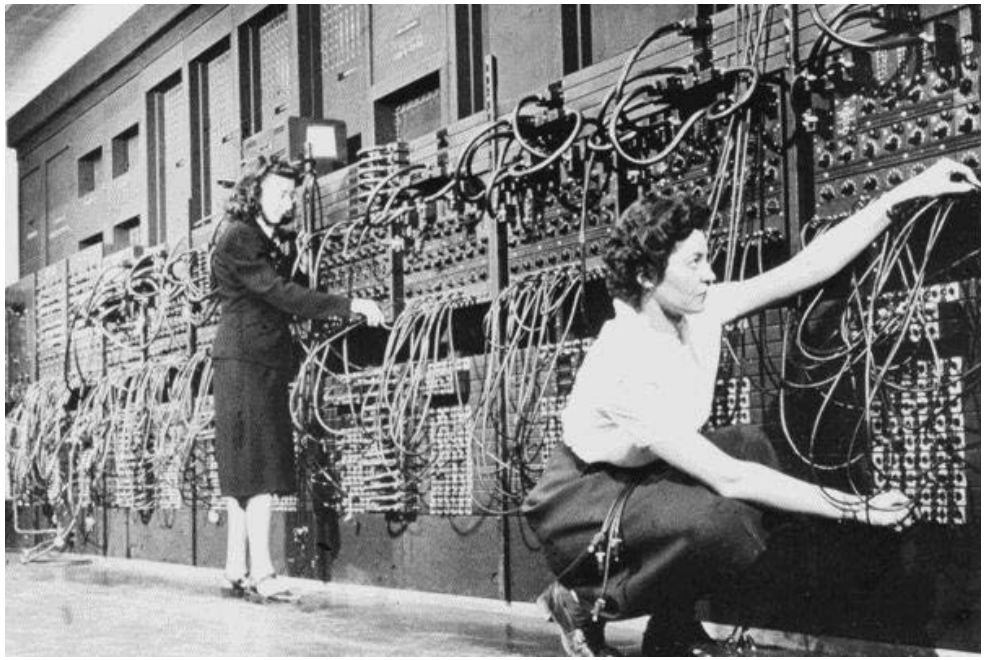


Image of ENIAC programmers in the 40s

N. Metropolis, *The beginning of the Monte Carlo Method*, Los Alamos Science, Special Issue (1987) pp.125-130
R. Eckhardt, *Stanislaw Ulam, John von Neumann, and the Monte Carlo Method*, Los Alamos Science, Special Issue (1987) pp.131-141

Advantages & Limitations

- The **advantages** of Monte Carlo simulation are:
  - *easy to understand and visualize*,
  - *flexible*, it doesn't need a lot of assumptions about what you're studying,
  - *widely applicable*: mechanics, aerodynamics, project management, finance, astrophysics, meteorology, *etc*.
- The main **disadvantage** is that it can require a lot of computer power, especially for more complex problems, but this is no longer an issue ☺

Where do we use Monte Carlo in pharmaceutical manufacturing?

- essentially where we want to understand/forecast "unpredictable" situations (*e.g.*, process optimization, risk assessment, DoE, supply chain management, *etc.*)

- as tool that forecasts the impact of changes to ingredients or manufacturing processes on the quality of the finished product before a product is ever manufactured. With Monte Carlo simulation we can in fact “produce” an almost limitless number of “virtual batches” of a drug.

Where do we use Monte Carlo in pharmaceutical manufacturing?

- We will now see five case studies relating to as many practical situations in the pharmaceutical field, in particular:
  - *a crystallization process*
  - *an API manufacturing process*
  - *a micronization process*
  - *robustness of an analytical method*
  - *stability analysis*

Where do we use Monte Carlo in pharmaceutical manufacturing?

- These examples, although not exhaustive of all possibilities and highly simplified, are intended to show the practical usefulness and versatility of "Monte Carlo Simulations" in different scenarios within the pharmaceutical field

- The R scripts for the case studies illustrated below can be downloaded at:

https://github.com/rbonfichi/monte-carlo-simulation

CASE STUDY 1 :

a crystallization process

CASE STUDY 1

- Let's assume we have an empirical equation for a given crystallization process:

$$\textit{Crystal Yield (kg)} = a \times (\textit{Solute Concentration}) + b \times (\textit{Temperature}) + c \times (\textit{Pressure})$$

where *a*, *b*, and *c* are constants determined through experimentation or literature.

For example, this equation, including constants, could result from a DoE study.

- Clearly this is just a simplified example. In real-world scenarios, it could be more complex.

CASE STUDY 1

- At this point, a Monte Carlo simulation would proceed as follows:

1. Randomly select a value for each input parameter within its range (*e.g.*, Solute Concentration between 0.5 M and 1.5 M, Temperature between 20°C and 30°C, *etc.*)

2. Use these values in the empirical equation to calculate the output value (*i.e.*, Crystal Yield)

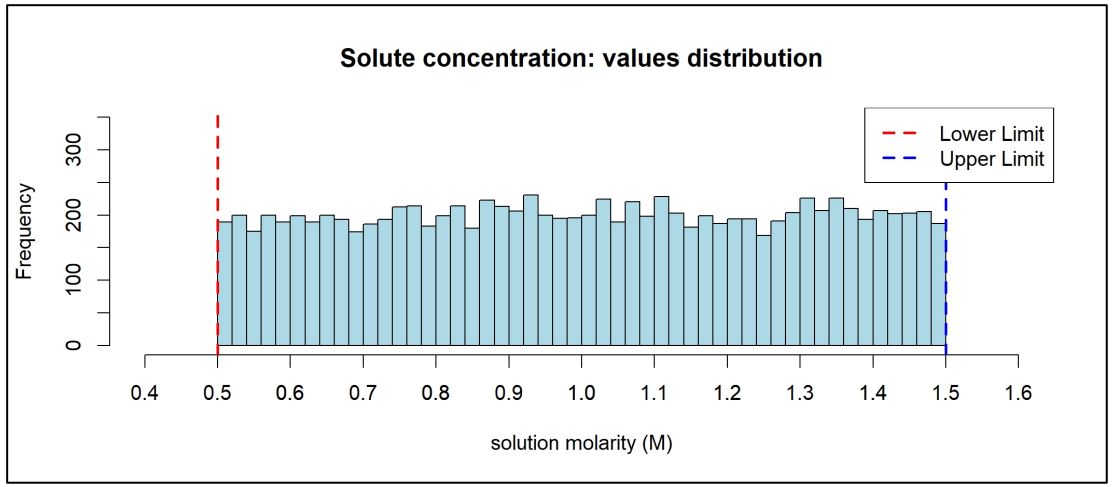
3. Repeat the process multiple times (*e.g.*, 10000) to get a distribution of the output.

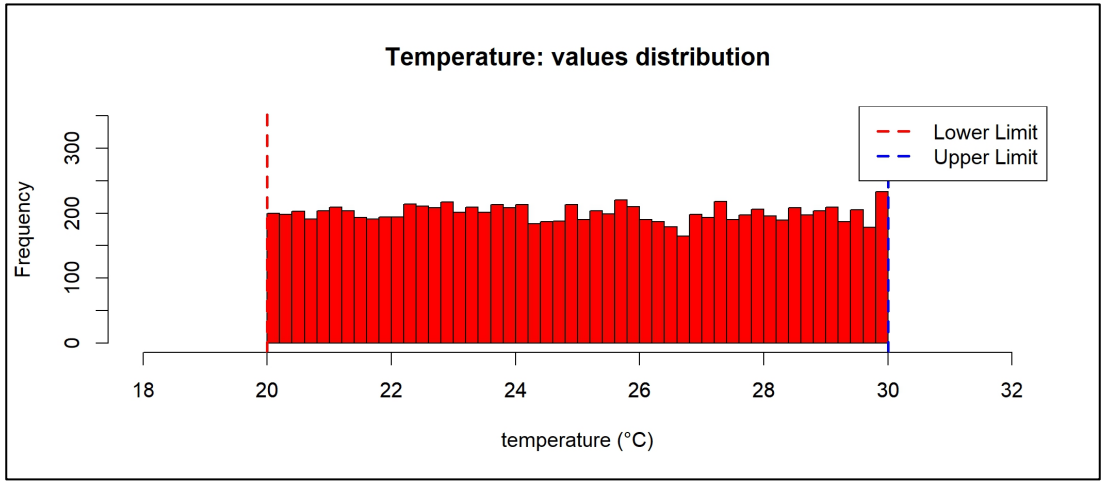
4. Aggregate the results and visualize them using a histogram, then calculate summary statistics such as the mean and standard deviation.

CASE STUDY 1

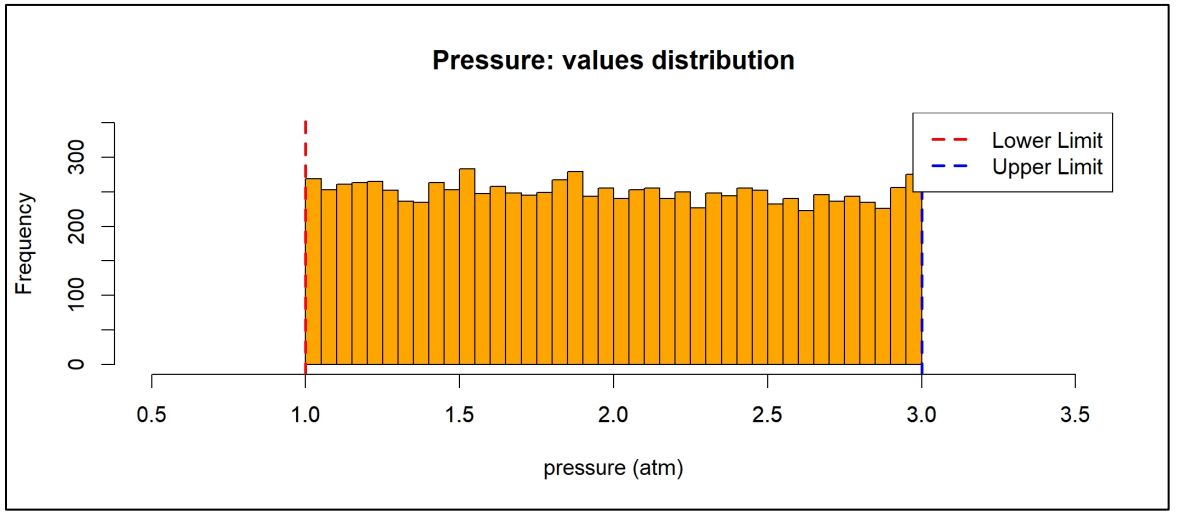
- As an example, let's imagine that:
  - a = 2 b = 1.5 c = 0.8
  - Solute Concentration: 0.5 M - 1.5 M
  - Temperature: 20°C - 30°C
  - Pressure: 1 atm - 3 atm

CASE STUDY 1





DISTRIBUTIONS OF VALUES



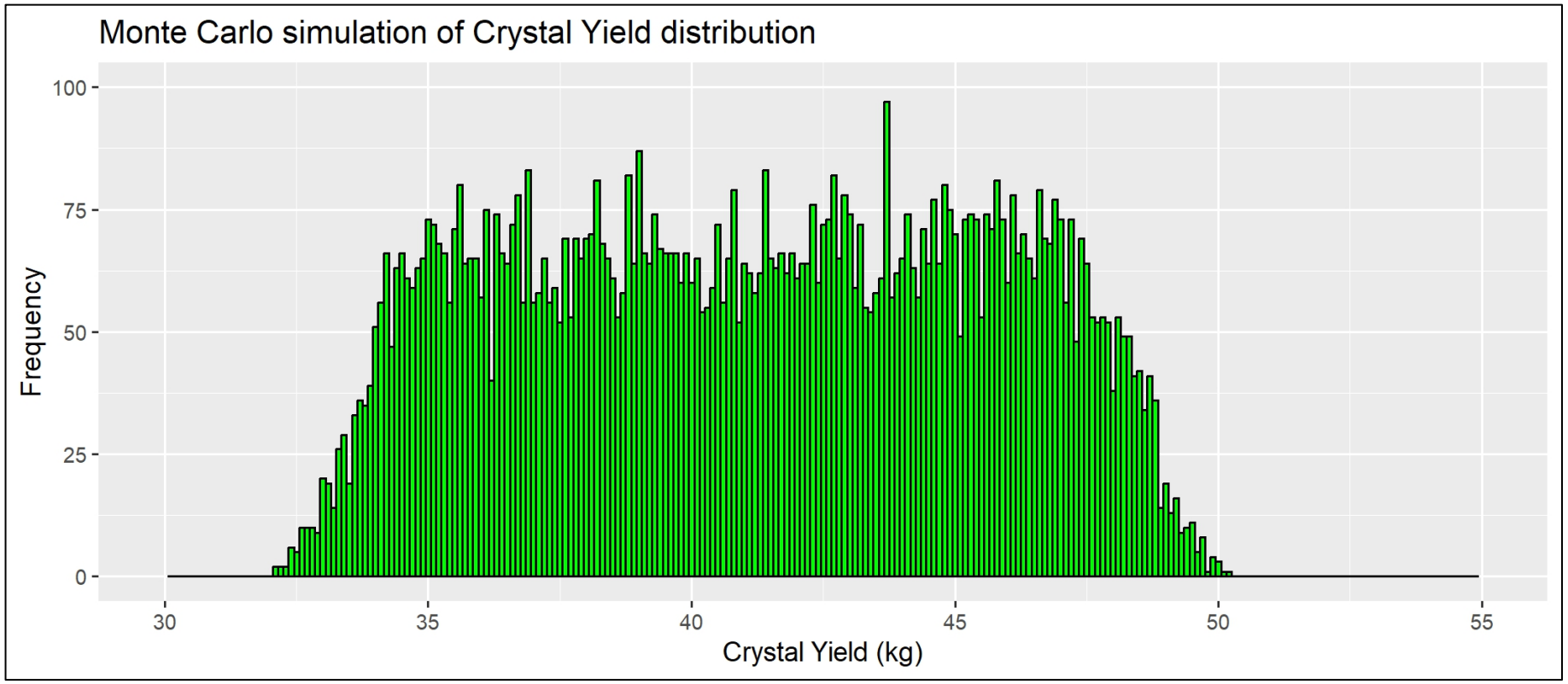
CASE STUDY 1

- Conducting a Monte Carlo simulation in R based on a hypothetical empirical equation:

Crystal Yield (kg)= 2 × (Solute Concentration) + 1.5 × (Temperature) + 0.8 × (Pressure)

we get the results shown in the next slide.

CASE STUDY 1



mean crystal yield	sd crystal yield	min crystal yield	max crystal yield
41.09	4.39	32.16	50.16

CASE STUDY 1

What does this simulation tell us and why is it useful ?

- The histogram and summary statistics give you an idea of the range and distribution of the crystal yield.
- This helps you understand the uncertainty associated with the process. For instance, knowing that the yield could vary between 32 kg and 50 kg is valuable information for Quality Control and Production planning.

CASE STUDY 1

What does this simulation tell us and why is it useful ?

- **Risk Assessment:** Knowing the distribution of possible outcomes helps in assessing risks. For example, if a minimum yield of 40 kg is required for profitability, we can calculate the probability of falling below this threshold.
  In fact, assuming that the data are normally distributed, it can be estimated that:

mean crystal yield	sd crystal yield	95% CI on mean	**Probability of Yield below 40kg**
41.04	4.41	40.96 - 41.13	**40.71 %**

CASE STUDY 1

What does this simulation tell us and why is it useful ?

- **Decision Support & Resource Allocation:** If one parameter/step has a significantly higher impact on yield, it may be worthwhile to invest in better control mechanisms for that parameter/step. This information aids in making informed decisions (see next case study).

- **Benchmarking:** The mean and standard deviation can serve as benchmarks for actual production. Deviations from these benchmarks could be indicators of anomalies or shifts in the process.

CASE STUDY 2 :

an API manufacturing process

CASE STUDY 2

- Suppose you have extracted from pilot studies data of a given API a relationship that links together:

  Assay, total impurities, residual humidity (LOD) and residual solvents content

  and which, by hypothesis, looks like:

$$Assay = 101.255 - (4.647 \times LOD) - (22.58 \times Total\ Impurities) - (210.4 \times Solv1) + (125.8 \times Solv2) + (171.0 \times Solv3)$$

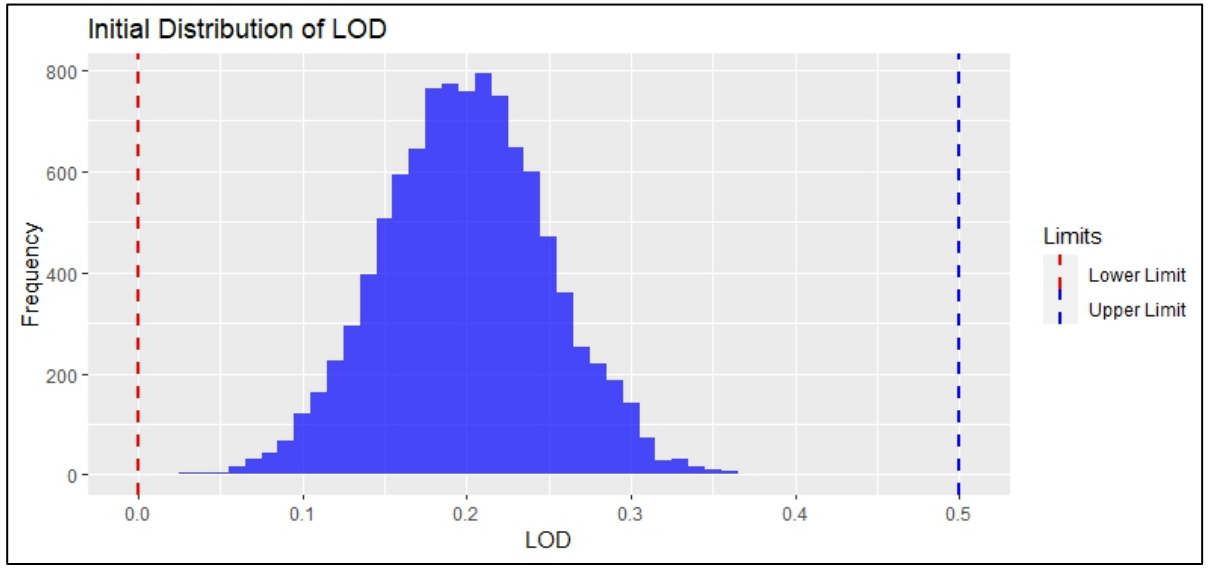
CASE STUDY 2

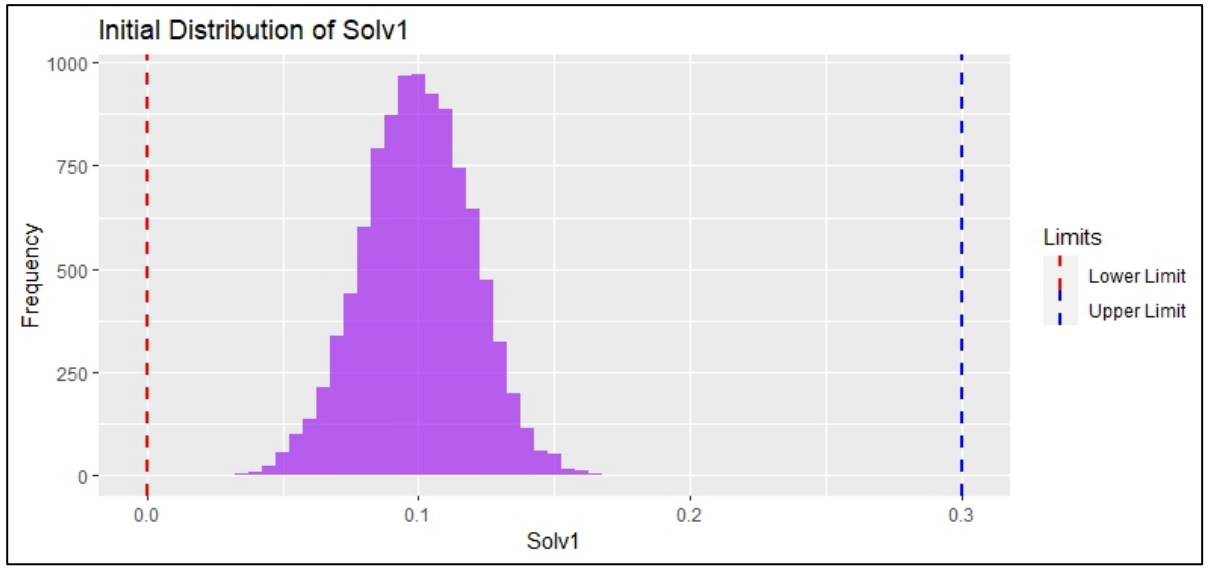
- Let's assume that, based on the data available to us, we know that:

	Spec. limits	Distribution
LOD	≤ 0.5%	~ N (0.2, 0.05)
Total Impurities	≤ 0.5%	~ N (0.1, 0.02)
Solv1	≤ 0.3%	~ N (0.1, 0.02)
Solv2	≤ 0.3%	~ N (0.1, 0.02)
Solv3	≤ 0.08%	~ N (0.02, 0.005)

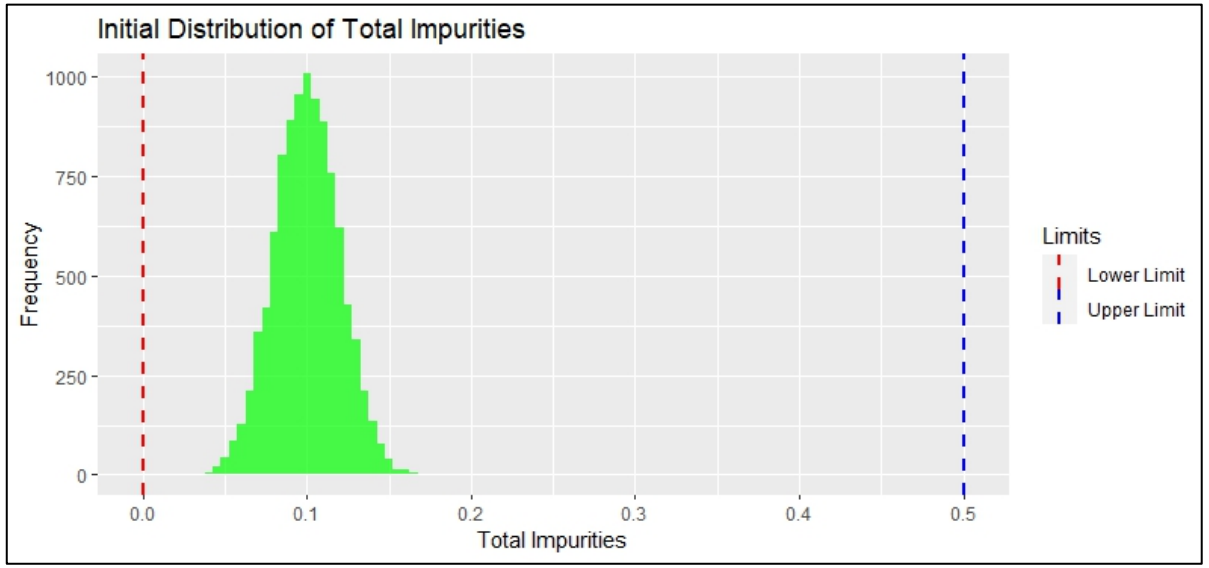
INITIAL DISTRIBUTIONS

CASE STUDY 2

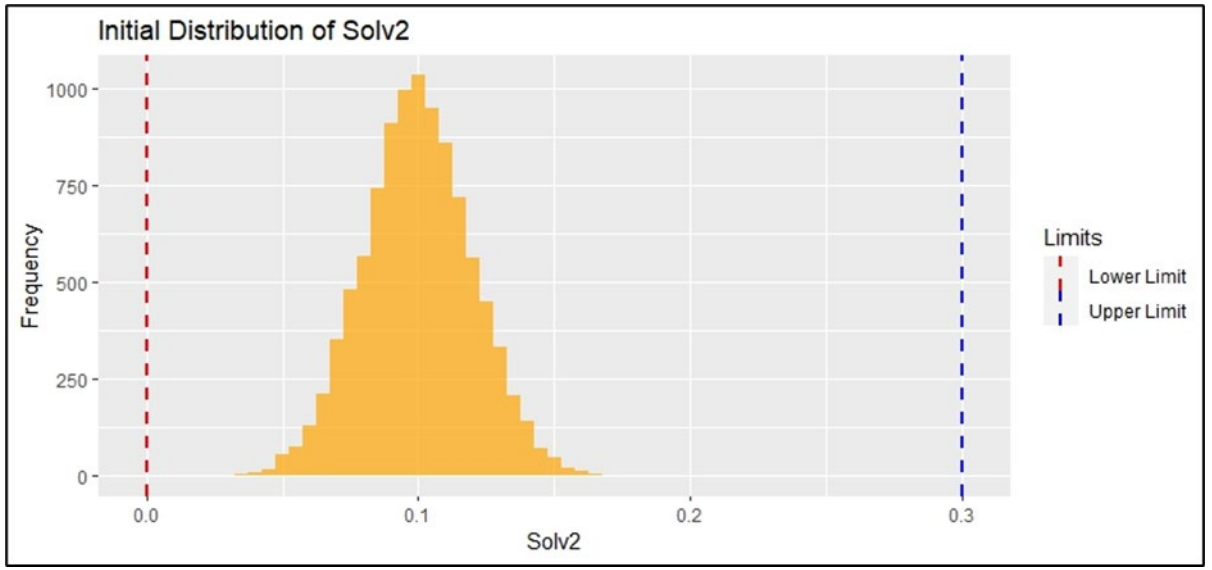




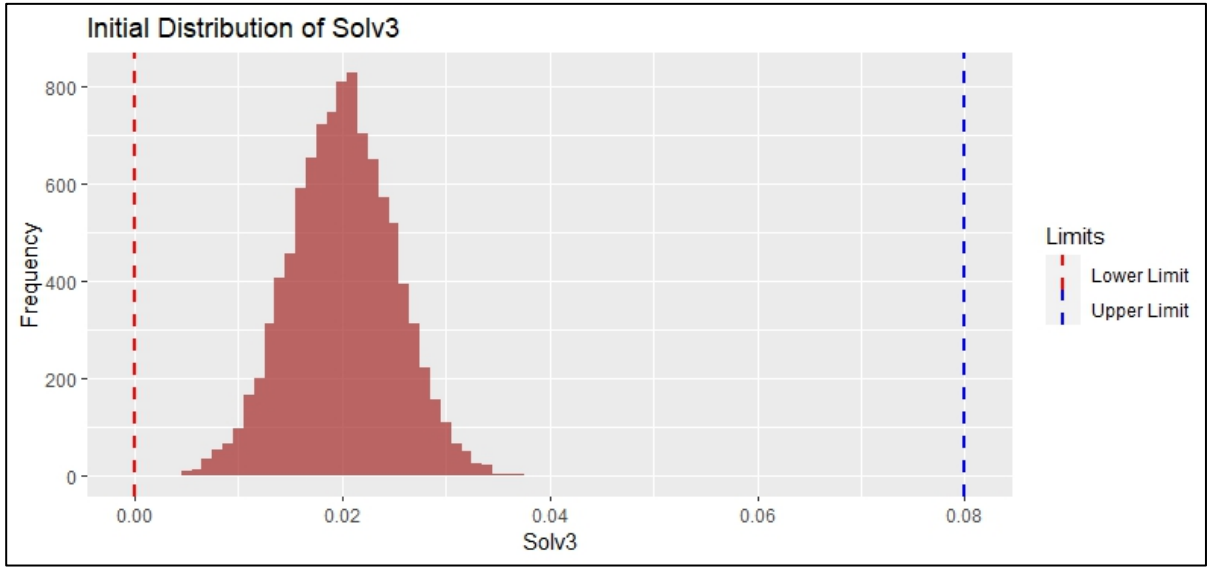
INITIAL DISTRIBUTIONS



CASE STUDY 2

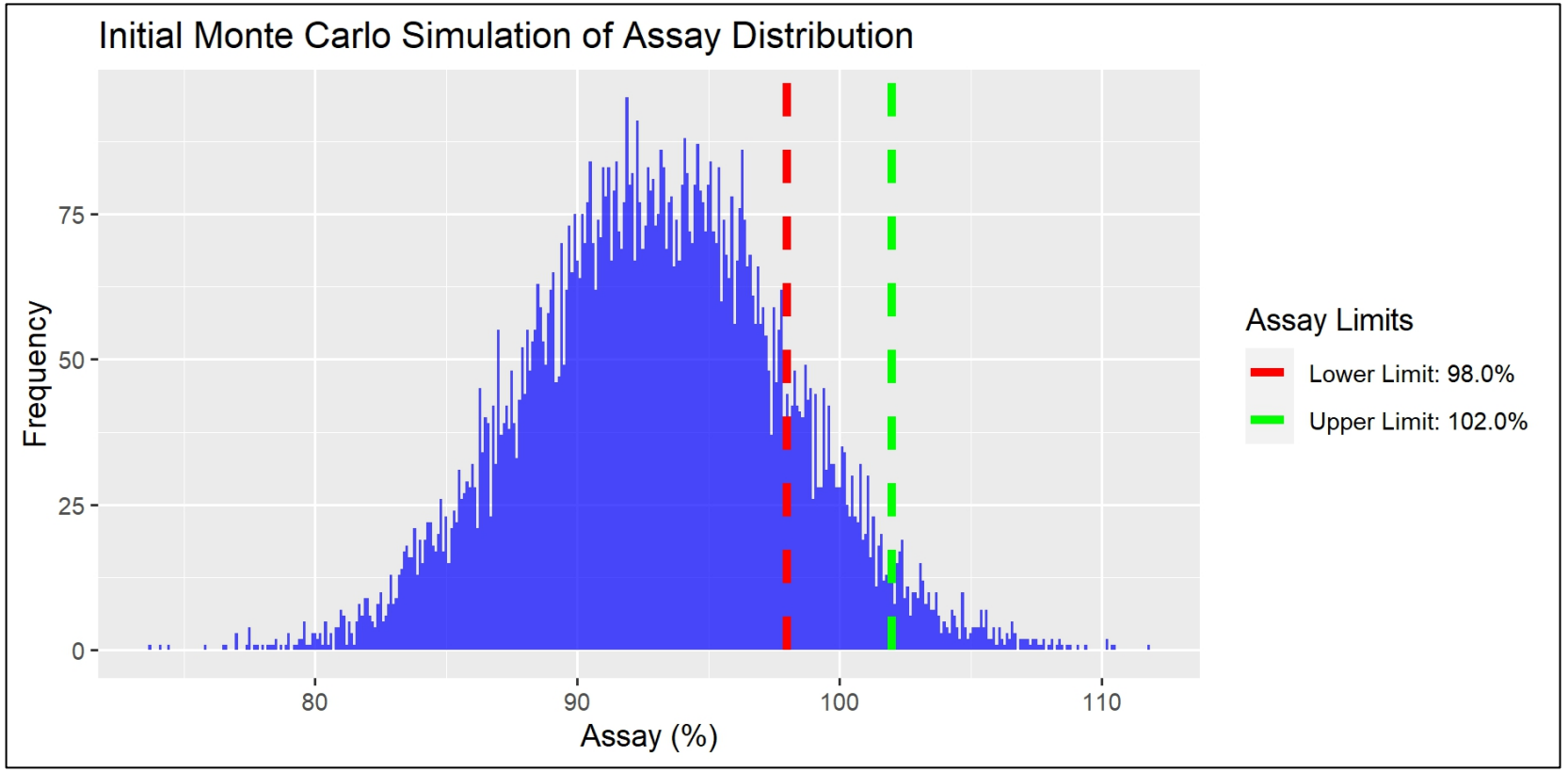


INITIAL DISTRIBUTIONS



CASE STUDY 2

- Conducting a Monte Carlo simulation in R we get the following:



Percentage within limits (Initial): 12.44

Percentage outside limits (Initial): 87.56

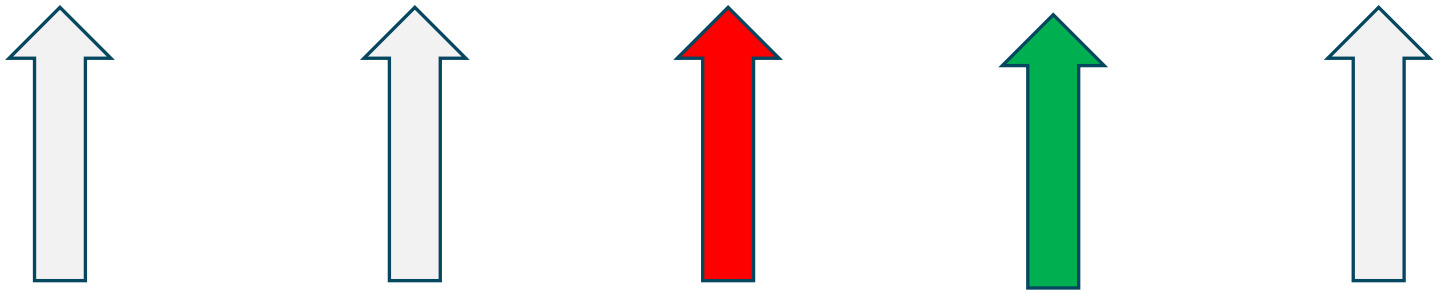
Cp (Initial): ~ 0.40

Cpk (Initial): - 0.97

CASE STUDY 2

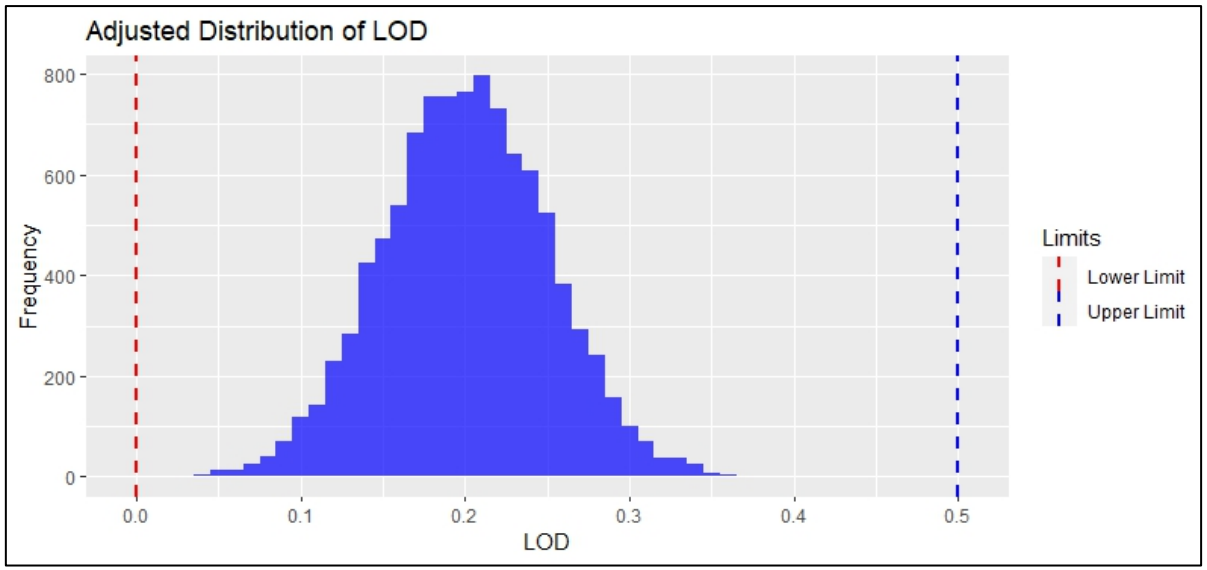
Spearman Rank Correlation with Assay

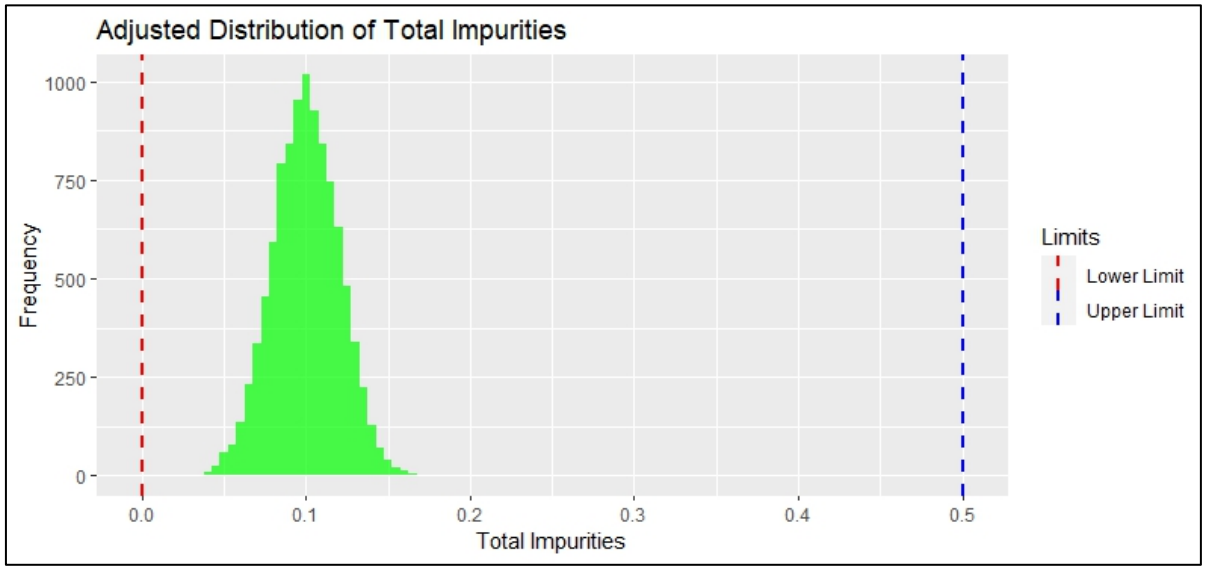
Assay	LOD	Total Imp.	Solv1	Solv2	Solv3
1.00	-0.04	-0.07	-0.84	0.50	0.17

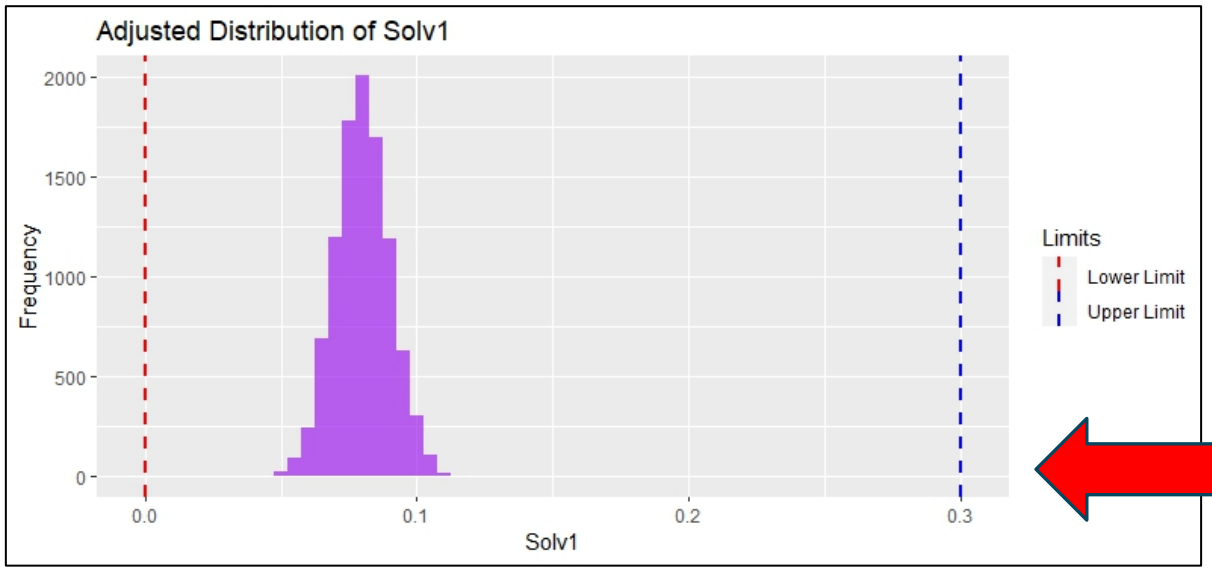


Weak correlation	Weak correlation	Strong negative correlation	Strong positive correlation	Weak correlation

ADJUSTED DISTRIBUTIONS

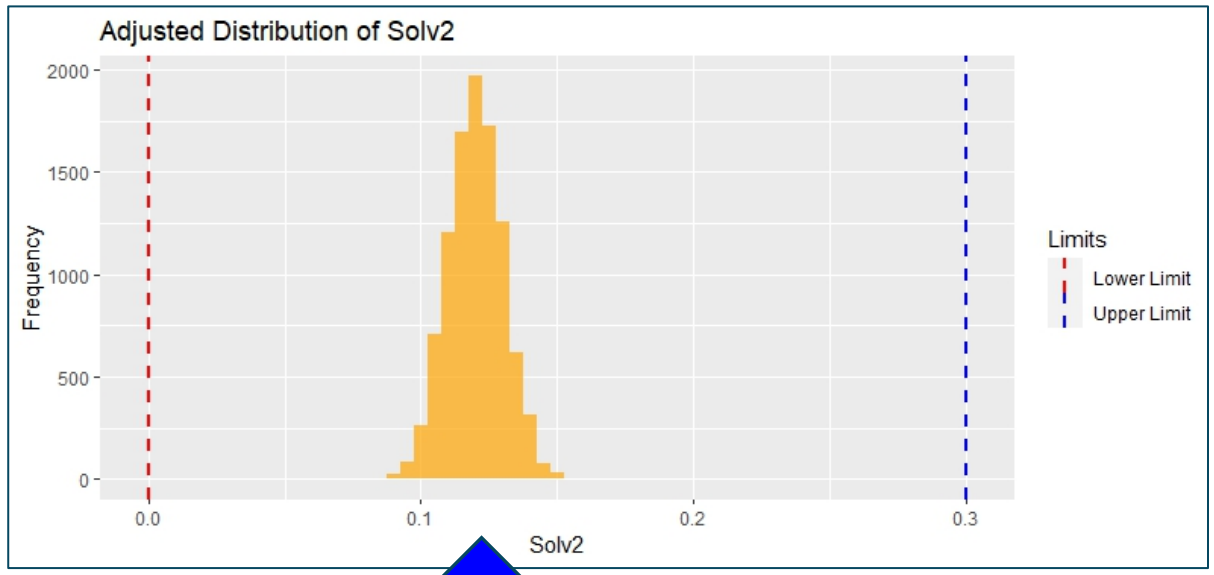




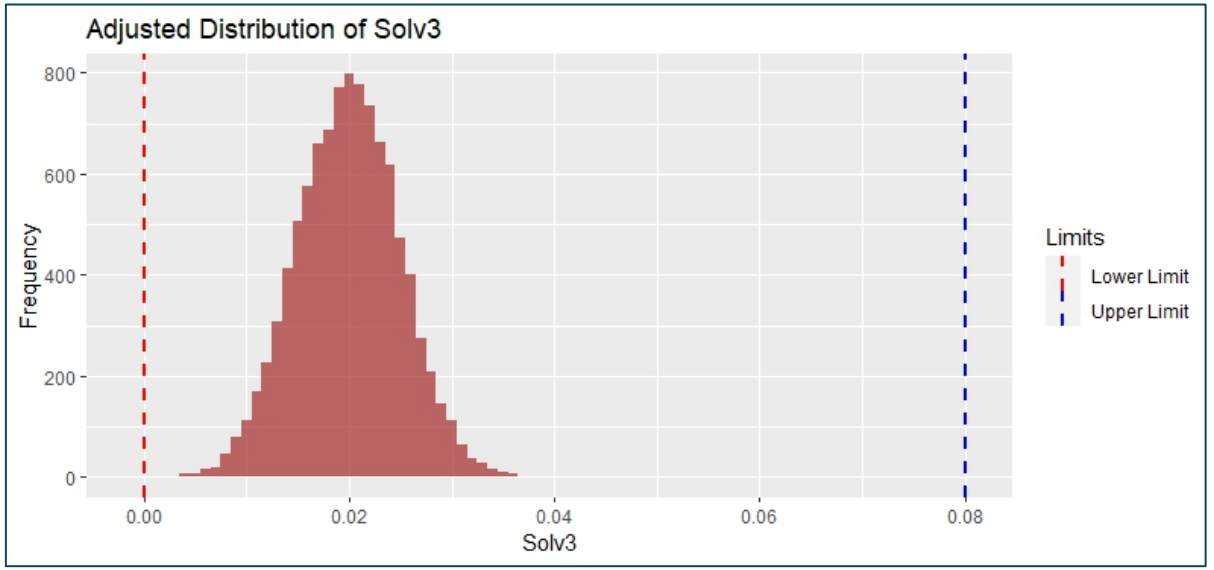


Since Solv1 is negatively correlated, it has been considered a distribution that focuses on lower values.

ADJUSTED DISTRIBUTIONS



Since Solv2 is positively correlated, it has been considered a distribution that focuses on higher values.



CASE STUDY 2

- Let's now estimate Assay based on the already established relationship, *i.e.*:

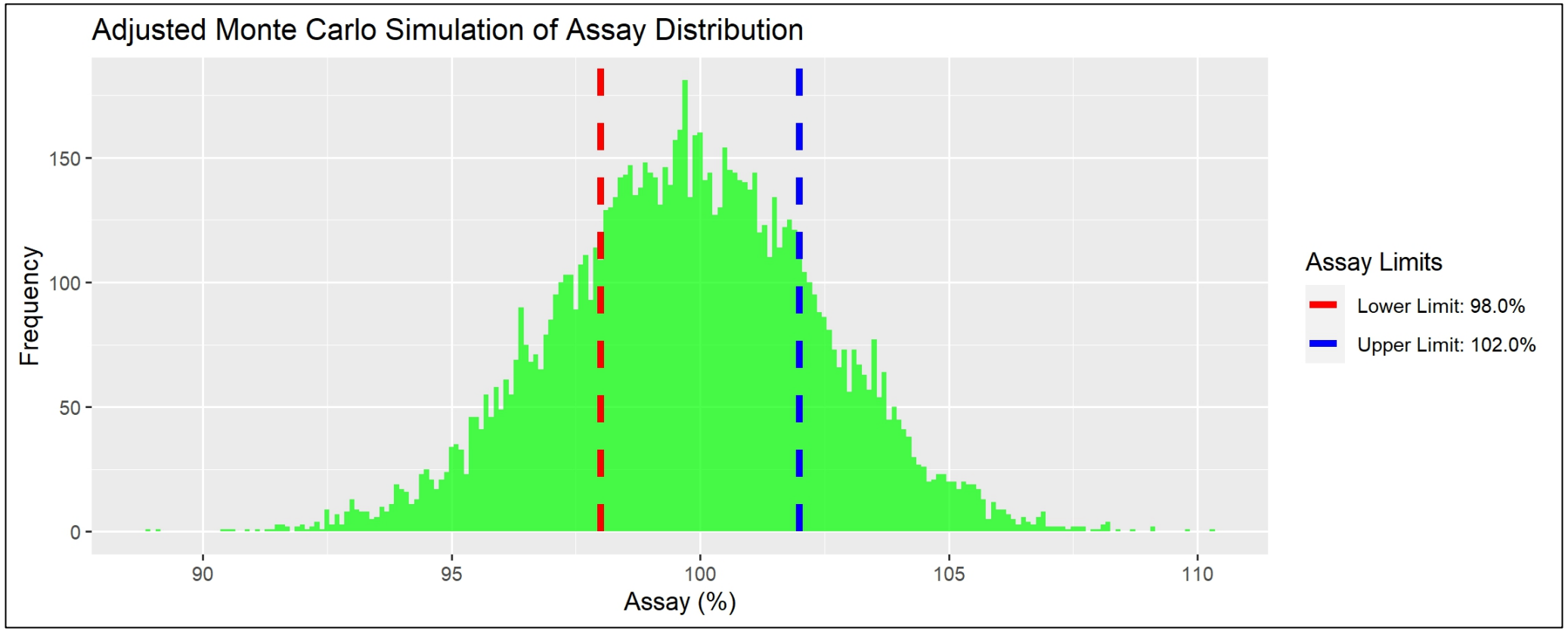
  Assay = 101.255 – (4.647 × LOD) – (22.58 × Total Impurities) – (210.4 × Solv1) + (125.8 × Solv2) + (171.0 × Solv3)

- But using what follows :

	Spec. limits	Distribution
LOD	≤ 0.5%	~ *N* (0.2, 0.05)
Total Impurities	≤ 0.5%	~ *N* (0.1, 0.02)
Solv1	≤ 0.3%	~ *N* (0.08, 0.01)
Solv2	≤ 0.3%	~ *N* (0.12, 0.01)
Solv3	≤ 0.08%	~ *N* (0.02, 0.005)

ADJUSTED DISTRIBUTIONS

CASE STUDY 2



Percentage within limits (Adjusted): 54.49

Percentage outside limits (Adjusted): 45.51

Cp (Adjusted): 0.75

Cpk (Adjusted): 0.65

CASE STUDY 3 :

a micronization process

CASE STUDY 3

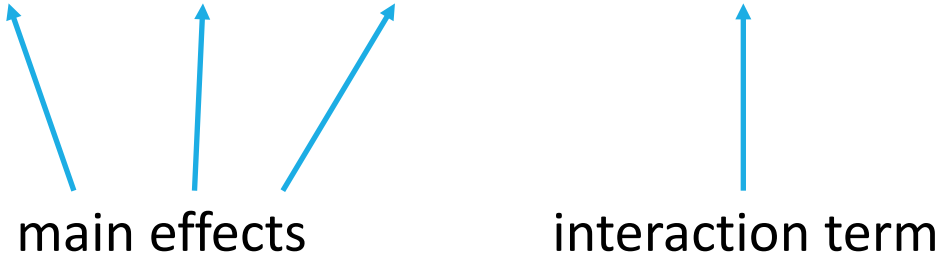
- Let’s consider the micronization: a pharmaceutical process which is crucial for achieving desired particle sizes, particularly for enhancing the dissolution rates and bioavailability of APIs .

- Many factors can influence the particle size during the micronization process in a jet mill, *e.g.*:
  - Feed Rate,
  - Air Pressure
  - Temperature and Humidity
  - Initial Particle Size Distribution and Hardness
  - Nozzle design, *etc*.

CASE STUDY 3

- Suppose we have extracted from pilot studies data of a given API a hypothetical *transfer function* that relates the particle size Y to the input variables feed rate X_1, air pressure X_2 and relative humidity X_3 as follows:

$$Y = 50 - 1.5\,X_1 - 0.5\,X_2 + 0.3\,X_3 + 0.005X_1X_2 + \epsilon$$



CASE STUDY 3

Explanation of Terms:

- X_1 : Feed rate (*e.g.*, in kg/h)
- X_2 : Air pressure (*e.g.*, in psi)
- X_3 : Relative Humidity (*e.g.*, in %)
- Y : Particle size (*e.g.*, in microns)

CASE STUDY 3

- The main feature of a *Monte Carlo simulation* is that it uses random samples (in this case the input variables X_1, X_2 and X_3) to explore the outcomes of a *process* or model (in this case, the particle size represented by the Y variable).

- This returns a distribution of *Y*, which will allow us to assess the impact of variability in the input parameters on the particle size.

- Using *optimization techniques,* we can then find a combination of input variables that would optimize (in this case maximize) the output variable (Y).

CASE STUDY 3

- Let's start with the relationship that correlates the output (particle size) to the input variables (Feed rate, Air pressure, Humidity, *etc.*):

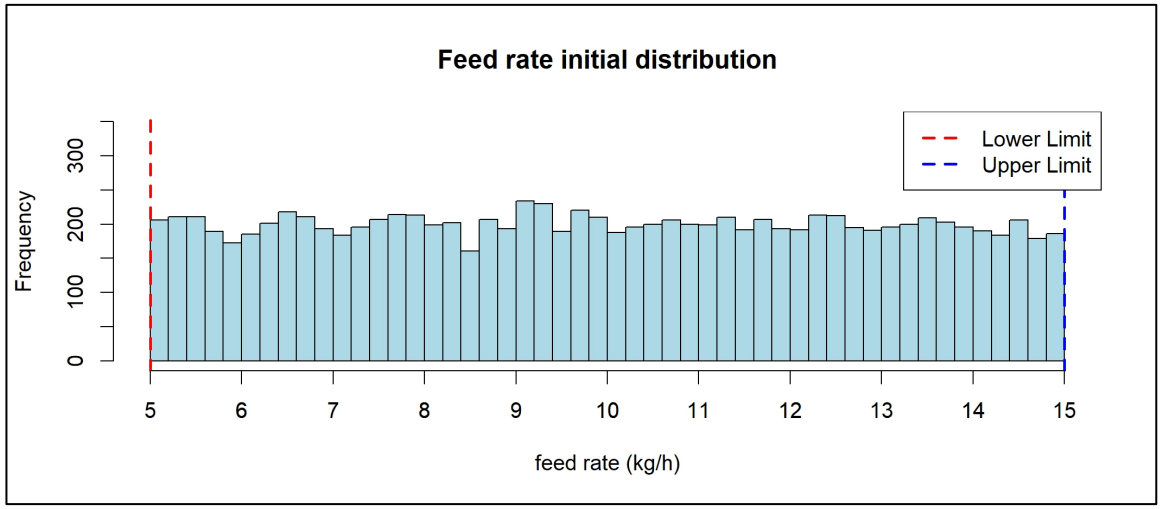
$$Y = 50 - 1.5\,X_1 - 0.5\,X_2 + 0.1\,X_3 + 0.005X_1X_2 + \epsilon$$

- and assume that:

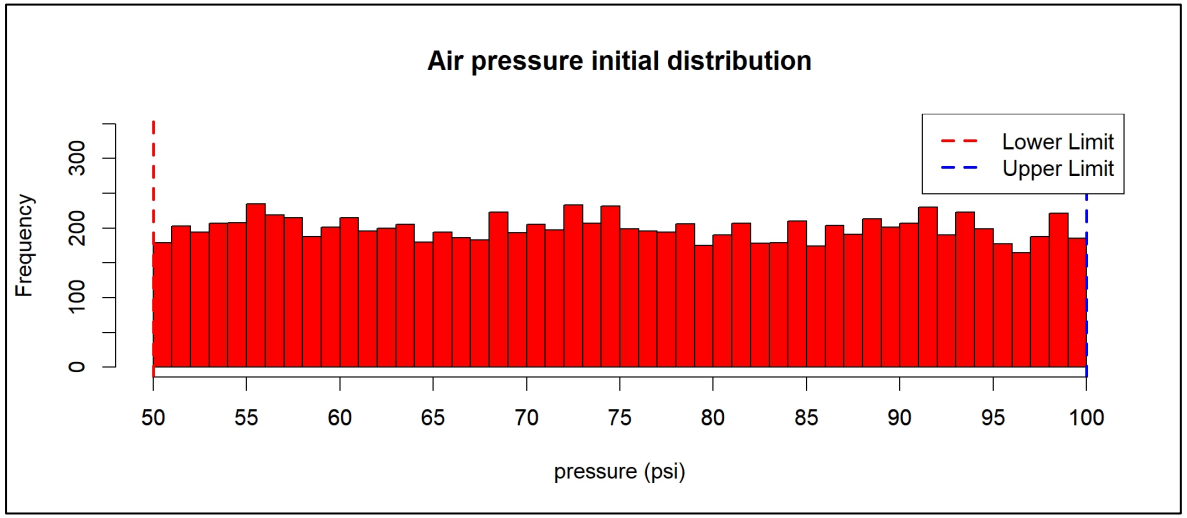
		Spec. Limits	Distribution
X_1	Feed rate	5-15 kg/h	~ *Unif* (5, 15)
X_2	Air pressure (psi)	50 – 100 psi	~ *Unif* (50, 100)
X_3	Relative Humidity (%)	20 – 80 %	~ *Unif* (20, 80)
ϵ	Random error term		~ *N* (0, 5)

INITIAL DISTRIBUTIONS

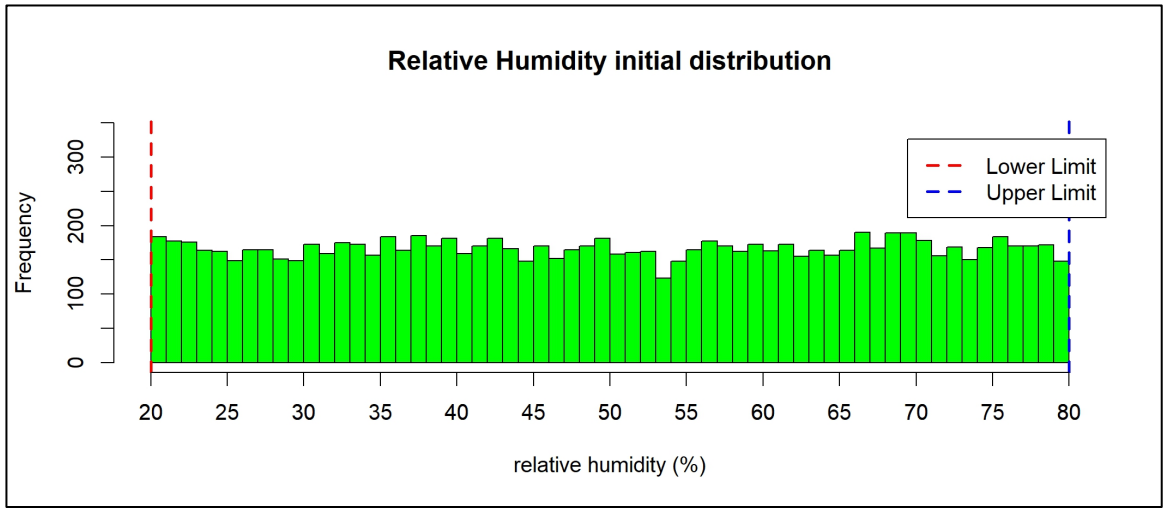
CASE STUDY 3



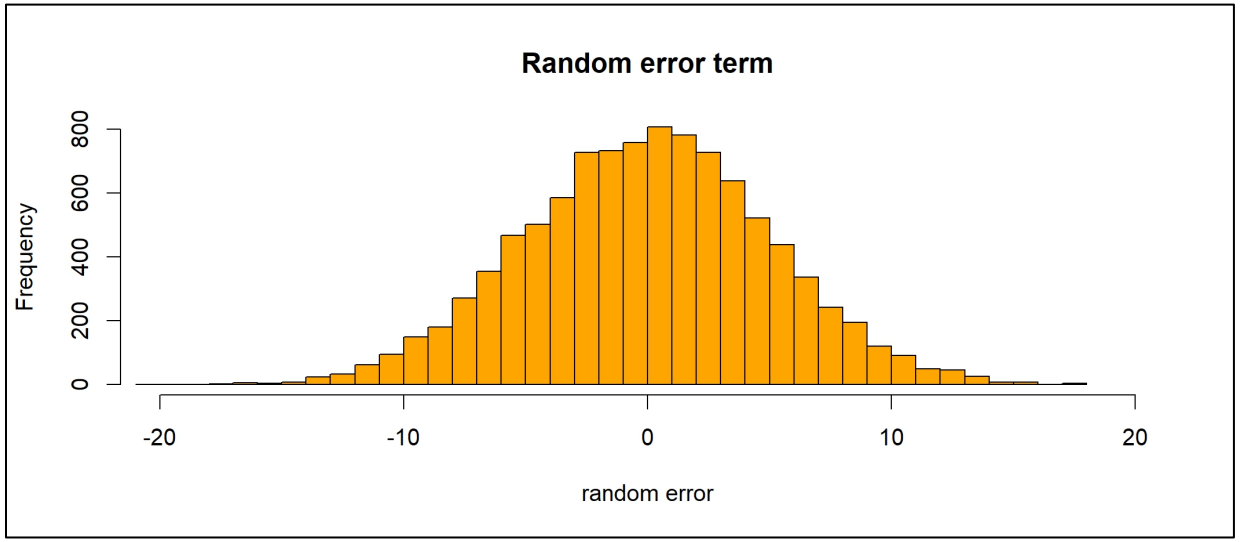
INITIAL DISTRIBUTIONS



CASE STUDY 3

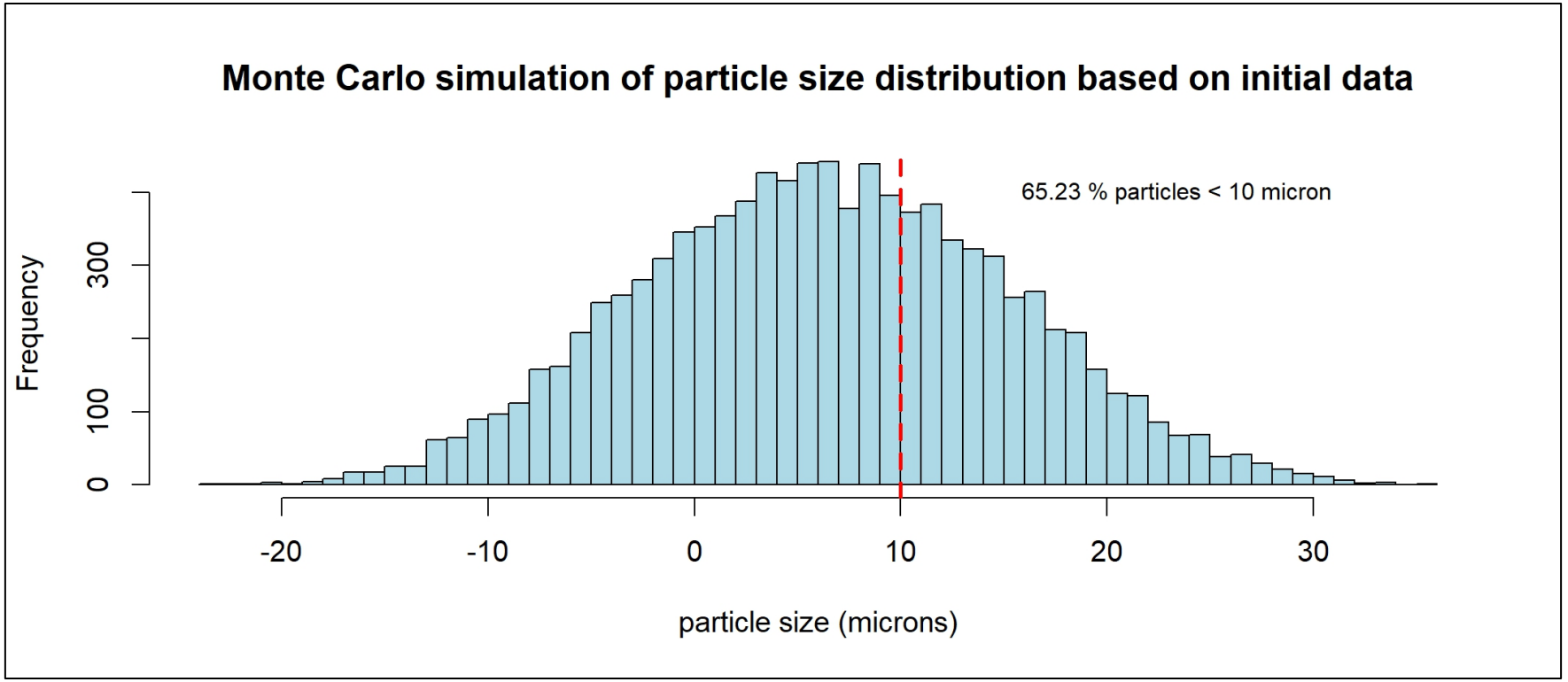


INITIAL DISTRIBUTIONS



CASE STUDY 3

- Conducting the Monte Carlo simulation based on the transfer function seen before and on the distributions of the input variables (*X1*, *X2*, *X3*) just illustrated, we get:



CASE STUDY 3

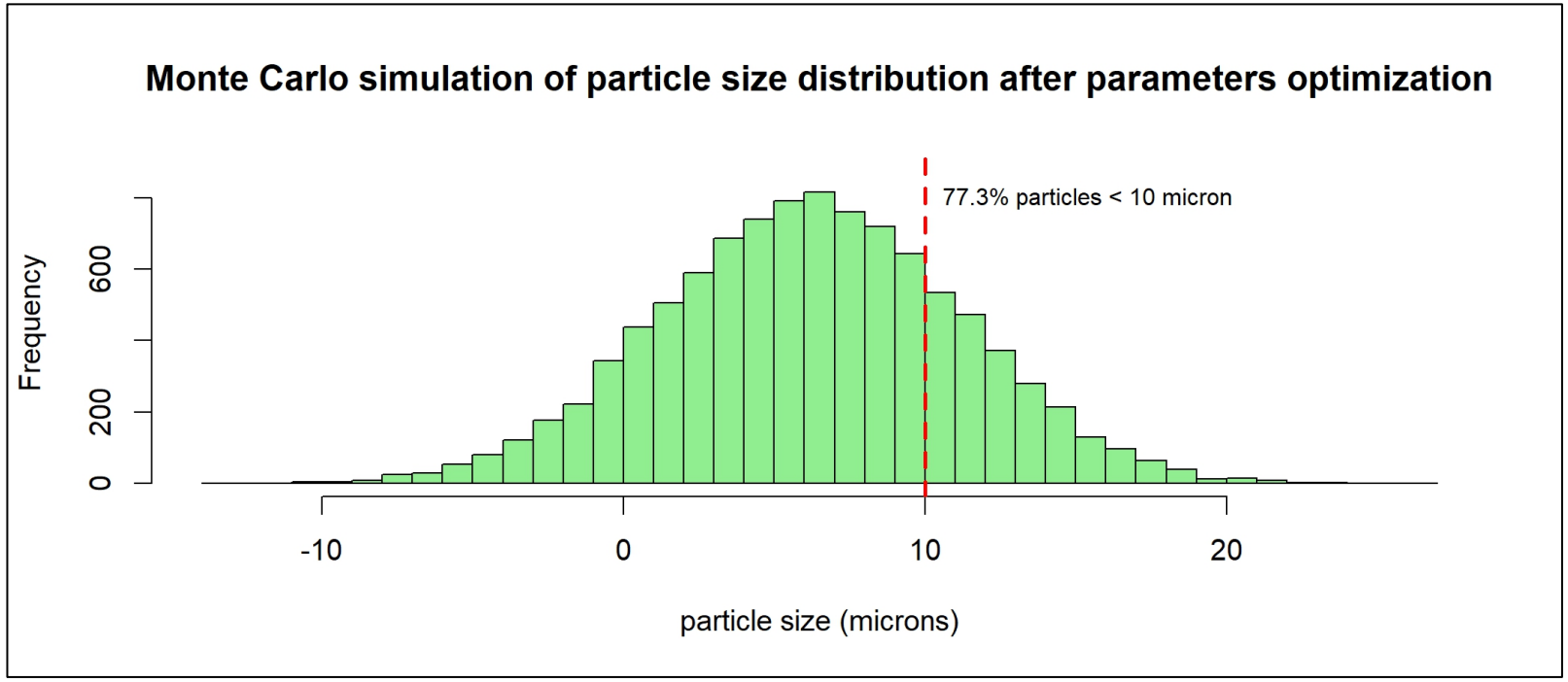
At this point:

- an *objective function* is defined that calculates the percentage of particles below 10 microns given a set of values for *X1*, *X2* and *X3*.
- an *optimization* is then conducted that considers finding the values of *X1*, *X2* and *X3* which minimize the *objective function* (and therefore maximize the percentage of particles under 10 microns) within the specified limits.
- this leads to the following optimized parameters:

$$X1 = 10.0 \qquad X2 = 74.8 \qquad X3 = 50.0$$

CASE STUDY 3

- A Monte Carlo simulation based on the transfer function and the optimized parameters leads to the following:



Particles < 10 microns	Before optimization	After optimization
	65.32%	77.3%

CASE STUDY 3

- It is clear that, as a result of parameters optimization (conducted using the *optimx* R package – version 2023-10.21):
  - the center position of the histogram has moved to a slightly lower value (~ 6.2 μm) than before the optimization (~ 6.4 μm) and
  - the percentage of particles smaller than 10 microns increased by more than 10%.

J. C. Nash, R. Varadhan, *Unifying Optimization Algorithms to Aid Software System Users: optimx for R*, Journal of Statistical Software, Vol. 43, Issue 9 (2011) pp. 1–14

CASE STUDY 4 :

analytical method robustness

CASE STUDY 4

- Suppose that a QC laboratory wants to evaluate a given HPLC method for measuring the concentration of an active pharmaceutical ingredient (API) in a drug product.

- Monte Carlo simulations can help in studying the analytical method by accounting for variability in *column temperature* and *flow rate*, which can affect the measurement accuracy.

  In other words, the Monte Carlo simulations can “model” the variability in measurements due to factors like *column temperature* and *flow rate variability*.

CASE STUDY 4

Suppose that we have:

- Column temperature: Normally distributed around 30°C with a standard deviation of 1°C.
- Flow rate: Normally distributed around 1 mL/min with a standard deviation of 0.05 mL/min.

CASE STUDY 4

- Each simulation run calculates the API concentration, and the output is a distribution of measured concentrations for a nominally 100 mg/mL solution.
- The results could “validate” the method by demonstrating its robustness across the simulated conditions, with 95% of the results within ±2% of the nominal concentration.

CASE STUDY 4

- Suppose the measured concentration C is nominally 100 mg/mL but varies with temperature and flow rate due to their effects on the separation efficiency and detection sensitivity according to the following hypothetical *transfer function*:

$$C = 100 + a \times \Delta T + b \times \Delta F$$

where:

- *C* is the measured API concentration in mg/mL
- ΔT is the deviation of the column temperature from its nominal value (30°C)
- ΔF is the deviation of the flow rate from its nominal value (1 mL/min)

CASE STUDY 4

- *a* and *b* are coefficients representing the sensitivity of the measured concentration to changes in temperature and flow rate, respectively.

  For the purposes of this example, let's assume:

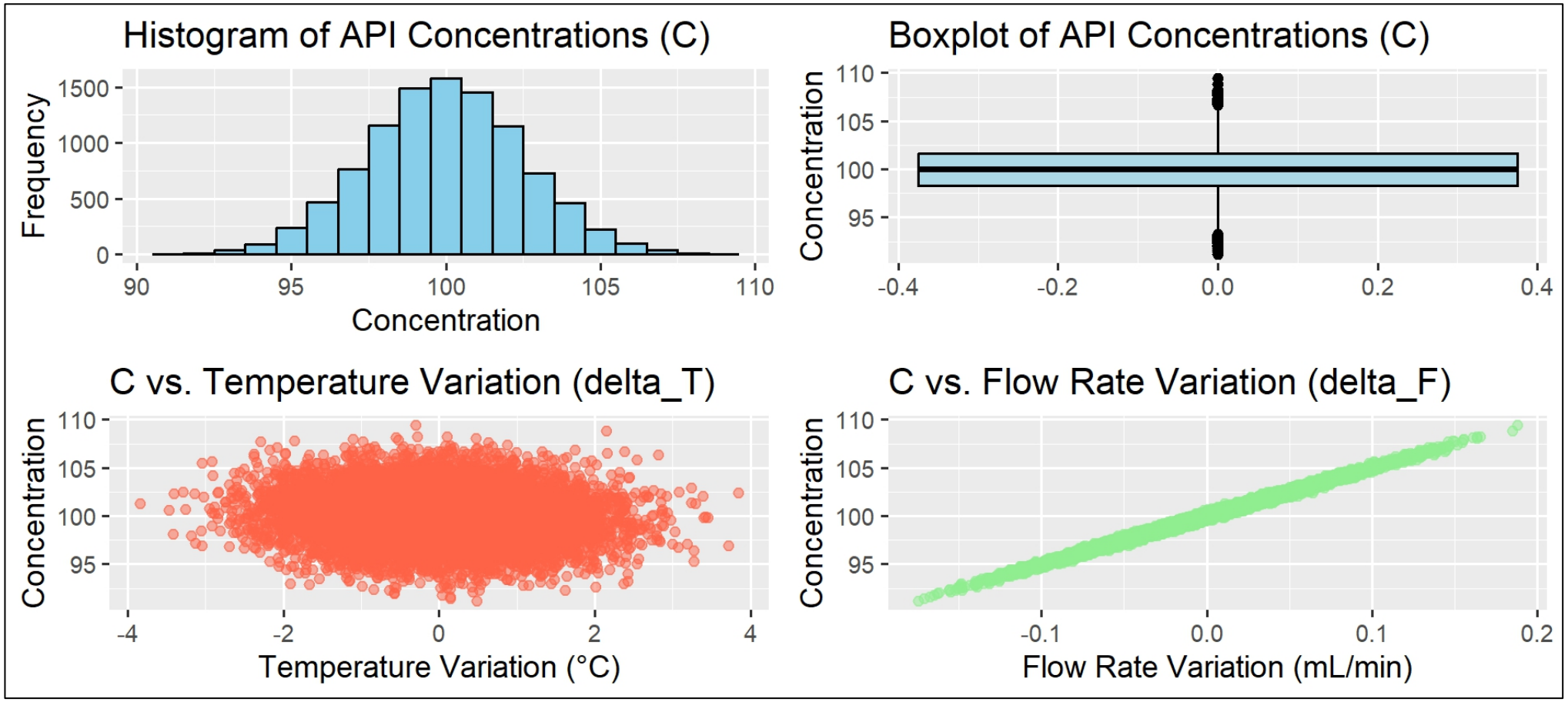
  a= −0.2 mg/mL/°C

  b= 50 mg/mL/(mL/min)

  to illustrate the concept.

Conducting a Monte Carlo simulation in R based on previous transfer function and operative conditions leads to:

CASE STUDY 4



Percentage of results within ±2% of the nominal concentration: **57.68 %**

CASE STUDY 4

What does this simulation tell us and why is it useful ?

- **API Concentrations (C) vs. Temperature Variation (ΔT)**: This scatterplot shows how the changes in temperature from the nominal value (30°C) affect the concentration of the API measured. Given the breadth of the spread in the concentration measurements as temperature varies, which seems relatively wide, it might suggest some degree of variability introduced by temperature changes, even if it's not a direct linear or clear relationship.

CASE STUDY 4

What does this simulation tell us and why is it useful ?

- **API Concentrations (C) vs. Flow Rate Variation (ΔF):** This scatter plot illustrates how the concentrations change with deviations in flow rate from the nominal (1 mL/min). The positive trend line indicates that there's a systematic change in the measured concentration with variations in flow rate. This could be a cause for concern as it suggests that the method's measurements are sensitive to the flow rate and that the flow rate must be precisely controlled.

CASE STUDY 4

What does this simulation tell us and why is it useful ?

- **Variability:** The result indicates that approximately 57.68% of the simulated measurements are within the range of 98 mg/mL to 102 mg/mL (given a nominal concentration of 100 mg/mL). This suggests a low level of precision of the method under the simulated variations of temperature and flow rate.
- **Method Performance:** The fact that a significant portion of the results (42.32%) are outside the ±2% range could point to a need for improvement in the method's performance or more stringent control of the operational parameters.

CASE STUDY 4

What does this simulation tell us and why is it useful ?

- **Robustness Concerns:** A robust analytical method would produce a higher percentage of results within the acceptable range. In this context, the outcome suggests that the method's robustness might be insufficient for routine use without further optimization.

CASE STUDY 5 :
stability analysis

CASE STUDY 5

- The stability of pharmaceutical products under various environmental conditions (temperature, humidity, light, *etc.*) can be predicted using Monte Carlo simulations.

- By simulating the degradation pathways and rates under different conditions, companies can better design stability studies and predict shelf life, ensuring that the product meets quality standards over its intended lifespan.

CASE STUDY 5

- The degradation of many pharmaceutical compounds can often be described by first-order kinetics, where the rate of degradation is proportional to the concentration of the drug that remains.

  The rate of degradation can be expressed by the first-order kinetics equation:

$$-\frac{dC}{dt} = kC$$

  where:

  - C is the concentration of the drug
  - k is the rate constant for the degradation process
  - t is time

CASE STUDY 5

The solution to this differential equation gives the concentration of the drug at any time t:

$$C(t) = C_0 \times e^{-kt}$$

where:

C_0 is the initial concentration of the drug

e is the base of the natural logarithm

CASE STUDY 5

- The rate constant *k* can increase with temperature and humidity.
- The Arrhenius equation provides a way to model the effect of temperature on the rate constant:

$$k = A \times e^{-E_a/RT}$$

where:

- *A* is a pre-exponential factor
- E_a is the activation energy
- *R* is the gas constant
- *T* is the temperature in Kelvin

CASE STUDY 5

For simplicity, let's model the increase in k with humidity (H) linearly, assuming a baseline rate constant k_0 at a specific temperature and humidity:

$$k = k_0 \times \left(1 + \alpha(T - T_0)\right) + \beta(H - H_0)$$

where:

- T_0 and H_0 are reference temperature and humidity conditions
- α and β are coefficients that determine how much k changes with temperature and humidity.

CASE STUDY 5

- The coefficients k_0 , α, and β represent the baseline rate constant and how sensitive the degradation rate is to changes in temperature and humidity, respectively.
- The choices for these values are critical and should ideally be based on experimental data.
- Temperature and humidity can be modeled with normal distributions to reflect natural variability in storage conditions. The means and standard deviations of:
  - mean = 25°C, SD = 5°C for temperature and
  - mean = 60%, SD = 10% for humidity

  are illustrative.

CASE STUDY 5

- A possible criterion for estimating shelf life could be based on whether the drug retains at least 90% of its original potency.

- The use of Monte Carlo simulation to assess the percentage of scenarios where the drug maintains the required potency allows for the consideration of the combined effects of variability in temperature and humidity. This approach provides a probabilistic assessment of shelf life rather than a single deterministic outcome, capturing the uncertainty inherent in real storage conditions.

CASE STUDY 5

- These choices allow us to build a simplified, but illustrative, example of how Monte Carlo simulations can be applied in stability studies for pharmaceutical products.

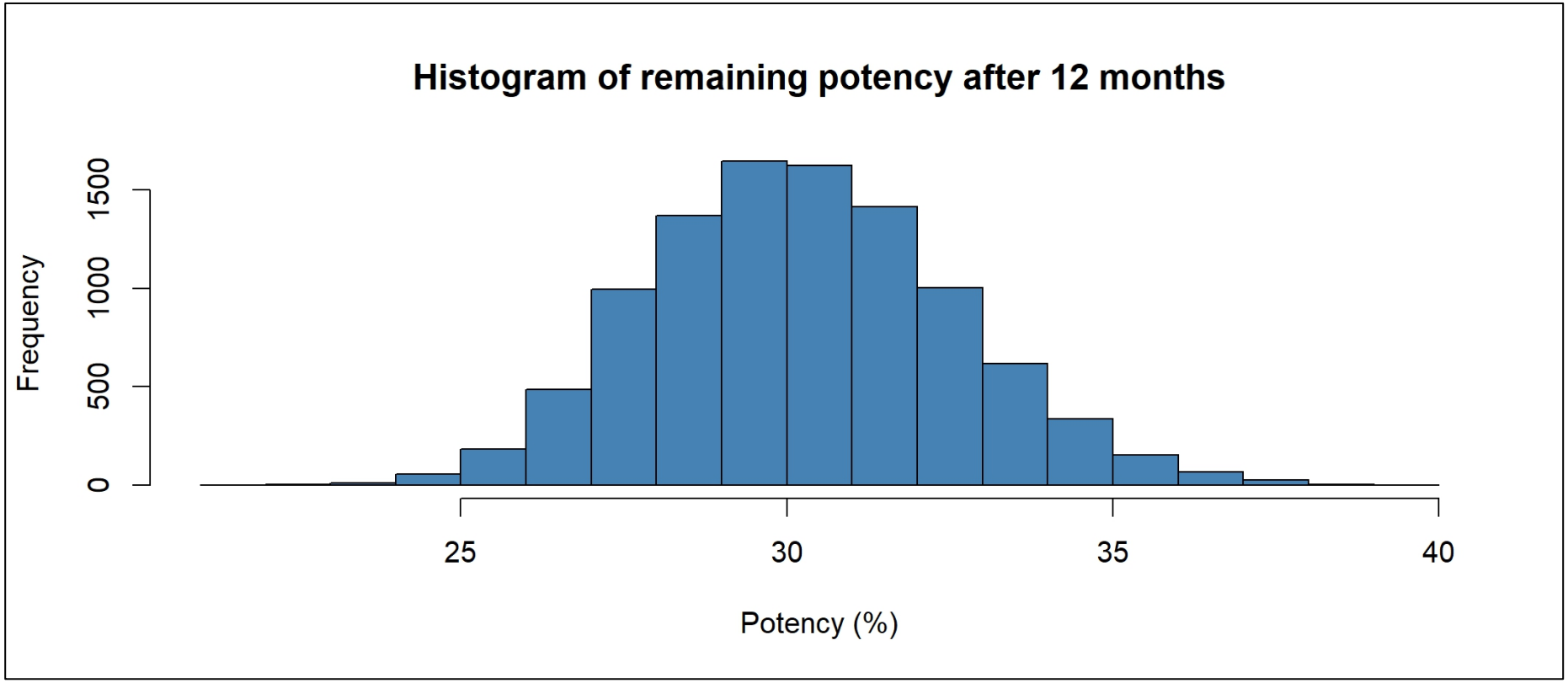
- It is clear that in real applications, the model parameters (k_0,α, β) should be validated against experimental data from accelerated stability studies. This ensures that the model accurately reflects the stability profile of the drug.
  In general, other adjustments may also be necessary.

CASE STUDY 5

- A Monte Carlo simulation can be run using the following hypothetical values for model parameters:

k_0	0.1	Baseline rate constant per month at reference conditions
α	0.02	Temperature coefficient
β	0.01	Humidity coefficient
T_0	25°C	Reference temperature (°C)
H_0	60%	Reference humidity (%)

CASE STUDY 5



Such a simulation leads to an average remaining potency after 12 months of 30% which is very low or, at least, far below a minimum of 90%.

CASE STUDY 5

This result, which might initially seem discouraging, instead highlights possible areas for improvement of the model such as, for example:

1. **Parameters Optimization:** Ensure that the baseline degradation rate (k_0) and the sensitivity coefficients (α and β) are optimized based on experimental stability studies under controlled conditions. These values should reflect the actual behavior of the drug substance or product.

CASE STUDY 5

2. **Extended Range of Conditions**: Explore a broader range of storage conditions, including lower temperatures or humidity levels that might be more conducive to maintaining drug potency. This could help identify more precise storage recommendations.

3. **Model Complexity**: Introduce more complexity into the model if necessary. For example, consider nonlinear effects of temperature and humidity on the degradation rate or include other factors such as light exposure or packaging type.

CASE STUDY 5

4. **Experimental Validation**: Compare the model predictions with actual experimental data from accelerated stability studies. This step is crucial for validating the model and adjusting it to better reflect real-world conditions.

CONCLUSIONS

CONCLUSIONS

- Monte Carlo method is a powerful tool for *simulating and predicting outcomes in pharmaceutical processes.*
- Its field of applicability is much broader than shown so far and can cover aspects such as:
  - ***Regulatory Compliance and Submission*** (*e.g.*, simulating the impact of variability in critical process parameters on the quality attributes of the final product can) demonstrate control over the manufacturing process and the consistency of the product quality)
  - ***Risk Assessment and Management*** (*e.g.*, evaluation of the risk of impurity contamination during manufacturing)

  and much more.

CONCLUSIONS

- These simulations help understand variability and the risk associated with it, thus facilitating better decision making.
- These methods bridge theoretical statistical concepts with practical applications, enhancing both efficiency and accuracy in the pharmaceutical industry.
- Monte Carlo simulation is a methodology that certainly deserves to be used for the powerful practical impact it has on both drug development processes and their management over time.

REFERENCES

- J.M. Hammersley, D.C. Handscomb, *Monte Carlo Methods*, Chapman and Hall, London (1983)
- C.P. Robert, G. Casella, *Monte Carlo Statistical Methods*, 2nd Ed., Springer (2004)
- C.P. Robert, G. Casella, *Introducing Monte Carlo Methods with R*, Springer (2009)
- D. Kroese, T. Taimre, Z.I. Botev, *Handbook of Monte Carlo Methods*, Wiley (2011)
- R.W. Shonkwiler, F. Mendivil, *Explorations in Monte Carlo Methods*, Springer (2009)
- M.L. Rizzo, *Statistical Computing with R*, 2nd Ed., CRC Press, Boca Raton (2019)
- J.C. Nash, *Nonlinear Parameter Optimization Using R Tools*, 1st Ed., Wiley (2014)
- https://permalink.lanl.gov/object/tr?what=info:lanl-repo/lareport/LA-UR-88-9068
- https://library.lanl.gov/cgi-bin/getfile?00326866.pdf