Lecture Notes: Random Number Generation

Our daily lives are filled with uncertainty, from the simple toss of a coin to the stock market's unpredictable rises and falls. This inherent unpredictability is often captured mathematically through random numbers and random variables. In many realms of science, engineering, and even entertainment, the ability to generate and harness randomness is invaluable. Random processes model phenomena as diverse as subatomic particle interactions, the growth of populations, and the strategies in a poker game.

In a world that's increasingly driven by computers — deterministic machines at their core — generating genuine randomness is challenging. Computers follow predefined instructions to produce predictable outcomes. So, how can such a machine create a random number, or simulate the whimsical dance of a snowflake, or the chaotic behavior of weather systems?

Random Number Generators

The answer lies in algorithms that can mimic randomness, even if they aren't truly random. These algorithms are the heart of Random Number Generators (RNGs). By inputting an initial value or 'seed', they produce sequences of numbers that, for all practical purposes, seem random.

Though it might seem counterintuitive, this 'pseudo-randomness' can be a desirable property. Imagine a scientist running a complex simulation of a galaxy. If something goes awry, it's invaluable to replay that simulation with the exact same sequence of 'random' events to debug and understand the problem. This reproducibility is only possible with pseudo-random numbers, not with truly random ones.

But let's step back a bit. Before delving into how we generate these numbers, it's crucial to understand the nature of randomness and how we represent it mathematically.

Definition 1 (Random number)**.** *A random number is an unpredictable value, generated independently from preceding or succeeding numbers. It lacks any discernible pattern or regularity, making it impossible to deduce without understanding the underlying random generation process. Additionally, a random number should accurately represent true randomness, ensuring an equitable chance for all potential outcomes.*

Moving from individual random numbers, it is essential to understand how we can generate a series of such numbers, which leads us to the concept of a random number generator.

Definition 2 (Random number Generator)**.** *A* ***Random Number Generator*** *(RNG) is an algorithm that produces a sequence of numbers that lacks any pattern, i.e., appears random.*

More formally, an RNG is defined as a function:

$$\begin{aligned} R : & S \to T \\ & (s) \mapsto t \end{aligned}$$

where:

- S *is the seed space, a finite set of initial states. An RNG is typically initialized with a value in* S*, known as the seed.*
- T *is the target space, typically the set of real numbers in the interval [0, 1) or a set of integer values.*
- *The function* R *maps each seed* $s \in S$ *to a target* $t \in T$ *in a manner that appears random.*

RNGs are essential in many areas of computing, including simulation, cryptography, and probabilistic algorithms. While the outputs of an RNG may appear random, they are determined entirely by the initial seed and are thus *pseudorandom*.

To get closer to "true" randomness in computer systems, one approach is to use some fundamentally unpredictable process as a source of randomness. These are known as hardware (or true) random number generators (HRNGs or TRNGs).

For example, they might use physical processes like atmospheric noise, radioactive decay, or even small variations in the timing of keyboard presses or mouse movements. These sources are inherently unpredictable and do not follow a deterministic algorithm, so the numbers generated in this way can be considered truly random.

However, HRNGs tend to be slower and more difficult to implement than PRNGs, and in many cases, the numbers generated by PRNGs are sufficiently random for the task at hand.

Property 1. *With the understanding of what random numbers and their generators are, we can now lay down some properties that a well-functioning random number generator should exhibit:*

- ***Unpredictability:*** *Without knowing the algorithm and seed, it should be impossible to predict future numbers.*
- ***Reproducibility:*** *Given the same seed, the RNG should produce the same sequence of numbers.*
- ***Representation of True Randomness:*** *The RNG should accurately represent true randomness, ensuring an equitable chance for all potential outcomes.*
- ***Long period:*** *The sequence of numbers should be long before repeating.*
- ***Efficiency:*** *The RNG should generate numbers quickly.*

It's important to note that not all random number generators will have all these properties. For instance, cryptographic random number generators prioritize unpredictability and may sacrifice reproducibility. The appropriate RNG for a given application depends on what properties are most important for that use case.

The Linear Congruential Generator (LCG) is a type of pseudorandom number generator, and it is one of the oldest and best-known pseudorandom number generator algorithms. The simplicity of its underlying mathematical structure, combined with its fast execution and the minimal memory it requires, have contributed to its widespread usage.

Definition 3 (Linear Congruential Generator)**.** *The LCG generates a sequence of random numbers via the following linear recurrence relation:*

$$X_{n+1} = (aX_n + c) \mod m \tag{1}$$

where:

- X_{n+1} *is the next number in the sequence.*
- X_n *is the current number.*
- a*,* c*, and* m *are constants, known as the multiplier, increment, and modulus, respectively.*
- mod *denotes the modulus operation.*
- *The initial or seed value* $X_0 = S$*, is also required to start the sequence.*

The LCG is designed to generate a sequence of numbers that appear random but are deterministically produced by the recurrence relation. This deterministic production makes the sequence reproducible, an essential property in many applications.

The key idea behind the LCG is the modulus operation, which allows the generator to produce a sequence of numbers in a specific range (0 to $m-1$), regardless of the values of a, c, and X_n.

The parameters a, c, and m can be carefully chosen to produce sequences with desirable properties. For example, with the right parameters, the LCG can achieve a long period (up to m) before repeating, which is another important characteristic for a good pseudorandom number generator.

The Linear Congruential Generator (LCG) has several key characteristics that shape its suitability as a random number generator.

Beginning with the simplest properties, the LCG is notably reproducible and efficient. Reproducibility is a crucial characteristic in many applications, such as simulations, where repeating the same sequence of numbers is vital for replicating results. With an LCG, one can always expect the same sequence of numbers when provided with the same seed and constants.

When it comes to efficiency, the LCG shines as well. The generation process involves merely multiplication, addition, and modulus operations, all of which are computationally inexpensive. The minimalistic requirement of state space, which is just the last generated number, further enhances this efficiency. This makes LCGs an ideal choice for systems burdened by limited computational resources or memory.

Unpredictability, another essential attribute of a good random number generator, is somewhat of a mixed bag for the LCG. While it's generally challenging to predict the output numbers without knowing the multiplier, increment, modulus, and seed, a person with knowledge of the algorithm and access to a sufficient number of sequential numbers from the sequence can potentially calculate the constants and forecast future numbers. Due to this, LCGs are not recommended for applications where a high level of unpredictability, such as in cryptography, is necessary.

To have an idea of the period, consider LCG with the recurrence relation:

$$X_{n+1} = (3X_n + 5) \mod 8 \tag{2}$$

where the seed value $X_0 = 1$. This LCG generates a sequence of integers between 0 and 7. See figure **??**. For this LCG, the sample space Ω for all possible output is the set $\{0, 1, 2, 3, 4, 5, 6, 7\}$. A period of 8, as in the previous example, is indeed quite small and could introduce noticeable patterns in the generated random numbers. It is clear that the maximum period of the LCG is m.

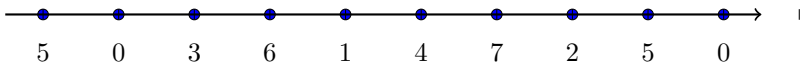

Figure 1: A sequence of realizations from a linear congruential generator (LCG). The x-axis represents the index in the sequence, and the y-coordinate of each point represents the value of the LCG at that index. The sequence is shown until it begins to repeat, including the first and second term of the repetition.

Choosing better values for the initial seed and the parameters of the RNG can lead to a longer period. For instance, the Mersenne Twister PRNG, which is commonly used in programming languages and statistical software, has a period of $2^{19937} - 1$ (a Mersenne prime), which is an astronomically large number.

The final property leads to the main concept of this book, the concept of probability.

Discrete probability distributions

Ensuring an equitable chance for all potential outcomes, as a guiding principle, embodies the essence of fairness and justice in diverse contexts. This idea offers a fundamental framework of equity, resonating with scenarios as simple as a game of chance, and as complex as social systems. Consider a coin toss, for instance. Ideally, the coin is fair, meaning that it offers an equal chance of landing on either heads or tails – a 50% chance for each. Similarly, a dice roll in a fair game should offer each face an equal likelihood of showing up, that is, about 16.67% for each of the six possible outcomes.

Mathematically, the chance of a situation is quantified using a measure called Probability.

Definition 4 (Random Variable)**.** *A random variable is a function that assigns a real number to each outcome of a random experiment. More formally, given an outcome space Ω, a random variable X is defined as a function*

$$X : \Omega \to \mathbb{R} \tag{3}$$

which maps each possible outcome $\omega \in \Omega$ to a real number $X(\omega) = x$. The set of all possible values of X, often denoted by $range(X)$ or simply the image of X, is a subset of the real numbers $\mathbb{R}$.

The sample space, denoted as Ω, for any given experiment signifies the collection of every conceivable outcome of that experiment. If the space Ω is discrete, we will say that X is a discrete R.V.

Definition 5 (Probability Mass Function (pmf))**.** *The probability mass function (pmf) asociated with a Random Variable X is represented as a function $p_X : range(X) \to [0, 1]$, defined by*

$$p_X(x) = P(\{\omega \in \Omega : X(\omega) = x\})$$

for every $x \in range(X)$, such that:

1. *For every outcome $x \in range(X)$, $0 \leq p_X(x) \leq 1$.*
2. *The sum of the probabilities for all possible outcomes is 1, i.e., $\sum_{x \in range(X)} p_X(x) = 1$.*

Any subset $E \subset \Omega$ is considered a situation or event. The probability of an event E, denoted as $P_X(E)$, represents the likelihood of that event occurring. It can also be denoted simply as $P(E)$ when the context is clear.

Let's consider the rolling dice process. Consider the scenario of rolling a fair six-sided dice. The sample space of this scenario is $\{1, 2, 3, 4, 5, 6\}$. Since the dice is fair, each outcome is equally likely. Therefore, the probability assigned to each outcome is $\frac{1}{6}$. The PMF of this discrete random variable is thus:

$$p(x) = \begin{cases} \frac{1}{6} & \text{if } x \in \{1, 2, 3, 4, 5, 6\}, \\ 0 & \text{otherwise} \end{cases}$$

The PMF here characterizes a uniform distribution, each outcome from the dice roll has an equal chance of occurring, demonstrating the fairness of the dice.

Experiment 1 (Probability and frequency)**.** *Consider the Linear Congruential Generator (LCG), a type of RNG algorithm, given by the recurrence relation:*

$$X_{n+1} = (1664525 \times X_n + 1013904223) \mod 2^{32}$$

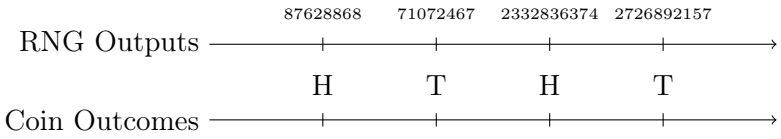

Figure 2: Mapping of LCG Outputs to Coin Toss Outcomes for First Four Iterations

To determine the probability of either outcome (H or T), we consider the frequency approach to probability. Given that the LCG produces uniform random numbers over its range, half of the numbers will be below $\frac{2^{32}}{2}$ *and half will be above. Using this midpoint, numbers below this value can be mapped to 'H' and numbers above to 'T'. Therefore, as* $n \to 2^{32}$*:*

$$P(H) = P(T) = 0.5$$

This ensures an almost equal probability for the LCG's output to represent either a 'Heads' or a 'Tails' for a sufficiently large number of trials.

The empirical validation of this theory can be observed in the figure below. As the sample size N *increases, the proportions of 'Heads' and 'Tails' converge to the theoretical value of* 0.5*, reinforcing the uniformity and reliability of the LCG in mapping outcomes.*

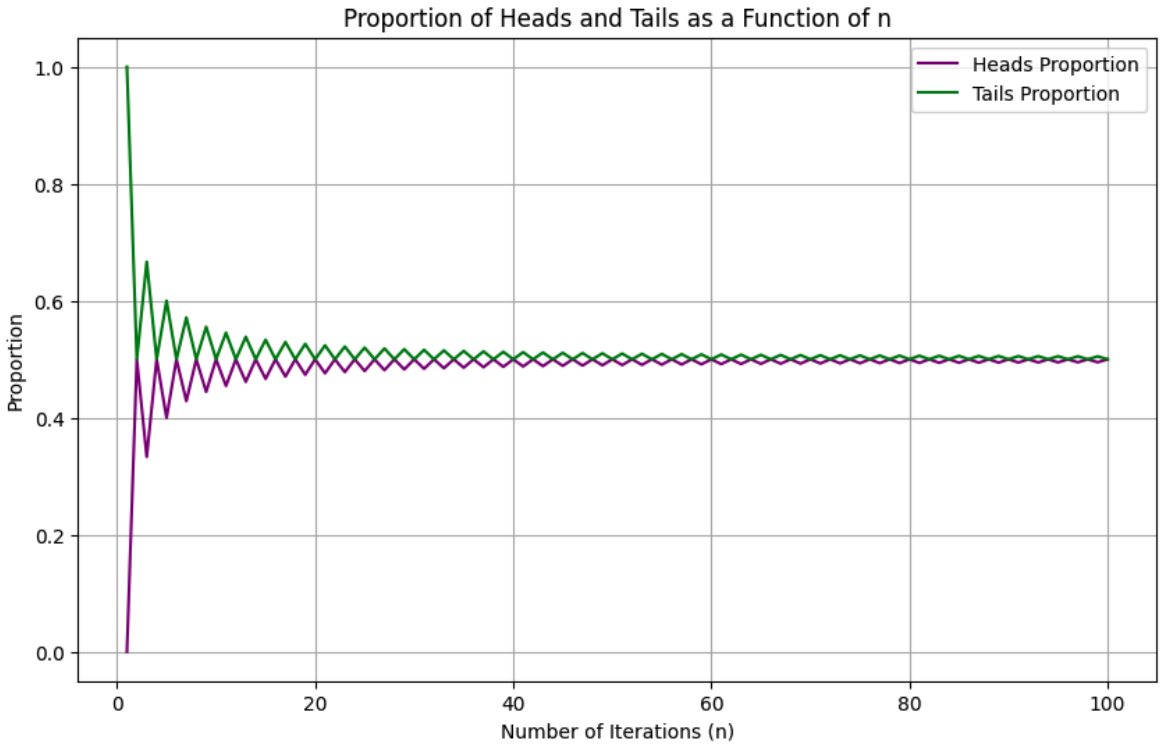

Figure 3: Proportions of 'Heads' and 'Tails' outcomes as generated by the Numerical Recipes LCG over a sample size of N. As N increases, the proportions converge to the expected value of 0.5, demonstrating the uniformity of the RNG in mapping outcomes.

One of the simplest forms of a discrete probability distribution is the discrete uniform distribution. As we just saw, if we let X represent the outcome of flipping a fair coin, X could take the values Heads, Tails each with probability 1/2. Similarly, if X represents the outcome of rolling a fair six-sided die, X can take the values 1,2,3,4,5,6 each with probability 1/6.

Example 1 (Discrete Uniform distribution)**.** *Let* X *be a random variable with outcomes equally likely in the set* $a, a+1, ..., b$ *has the following probability mass function (PMF):*

$$p(X = k) = \frac{1}{b - a + 1}, \quad for \quad k = a, a + 1, ..., b. \tag{4}$$

This, is called the discrete uniform distribution

The preceding examples primarily highlighted uniform distributions, where each outcome had an equal likelihood of occurrence. As we transition further, we will delve into some of the most renowned distributions that exhibit varied probabilities for different outcomes, breaking away from the uniformity principle.

Example 2 (Bernoulli Distribution)**.** *Consider a binary experiment with outcomes that do not share the same probability. The Bernoulli distribution provides a mathematical model for such situations.*

Let X be a random variable representing the outcome of this binary experiment. X is said to follow a Bernoulli distribution if it takes on two possible outcomes: 1 (success) with probability p and 0 (failure) with probability $1 - p$. Formally, the PMF is defined as:

$$P(X = k) = \begin{cases} p & \text{if } k = 1, \\ 1 - p & \text{if } k = 0. \end{cases}$$

Here, $0 \leq p \leq 1$ is the probability of success, and k can only take values 0 or 1.

Example 3 (Binomial Distribution)**.** *Following the idea of the Bernoulli distribution, suppose we conduct a series of n Bernoulli trials. Each trial has two possible outcomes: "success" (with probability p) and "failure" (with probability $1 - p$). The probability of obtaining exactly k successes in these n trials is given by:*

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

for $k = 0, 1, \ldots, n$. The random variable X representing the number of successes in n trials follows a binomial distribution.

Example 4 (Poisson Distribution)**.** *A discrete random variable X is said to have a Poisson distribution with parameter $\lambda > 0$ (where λ is the average number of occurrences in a fixed interval or region) if its PMF is given by:*

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

for $k = 0, 1, 2, \ldots$ where e is the base of the natural logarithm.

Expectation

Imagine you play a game where you roll a fair six-sided dice, and you money equal to the number that shows up on the dice. If you play this game once, you could end up with any amount between 1 CHF and 6 CHF. But if you were to play this game many times, what would be the "average" amount you'd expect to win on each roll? This average value is the expected value.

Definition 6. *For a **discrete** random variable X with probability mass function $p(x)$, the expected value $E[X]$ is defined as:*

$$E[X] = \sum_x x \cdot p(x)$$

where the summation is over all possible values of x.

Essentially, it is the weighted average of the possible values the random variable can take, where the weights are given by their respective probabilities.
For a fair six-sided dice, the expected value would be:

$$E[X] = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) = \frac{21}{6} = 3.5$$

This means that, on average, you would expect to win 3.50 CHF per roll, even though it's impossible to roll a 3.5 on a dice.

Expectation is a key concept in statistics and probability. It plays a central role in decision theory for the evaluation and comparison of strategies. Additionally, it is instrumental in defining other statistical measures such as variance and covariance. By providing a method to summarize the potential outcomes of a random variable into a single value, expectation facilitates predictions based on probabilistic models. Its understanding is essential for advanced applications in statistics and probabilistic analysis.

Property 2 (Linearity of Expectation)**.** *Let $X_1, X_2, \ldots, X_n$ be random variables (not necessarily independent). For any constants $a_1, a_2, \ldots, a_n$, the expectation of their linear combination is given by:*

$$\mathbb{E}\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i \mathbb{E}[X_i]$$

Linearity of expectation, as described above, plays a critical role in many probabilistic analyses, enabling us to break down complex expressions into simpler components. It also serves as foundational groundwork for understanding more intricate theorems, such as the Law of Large Numbers.

Example 5 (Expectation of the Bernoulli Distribution)**.** *Let X be a random variable following a Bernoulli distribution with success probability p.*

Using the expectation formula for the Bernoulli distribution, we have:

$$E[X] = 1 \cdot p + 0 \cdot (1 - p) = p$$

Hence, the expected value of X is p.

Example 6 (Expectation of the Poisson Distribution)**.** *Let X be a random variable following a Poisson distribution with parameter λ. Its probability mass function is given by:*

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

To find the expectation $E[X]$, we compute:

$$E[X] = \sum_{k=0}^{\infty} k \cdot \frac{e^{-\lambda}\lambda^k}{k!}$$

This can be split into two parts: the $k = 0$ term and the sum from $k = 1$ to infinity. The $k = 0$ term gives a contribution of 0 to the sum.

So,

$$E[X] = \sum_{k=1}^{\infty} k \cdot \frac{e^{-\lambda}\lambda^k}{k!}$$

Let's consider the sum S defined by:

$$S = \sum_{k=1}^{\infty} \frac{e^{-\lambda}\lambda^k}{(k-1)!}$$

Multiplying and dividing by λ in our expression for $E[X]$, we notice that:

$$E[X] = \lambda \cdot \sum_{k=1}^{\infty} \frac{e^{-\lambda}\lambda^{k-1}}{(k-1)!} = \lambda \cdot S$$

Recognizing the Taylor series expansion for e^x, which is $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$, our sum S becomes the expansion for e^λ, excluding the $k = 0$ term. This gives:

$$S = e^\lambda - 1$$

Thus, $E[X] = \lambda \cdot (e^\lambda - 1) + \lambda$.
But since $e^{-\lambda} \cdot (e^\lambda - 1)$ is just $-e^{-\lambda}$ added to e^λ, and $-e^{-\lambda} + e^\lambda = \lambda$, we find:

$$E[X] = \lambda$$

Hence, the expected value of X is λ.

Independence

The concept of independence is crucial in the study of probability and statistics. When dealing with random variables, independence ensures that the realization of one random variable doesn't give any information about the realization of another.

Definition 7. *Let X and Y be two random variables. They are said to be **independent** if and only if for every pair of events (or situations) E_1 and E_2, we have:*

$$P(X \in E_1, Y \in E_2) = P(X \in E_1) \cdot P(Y \in E_2)$$

This definition ensures that our understanding of randomness is preserved; knowledge about the outcome of one random process shouldn't provide any information about the outcome of another if they are indeed independent.

Example 7 (Independence with RNG-simulated Coin Tosses)**.** *Building on our previous experiment with the LCG and the coin toss simulation, let's now formalize this using random variables.*

Let:

- *X_1 be a random variable representing the outcome of the first RNG run.*
- *X_2 be a random variable representing the outcome of the second RNG run.*

The RNG can produce outcomes corresponding to 'Heads' (H) or 'Tails' (T). For simplicity, let's denote 'Heads' as 1 and 'Tails' as 0. So, the random variables X_1 and X_2 can take values from the set {0, 1}.

Based on our earlier exploration:

$$P(X_1 = 1) = P(H) = 0.5$$

$$P(X_2 = 0) = P(T) = 0.5$$

If X_1 and X_2 are independent, then:

$$P(X_1 = 1, X_2 = 0) = P(X_1 = 1) \cdot P(X_2 = 0)$$

Plugging in our values:

$$P(X_1 = 1, X_2 = 0) = 0.5 \cdot 0.5 = 0.25$$

Thus, if our simulation yields a joint probability of $X_1 = 1$ and $X_2 = 0$ close to 0.25, it reinforces that X_1 and X_2 are independent when simulated using our RNG method.

In practical scenarios, such independence can be vital to ensure that simulations or experiments aren't inadvertently biased by intertwined RNGs.

Experiment 2. *Let's illustrate dependence using a single RNG, specifically the Linear Congruential Generator (LCG). Recall the LCG's formula:*

$$X_{n+1} = (aX_n + c) \mod m$$

where X_n is the current random number, a and c are constants, and m is the modulus.

For this illustration, we will define two situations based on the same Ω:

- $X_1 : \Omega \to \mathbb{R}$ *where* $X_1(\omega) = \frac{\omega}{m}$
- $X_2 : \Omega \to \mathbb{R}$ *where* $X_2(\omega) = 1 - \frac{\omega}{m}$

Consider two situations:

- A = *Situation where* $X_1(\omega) > \frac{1}{2}$
- B = *Situation where* $X_2(\omega) > \frac{1}{2}$

By our definitions:

- $P(A) = \frac{m/2}{m} = \frac{1}{2}$
- $P(B) = P(1 - \frac{\omega}{m} > \frac{1}{2}) = P(\omega < \frac{m}{2}) = \frac{1}{2}$

However, if A occurs, i.e., $X_1(\omega) > \frac{1}{2}$, then $X_2(\omega)$ will necessarily be less than $\frac{1}{2}$. Thus, $P(A \cap B) = 0$.

This implies:

$$P(A \cap B) \neq P(A)P(B)$$

showing that the situations A and B are dependent.

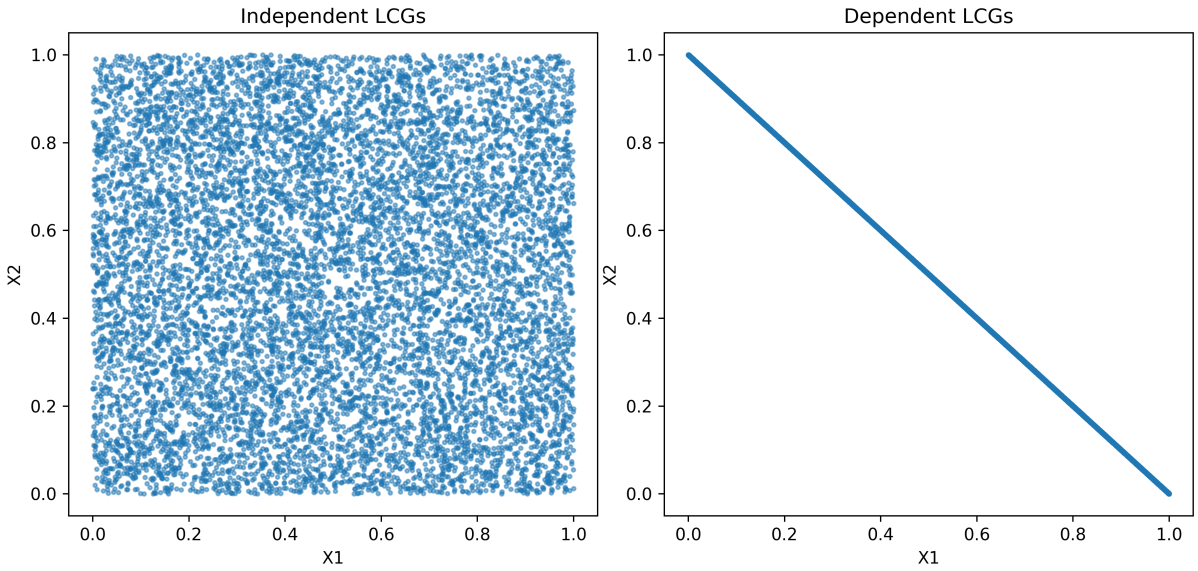

Figure 4: Simulation of dependent and independent LCG.

Definition 8 (Cumulative Distribution Function (Discrete Case))**.** *Let X be a discrete random variable with a probability mass function $p(x)$. The cumulative distribution function (CDF) of X is defined by:*

$$F(x) = P(X \leq x) = \sum_{t \leq x} p(t)$$

where the sum runs over all possible values t of X that are less than or equal to x.

Example 8 (CDF of a Discrete Uniform Distribution)**.** *Consider a random variable X that follows a discrete uniform distribution over the interval $[a, b]$, where a and b are integers with $a \leq b$.*

The Cumulative Distribution Function (CDF) for any value k in this interval is:

$$F(k) = \begin{cases} 0 & \text{if } k < a \\ \frac{k-a+1}{b-a+1} & \text{if } a \leq k \leq b \\ 1 & \text{if } k > b \end{cases}$$

For values in the interval $[a, b]$, the CDF $F(k)$ represents the probability that X takes on a value less than or equal to k. It increases by $\frac{1}{b-a+1}$ for each integer increment in k within the interval.