Lecture 02

FOUNDATIONS OF RANDOMNESS (PART 2)

Probability & Statistics

Francisco Richter, Martina Boschi and Ernst Wit

1 INTRODUCTION

In this lecture, we will introduce several fundamental concepts in probability theory, beginning with **independence** and **expectation**. These two concepts are essential for understanding the **Law of Large Numbers**, a crucial element in our simulation-based approach. We will then explore the **Cumulative Distribution Function** and **Inverse Transform Sampling** for discrete distributions, which offer methods for simulating discrete random variables based on their probability distributions. Together, these concepts provide a strong foundation for simulation-based approach.

2 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 1 *Let X and Y be two random variables. They are said to be **independent** if and only if:*

$$P(u|X(u) = k \cap Y(u) = h) = P(u|X(u) = k) \times P(u|Y(u) = h)$$

where u is a random draw from a random number generator in $[0, 1]$. 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 1 (Independence with RNG-simulated Coin Tosses)

- *Let X represent the outcome of the first coin toss.*
- *Let Y represent the outcome of the second coin toss.*

The RNG produces outcomes between 0 and 1. For each random variable, we can map half of these outcomes to "Heads" and the other half to "Tails." We assign the value 1 to "Heads" and 0 to "Tails." Thus, the random variables X and Y can take values from the set {0, 1}.
From our earlier analysis, we know:

$$P(X = 1) = P(u \,|\, X(u) = 1) = P(Y = 1) = P(u \,|\, Y(u) = 1) = 0.5$$

$$P(X = 0) = P(u \,|\, X(u) = 0) = P(Y = 0) = P(u \,|\, Y(u) = 0) = 0.5$$

If X and Y are independent, then:

$$P(X = 1, Y = 0) = P(X = 1) \cdot P(Y = 0)$$

Substituting the known probabilities:

$$P(X = 1, Y = 0) = 0.5 \cdot 0.5 = 0.25$$

To simulate these two independent variables, we need to ensure that, among the values u which map X to 1, half of them map Y to 0 and half Y to 1. Similarly, for the values of u which map X to 0, half should map Y to 0 and the other half to 1. This will ensure that the joint probability $P(X = 1, Y = 0)$ approaches 0.25, reinforcing that X and Y are independent when simulated using the RNG method.

3 EXPECTATION AND LAWS

Imagine you play a game where you roll a fair six-sided dice, and you win 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 2 *For a discrete random variable X with probability mass function $p_X(x)$, the expected value $E[X]$ is defined as:*

$$E[X] = \sum_{x \in R(X)} x \cdot p_X(x)$$

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. This is actually an important point: the expectation of a RV is not required to belong to the range of that variable.

Property 1 (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:*

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

Example 2 (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 3 (Expectation of the Poisson Distribution) *Let X be a random variable following a Poisson distribution with parameter λ. 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.

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

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

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

By replacing $h = k - 1$ and recognizing the Taylor series expansion for e^{λ}, which is $e^{\lambda} = \sum_{h=0}^{\infty} \frac{\lambda^h}{h!}$, $E[X]$ becomes:

$$E[X] = \lambda \cdot e^{-\lambda} \cdot e^{\lambda} = \lambda$$

Hence, the expected value of X coincides with its rate parameter λ.

3.1 LAW OF LARGE NUMBERS

The Law of Large Numbers (LLN) is a fundamental theorem in probability theory that describes the result of performing the same experiment a large number of times. It states that the average of the results obtained from a large number of trials should be close to the expected value and will tend to become closer as more trials are performed.

Theorem 1 (Law of Large Numbers) *Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed (i.i.d.) random variables with expected value $E[X_i] = E[X]$. Then, for any $\epsilon > 0$,*

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - E[X]\right| \geq \epsilon\right) \to 0 \text{ as } n \to \infty.$$

This theorem guarantees that the sample average converges to the expected value as the number of trials increases.

Example 4 (LLN with Dice Rolls) *Consider rolling a fair six-sided dice. Let X_i represent the outcome of the i-th roll. The expected value $E[X]$ is 3.5, as calculated earlier. According to the LLN, as the number of rolls n increases, the sample average $\frac{1}{n}\sum_{i=1}^{n} X_i$ will converge to 3.5.*

4 CUMULATIVE DISTRIBUTION FUNCTION

Definition 3 (Cumulative Distribution Function - Discrete Case) *Let X be a discrete random variable with a probability mass function $p_X(x)$. The cumulative distribution function (cdf) of X is defined by:*

$$F_X(x) = P(X \leq x) = \sum_{k \leq x} p_X(k)$$

Example 5 (Discrete Uniform Cdf) *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 cdf for any value k in this interval is:

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

For values in the interval $[a, b]$, the cdf $F_X(k)$ 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.

4.1 INVERSE TRANSFORM SAMPLING FOR DISCRETE VARIABLES

Up until now, we have been simulating values from a random number generator within the range of 0 and 1. Our aim was to generate random values that reflect the characteristics of true randomness. Essentially, what we have learned so far is how to simulate values from a uniform distribution, by means of the linear congruential generator.
Nevertheless, we already introduced several types of random experiments, such as the Bernoulli random experiment, that we may be interested in simulating in practice. The knowledge of the cdf allows to reach this goal, by means of an adaptation of the inverse transform sampling method for discrete variables.
Inverse transform sampling consists of a technique that allows to simulate random values according to a particular probability distribution by means of the inverse of the corresponding cdf.
We can thus have a look to the recipe for simulating a discrete random variable:

1. Let random variable X have cdf F_X (given).
2. Simulate a random number u between 0 and 1 by means of a random number generator.
3. Then:
$$x = \min\{k | F_X(k) \geq u\}$$
is a random draw from the random variable X.