Statistical Distributions via Recursive Convolution: An Abstract Notation Framework

Abstract

This paper presents a unified abstraction framework for understanding and constructing statistical distributions using a new recursive convolution-based notation. Traditional probability distributions, such as the binomial, geometric, negative binomial, exponential, gamma, and Poisson distributions, are often defined using final algebraic formulas that obscure the underlying processes. Our approach shifts the focus to decomposition through recursive convolution and core rules, allowing for a deeper understanding of their structure and the natural emergence of combinatorial patterns.

We demonstrate how this abstraction can generalize key distributions and potentially extend to new probability models through flexible core rules. By consistently applying recursive convolution, we show that the factorial terms, binomial coefficients, and exponential decay functions emerge organically from the construction process.

1. Introduction

Statistical distributions play a central role in probability theory, but their traditional presentation often relies on polished formulas, such as:

$$P(X = k) = \binom{n}{k} p^k q^{n-k} \quad \text{(binomial distribution)}$$

These formulas are valuable for computation but obscure the underlying structure of the distributions and how they arise. In this paper, we propose a recursive convolution-based abstraction framework that provides a conceptual understanding of how distributions are built from atomic components. Our approach abstracts the construction process through new notation that highlights recursive accumulation and core rules.

The primary objective of this framework is to show that key statistical distributions, both discrete and continuous, can be derived and understood through recursive convolution layers without explicit reliance on combinatorial or algebraic proofs. Additionally, the flexible core rules in this framework open the door to exploring new and hybrid distributions.

2. Core Building Blocks

2.1 Probability Atoms

We define two primitive probability atoms:

- **Success atom** p: Represents the probability of a success event.
- **Complement atom** $q = 1 - p$: Represents the probability of a failure or complement event.

These atoms form the building blocks for discrete distributions such as the binomial and geometric distributions. For continuous distributions (e.g., exponential and gamma), we generalize these to infinitesimal survival probabilities.

2.2 Recursive Convolution

The key operation in our framework is recursive convolution, denoted as:

$$\mathrm{conv}(c_1, q(i))^k$$

This operation recursively combines an initial core c_1 with flowing probabilities $q(i)$ across layers to build the desired distribution. Each layer represents an incremental step (discrete or continuous), and core rules govern how the recursion progresses.

2.3 Core Rules

Core rules determine when and how to switch between probability atoms during the recursive process. For example, in the geometric distribution, the rule specifies that failures (q) occur until the first success (p), while in the gamma distribution, the rule specifies that survival probabilities accumulate until k events have occurred.

3. New Notation for Distributions

We formalize statistical distributions using the notation:

$$\text{Distribution} = \text{conv}(c_1, q(i))^k \quad \text{where core rules apply.}$$

This notation builds distributions recursively by convolving an initial core c_1 with a flowing probability $q(i)$ across k steps, where the core rule defines how and when $q(i)$ or c_1 is applied.

---

4. Applications to Common Distributions

4.1 Binomial Distribution

The binomial distribution models the probability of k successes in n independent trials, each with success probability p.

Notation:

$$\text{Binomial}(n, p) = \text{conv}(c_1)^n \quad \text{where } c_1 = \begin{cases} p & \text{success} \\ q & \text{failure} \end{cases}$$

The recursive convolution naturally accumulates the combinations of successes and failures, and the binomial coefficient $\binom{n}{k}$ emerges from the structure.

4.2 Geometric Distribution

The geometric distribution models the probability of waiting for the first success after a sequence of failures.

Notation:

$$\text{Geometric}(p) = \text{RCDF}\left(\text{conv}(c_1, c_2)^n \text{ where } c_1 = q, \ c_2 = \begin{cases} q & i < n \\ p & i = n \end{cases}\right)$$

The core rule specifies that failures q occur until the final success p at the n-th step.

4.3 Negative Binomial Distribution

The negative binomial distribution generalizes the geometric distribution by modeling the probability of k successes in a sequence of trials.

Notation:

$$\text{NegativeBinomial}(k, p) = \text{RCDF}\left(\text{conv}(\text{conv}(c_1, c_2)^n)^k\right)$$

The recursive convolution layers handle both the sequence of failures before each success and the accumulation of multiple successes.

4.4 Exponential Distribution

The exponential distribution models the waiting time for the first event in a Poisson process.

Notation:

$$\text{Exponential}(\lambda) = \text{RCDF}\left(\text{conv}(c_1, q(dt))^t \text{ where } c_1 = \lambda, \ q(dt) = 1 - \lambda \cdot dt\right)$$

The recursive convolution over infinitesimal time increments dt naturally produces the exponential decay $e^{-\lambda t}$.

4.5 Gamma Distribution (Corrected as Nested Convolution)

The gamma distribution generalizes the exponential distribution by modeling the waiting time for k events in a Poisson process. Instead of directly accumulating waiting times, we convolve waiting intervals for k events recursively.

Notation:

$$\text{Gamma}(k, \lambda) = \text{RCDF}\left(\text{conv}(\text{conv}(c_1, q(dt))^t)^k \text{ where } c_1 = \lambda, \ q(dt) = 1 - \lambda \cdot dt\right)$$

- The inner convolution $\text{conv}(c_1, q(dt))^t$ models the waiting time for a single event (like the exponential distribution).
- The outer convolution $\text{conv}(\cdot)^k$ recursively accumulates waiting times for k events.

4.6 Poisson Distribution

The Poisson distribution models the number of events occurring in a fixed interval given an average rate λ.

Notation:

$$\text{Poisson}(\lambda) = \text{conv}(c_1, q(i))^k \text{ where } c_1 = e^{-\lambda}, \ q(i) = \frac{\lambda}{i}$$

The recursive convolution counts events directly without accumulating waiting times, and the factorial term $\frac{1}{k!}$ emerges naturally.

---

5. Generalization and Future Applications

The flexibility of this framework lies in its core rules, which can be modified to generate hybrid or new distributions. By altering when and how success and complement atoms are applied, one can construct distributions that model complex real-world phenomena beyond the standard statistical models.

Possible future directions include:

- **Non-homogeneous Poisson processes** with time-varying rates.
- **Multivariate distributions** constructed from interacting convolution layers.
- **Hybrid models** combining discrete and continuous processes.

---

6. Conclusion

This paper presents a novel abstraction framework for statistical distributions using recursive convolution and core rules. By shifting the focus from static formulas to dynamic processes, we have shown how key distributions naturally emerge through recursive constructions. Our notation is consistent, generalizable, and provides a deeper understanding of the underlying structure of probability distributions.