The functional CLT (Donsker's invariance principle): Proof, Simulations

Giacomo Babudri

January 2024

1 Introduction

In probability theory, Donsker's theorem (also known as Donsker's invariance principle, or the functional central limit theorem), named after Monroe D. Donsker, is a functional extension of the central limit theorem for empirical distribution functions. Specifically, the theorem states that an appropriately centered and scaled version of the empirical distribution function converges to a Gaussian process.

1.1 Importants Notions

Empirical Process

Consider a sequence of independent and identically distributed (i.i.d.) random variables $X_1, X_2, \ldots$ with a common distribution function $F(x)$. Define the empirical distribution function $F_n(x)$ based on the first n observations:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{X_i \leq x\}}$$

The empirical process is given by $G_n(x) = \sqrt{n}(F_n(x) - F(x))$.

Donsker's Invariance Principle

Donsker's theorem states that, under certain conditions, the empirical process $G_n(x)$ converges in distribution to a limiting process called the Brownian bridge.

Brownian Bridge

The Brownian bridge is a Gaussian process with mean zero and covariance function $Cov(B_s, B_t) = s(1-t)$ for $0 \leq s \leq t \leq 1$.

1.2 Proof

1.2.1 Intuition:

The main concept is to cleverly embed a sequence of random variables $X_1, X_2, \ldots, X_n$ in the same "world" as a Brownian motion. This ensures that when we look at the scaled partial sums S_n^*, they closely mimic the behavior of a scaled Brownian motion.

1.2.2 Step-by-Step Explanation:

1. Start with Brownian Motion:

Begin with a standard Brownian motion B_t and introduce stopping times T_n where the motion intersects horizontal integer lines.

2. Define Stopping Times:

Formally define stopping times $T_1 := \inf\{t : |B_t| = 1\}$ and $T_{n+1} := \inf\{t > T_n : |B_t - B_{T_n}| = 1\}$. These stopping times ensure we capture the moments when the Brownian motion hits integer lines.

3. Skorokhod Embedding:

Use Skorokhod's Embedding Theorem, guaranteeing the existence of a stopping time T such that B_T follows the same distribution as a random variable X with mean 0, variance 1, and finite second moment. In simpler terms, X behaves like a "scaled" version of B_T.

4. Recursive Construction:

Build a sequence of stopping times $T_1 < T_2 < \ldots < T_n$ inductively, such that $S_n = B_{T_n}$. This means the Brownian motion with these stopping times has the same distribution as a simple random walk described by S_n.

5. Scaling and Convergence:

Rescale the Brownian motion, and observe that as n approaches infinity, the difference $\sqrt{n}(B_{n/t}/\sqrt{n} - S_n^*)$ becomes negligible. This implies the convergence in distribution of the scaled empirical process to a standard Brownian bridge.

2 Simulation

2.1 Code

```
import numpy as np
import matplotlib.pyplot as plt
```

```
# Number of sequences
num_sequences = 500

# Number of observations in each sequence
sequence_length = 100

# Generate random sequences
sequences = np.random.randn(num_sequences, sequence_length)

# Calculate cumulative sums for each sequence
cumulative_sums = np.cumsum(sequences, axis=1)

# Rescale the cumulative sums according to Donsker's Invariance Principle
scaled_cumulative_sums = cumulative_sums / np.sqrt(sequence_length)

# Plot the rescaled cumulative sums
plt.figure(figsize=(10, 6))
plt.plot(scaled_cumulative_sums.T, color='blue', alpha=0.2)
plt.title("Simulation of Donsker's Invariance Principle")
plt.xlabel("Time")
plt.ylabel("Scaled Cumulative Sums")
plt.show()
```

Output

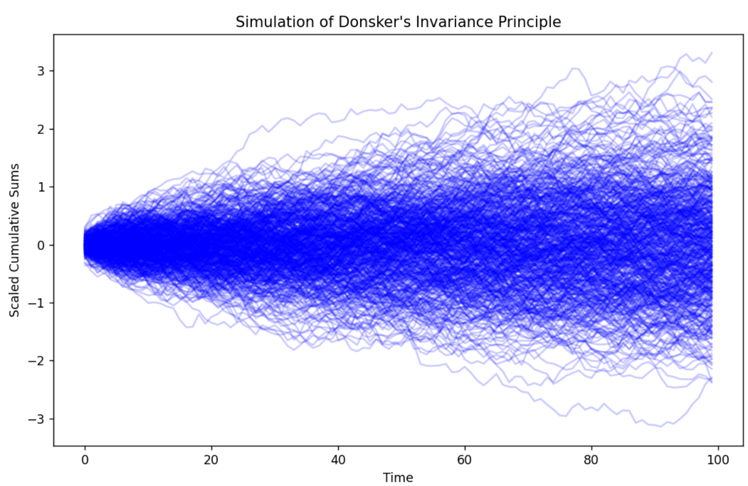

Figure 1: Simulation's output