Empirical Processes: Theory and Application

Zhe Gao

School of Management
University of Science and Technology of China

20, March, 2025

Outline

Overview

- Empirical processes arise naturally in the study of statistics as a way to understand the behavior of sample data relative to the underlying population distribution.
- They are essential in fields that require robust, non-parametric methods where traditional parametric assumptions cannot be satisfactorily met.
- This presentation explores the theoretical foundations of empirical processes, their practical applications, and how they inform modern statistical practice.
- Understanding these concepts is crucial for professionals in data-intensive fields such as data science, biostatistics, and financial analytics.

Basic Concepts - Empirical Distribution Function

- Empirical Distribution Function (EDF): For a sample $X_1, X_2, \ldots, X_n$ from a distribution F, the EDF is defined as follows:

$$\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq t),$$

where I is the indicator function, which equals 1 if the condition inside the parentheses is true, and 0 otherwise.

- EDF is a step function that jumps $1/n$ at each sample point.
- Properties:
  - Right-continuous
  - Converges pointwise to the CDF as $n \to \infty$

Basic Concepts - Glivenko-Cantelli Theorem

- The Glivenko-Cantelli Theorem, a fundamental result in the theory of empirical processes, states that the EDF converges uniformly to the true distribution function as the sample size increases:

$$\sup_{t\in\mathbb{R}} |\hat{F}_n(t) - F(t)| \to 0 \text{ almost surely as } n \to \infty.$$

- This theorem assures us that the empirical distribution function is a good estimator of the true distribution function in a very strong sense.
- The Glivenko–Cantelli classes arise in Vapnik–Chervonenkis theory, with applications to machine learning.

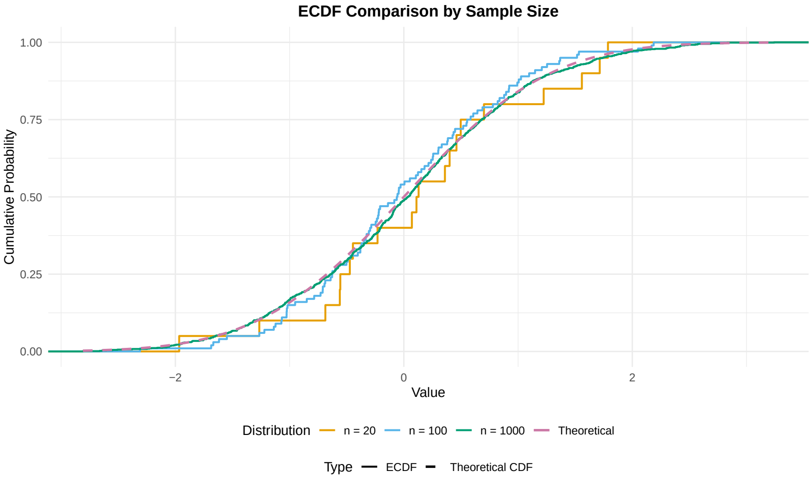
Cumulative Distribution Function

- The CDF F of a random variable X is defined as:

$$F(t) = P(X \le t),$$

- F is right-continuous with left limits and increases monotonically.
- Properties:
  - Bounded: $0 \le F(t) \le 1$
  - Non-decreasing: If $a \le b$, then $F(a) \le F(b)$

Toy Example


ECDF Comparison by Sample Size
Cumulative Probability
1.00
0.75
0.50
0.25
0.00
−2
0
2
Value
Distribution
n = 20
n = 100
n = 1000
Theoretical
Type
ECDF
Theoretical CDF

Empirical process

- The empirical process $\alpha_n(t)$ associated with $\hat{F}_n$ is then given by:

$$\alpha_n(t) = \sqrt{n}(\hat{F}_n(t) - F(t))$$

- This process measures the fluctuation of the EDF around the true distribution F.
- The empirical process provides a mathematical framework for understanding and quantifying how sample data approximates its true distribution. It reveals large sample properties, especially in the context of nonparametric statistics.

Outline

Introduction

- The Glivenko-Cantelli Theorem, also known as the "Fundamental Theorem of Statistics," is crucial for validating the empirical distribution function (EDF) as a consistent estimator of the cumulative distribution function (CDF).
- It guarantees that the EDF converges uniformly to the CDF across all points as the sample size increases indefinitely.

Glivenko-Cantelli Theorem

Theorem 2.1

For i.i.d. real-valued random variables $X_1, X_2, \ldots, X_n$ with distribution function F, we have almost sure convergence:

$$\|F_n - F\|_\infty = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{a.s.} 0 \quad as\ n \to \infty$$

This implies uniform convergence of the EDF to the CDF over the entire real line.

Glivenko–Cantelli class

Definition 2.2

A class $\mathcal{F}$ is called a Glivenko–Cantelli class with respect to a probability measure P if

$$\left\|P_n - P\right\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}}\left|P_n f - Pf\right| \to 0,$$

where $Pf = \int_S f d\mathbb{P}$.

If convergence is:

- Almost surely: Strong GC class;
- In probability: weak GC class.

The GC Theorem is a special case, with $\mathcal{F} = \{I(x \leq t) : t \in \mathbb{R}\}$.

Proof Outline

- Concentration: with probability at least $1 - \exp\left(-2\epsilon^2 n\right)$,

$$\|P - P_n\|_G \le \mathbf{E}\,\|P - P_n\|_G + \epsilon.$$

- Symmetrization: $\mathbf{E}\,\|P - P_n\|_G \le 2\mathbf{E}\,\|R_n\|_G$, where we've defined the Rademacher process $R_n(g) = (1/n)\sum_{i=1}^{n} \epsilon_i g\left(X_i\right)$.
- Restrictions.

Proof - Concentration

- Fix $-\infty = x_0 < x_1 < \cdots < x_{n-1} < x_n = \infty$ such that $F\left(x_j\right) - F\left(x_{j-1}\right) = \frac{1}{n}$ for $j = 1, \ldots, n$. Now for all $x \in \mathbb{R}$ there exists $j \in \{1, \ldots, m\}$ such that $x \in \left[x_{j-1}, x_j\right]$.

$$
\begin{aligned}
F_n(x) - F(x) &\leq F_n\left(x_j\right) - F\left(x_{j-1}\right) = F_n\left(x_j\right) - F\left(x_j\right) + \frac{1}{n} \\
F_n(x) - F(x) &\geq F_n\left(x_{j-1}\right) - F\left(x_j\right) = F_n\left(x_{j-1}\right) - F\left(x_{j-1}\right) - \frac{1}{n}
\end{aligned}
$$

Therefore,

$$
\|F_n - F\|_\infty = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \leq \max_{j \in \{1, \ldots, n\}} \left|F_n\left(x_j\right) - F\left(x_j\right)\right| + \frac{1}{n}
$$

Proof - Concentration

- Let $G = \{I[x \leq t] : t \in \mathbf{R}\}$, then

$$\|F_n - F\|_\infty = \|P - P_n\|_G = \sup_{g \in G} \|Pg - P_n g\|.$$

- The concentration inequality implies that,

$$P(\|F_n - F\|_\infty \leq \mathbf{E}[\|F_n - F\|_\infty] + \epsilon) \leq 1 - \exp\left(-2\epsilon^2 n\right).$$

Proof - Symmetrization

We symmetrize by replacing Pg by $P'_n g = \frac{1}{n}\sum_{i=1}^{n} g\left(X'_i\right)$,

$$\begin{aligned}
\mathbf{E}[\|P-P_n\|_G] &= \mathbf{E}\left[\sup_{g\in G}\left|\mathbf{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left(g\left(X'_i\right)-g\left(X_i\right)\right)\middle| X_1^n\right]\right|\right] \\
&\leq \mathbf{E}\left[\mathbf{E}\left[\sup_{g\in G}\left|\frac{1}{n}\sum_{i=1}^{n}\left(g\left(X'_i\right)-g\left(X_i\right)\right)\right|\middle| X_1^n\right]\right] \\
&= \mathbf{E}\left[\sup_{g\in G}\left|\frac{1}{n}\sum_{i=1}^{n}\left(g\left(X'_i\right)-g\left(X_i\right)\right)\right|\right] \\
&= \mathbf{E}\left\|P'_n-P_n\right\|_G .
\end{aligned}$$

Proof - Symmetrization

We symmetrize again: for any $\epsilon_i \in \{+1, -1\}$,

$$\mathbf{E}\left[\sup_{g\in G}\left|\frac{1}{n}\sum_{i=1}^{n}\left(g\left(X_i'\right)-g\left(X_i\right)\right)\right|\right]=\mathbf{E}\left[\sup_{g\in G}\left|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i\left(g\left(X_i'\right)-g\left(X_i\right)\right)\right|\right]$$

Then we have

$$\begin{aligned}
&\mathbf{E}\left[\sup_{g\in G}\left|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i\left(g\left(X_i'\right)-g\left(X_i\right)\right)\right|\right]\\
\leq&\mathbf{E}\left[\sup_{g\in G}\left|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i g\left(X_i'\right)\right|\right]+\mathbf{E}\left[\sup_{g\in G}\left|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i g\left(X_i\right)\right|\right]\\
\leq&2\mathbf{E}\left\|R_n\right\|_G,
\end{aligned}$$

where $R_n(g)=(1/n)\sum_{i=1}^{n}\epsilon_i g\left(X_i\right)$ is the Rademacher process .

Proof - Restrictions

Lemma 2.3

For $A \subseteq \mathbb{R}^n$ *with* $R = \max_{a \in A} \|a\|_2$,

$$\mathbf{E} \sup_{a \in A} \langle \epsilon, a \rangle \leq \sqrt{2R^2 \log |A|}.$$

Hence

$$\mathbf{E} \sup_{a \in A} |\langle \epsilon, a \rangle| = \mathbf{E} \sup_{a \in A \cup -A} \langle \epsilon, a \rangle \leq \sqrt{2R^2 \log(2|A|)}.$$

Proof - Restrictions

For the class G of step functions, $R \leq 1/\sqrt{n}$ and $|A| \leq n+1$. Thus, with probability at least $1 - \exp\left(-2\epsilon^2 n\right)$,

$$\|P - P_n\|_G \leq \sqrt{\frac{8\log(2(n+1))}{n}} + \epsilon$$

By Borel-Cantelli, $\|P - P_n\|_G \xrightarrow{as} 0$.

Empirical Risk Minimization

We define a loss function $l(\theta, z)$ which measures how bad it is to choose θ when the outcome is z. For $Z \sim P$, the risk is $L(\theta) = Pl(\theta, z)$.

- Pattern classification: $\theta : \mathcal{X} \to \{0,1\}, z = (x,y) \in \mathcal{X} \times \{0,1\}$, $\ell(\theta,(x,y)) = 1[\theta(x) \neq y]$. Then we aim to choose $\theta \in \Theta$ to minimize the probability of misclassification.
- Density estimation: p_θ is a density, $X \sim P, p_{\theta^*}, \ell(\theta, z) = -\log p_\theta(z)$. Then we aim to choose θ to minimize

$$\mathbf{E} \log \frac{p_{\theta^*}(X)}{p_\theta(X)} = D_{KL}(p_{\theta^*} \| p_\theta)$$

- Regression: $\theta \in \mathbb{R}^p, z = (x,y), \ell(\theta,(x,y)) = |\theta' x - y|$. Then we aim to choose θ to minimize expected absolute error.

Empirical Risk Minimization

Suppose $Z_1, \ldots, Z_n$ are i.i.d. according to P. Define the empirical risk as

$$L_n(\theta) = P_n \ell(\theta, Z) = \frac{1}{n} \sum_{i=1}^{n} \ell(\theta, Z_i)$$

Empirical risk minimization chooses θ to minimize $L_n(\theta)$.

We are interested in controlling the excess risk,

$$L(\hat{\theta}) - \inf_{\theta \in \Theta} L(\theta) = L(\hat{\theta}) - L(\theta^*)$$

where θ^* minimizes L on Θ. We can decompose it as

$$L(\hat{\theta}) - L(\theta^*) = \left[L(\hat{\theta}) - L_n(\hat{\theta})\right] + \left[L_n(\hat{\theta}) - L_n(\theta^*)\right] + \left[L_n(\theta^*) - L(\theta^*)\right],$$

with approximation error and statistical error.

Empirical Risk Minimization

For statistical error, we have

$$L_n(\theta^*) - L(\theta^*) = \frac{1}{n}\sum_{i=1}^{n} \ell(\theta^*, Z_i) - P\ell(\theta^*, Z).$$

The law of large numbers shows that this term converges to zero. But more generally, we need to study the uniform laws of large numbers

$$L(\hat{\theta}) - L_n(\hat{\theta}) \le \sup_{\theta\in\Theta} |L(\theta) - L_n(\theta)| = \sup_{\theta\in\Theta} |P\ell_\theta - P_n\ell_\theta|.$$

We need to show ℓ_θ is a GC class (or prove a general form of GC Theorem).

Empirical Risk Minimization

Recall that

Definition 2.4

The Rademacher complexity of F is $\mathbf{E}\|R_n\|_F$, where the empirical process R_n is defined as

$$R_n(f) = \left| \frac{1}{n} \sum_{i=1}^{n} \epsilon_i f(X_i) \right|$$

where the $\epsilon_1, \ldots, \epsilon_n$ are Rademacher random variables: i.i.d. uniform on $\{\pm 1\}$.

Note that this is the expected supremum of the alignment between the random $\{\pm 1\}$-vector ϵ and $F\left(X_1^n\right)$, the set of n-vectors obtained by restricting F to the sample $X_1, \ldots, X_n$.

Uniform laws and Rademacher complexity

Theorem 2.5

For any $F, \mathbf{E}\|P-P_n\|_F \leq 2\mathbf{E}\|R_n\|_F$*. If* $F \subset [0,1]^{\mathcal{X}}$,

$$\frac{1}{2}\mathbf{E}\|R_n\|_F - \sqrt{\frac{\log 2}{2n}} \leq \mathbf{E}\|P-P_n\|_F \leq 2\mathbf{E}\|R_n\|_F$$

and, with probability at least $1-2\exp\left(-2\epsilon^2 n\right)$,

$$\mathbf{E}\|P-P_n\|_F - \epsilon \leq \|P-P_n\|_F \leq \mathbf{E}\|P-P_n\|_F + \epsilon$$

Thus, $\mathbf{E}\|R_n\|_F \to 0$ *iff* $\|P-P_n\|_F \xrightarrow{as} 0$.

The sup of the empirical process $P-P_n$ is concentrated about its expectation, and its expectation is about the same as the expected sup of the Rademacher process R_n.

Controlling Rademacher complexity

Control $\mathbf{E}\|R_n\|_F$:

- $\left|F\left(X_1^n\right)\right|$ small.
- For binary-valued functions: Vapnik-Chervonenkis dimension. Bounds rate of growth function. Can be bounded for parameterized families.
- Structural results on Rademacher complexity: Obtaining bounds for function classes constructed from other function classes.
- Covering numbers: Dudley entropy integral, Sudakov lower bound.
- For real-valued functions: scale-sensitive dimensions.

Extension: Glivenko-Cantelli Theorem of MDF

For $\forall \mathbf{u}, \mathbf{v} \in \mathcal{M}$, let

$$\delta(\mathbf{u}, \mathbf{v}, \mathbf{x}) = \prod_{k=1}^{K} I\left\{x_k \in \bar{B}\left(u_k, r_k\right)\right\} = \prod_{k=1}^{K} I\left\{x_k \in \bar{B}\left(u_k, d_k\left(u_k, v_k\right)\right)\right\}.$$

Definition 2.6 (Metric distribution function)

Given a probability measure μ, we define the metric distribution function $F_{\mu}^{M}(u, v)$ of μ on $\mathcal{M}$: $\forall \mathbf{u}, \mathbf{v} \in \mathcal{M}$,

$$F_{\mu}^{M}(\mathbf{u}, \mathbf{v}) = \mu\left[\prod_{k=1}^{K} \bar{B}\left(u_k, r_k\right)\right] = E[\delta(\mathbf{u}, \mathbf{v}, \mathbf{X})]$$

Extension: Glivenko-Cantelli Theorem of MDF

Suppose that $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ are iid samples generated from a probability measure μ on a product metric space $\mathcal{M} = \prod_{k=1}^{K} \mathcal{M}_k$. We define the empirical metric distribution function (EMDF) associated with μ by the following formula naturally:

$$F_{\mu,n}^{M}(\mathbf{u}, \mathbf{v}) = \frac{1}{n} \sum_{l=1}^{n} \delta\left(\mathbf{u}, \mathbf{v}, \mathbf{X}_l\right)$$

Extension: Glivenko-Cantelli Theorem of MDF

we define the collection of the indicator functions of closed balls on $\mathcal{M}$: $\mathcal{F} = \{\delta(\mathbf{u}, \mathbf{v}, \cdot) : \mathbf{u} \in \mathcal{M}, \mathbf{v}\}$.

Theorem 2.7

Let $\mathcal{M} = \prod_{k=1}^{K} \mathcal{M}_k$ be a product space and μ be a probability measure on it. Suppose that $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ is a sample of iid observations from μ. Define $\mathcal{F}\left(\mathbf{X}_1^n\right) := \{(f(\mathbf{X}_1), \ldots, f(\mathbf{X}_n)) \mid f \in \mathcal{F}\}$. If μ satisfies that

$$\frac{1}{n} E_{\mathbf{X}} \left[\log\left(\operatorname{card}\left(\mathcal{F}\left(\mathbf{X}_1^n\right)\right)\right)\right] \to 0$$

where $\operatorname{card}(\cdot)$ is the cardinality of a set, we have the Glivenko-Cantelli property of our empirical metric distribution function:

$$\lim_{n \to \infty} \sup_{\mathbf{u} \in \mathcal{M}, \mathbf{v} \in \mathcal{M}} \left| F_{\mu,n}^{M}(\mathbf{u}, \mathbf{v}) - F_{\mu}^{M}(\mathbf{u}, \mathbf{v}) \right| = 0, \; a.s.$$

Remark

The conditions of Theorem are often satisfied in practice.

- The first example is $\mathcal{M} = \mathbb{R}^q$ with the ℓ_p-norm (where p is a positive integer or ∞), and μ is an arbitrary probability measure because the set of ℓ_p ball has a finite VC-dimension. Since the VC-dimension of closed balls in Euclidean space R^q is $q+2$, if $q = o\left(\frac{n}{\log n}\right)$ the Glivenko-Cantelli property still holds.
- The second example is that $\mathcal{M}$ is a smooth regular curve in Euclidean space or a sphere in $\mathbb{R}^q$ with the geodesic distance, and μ is an arbitrary probability measure.
- The third example is that $\mathcal{M}$ is a set of polygonal curves in $\mathbb{R}^d$ with the Hausdorff distance for the Fréchet distance and μ is an arbitrary probability measure.
- Another example is that $\mathcal{M}$ is a separable Hilbert space with a probability measure μ with support on a finite-dimensional subspace because the set of balls on the support of μ has a finite VC-dimension.

Outline

Introduction

- Donsker's Theorem is a fundamental result in the field of probability theory and statistical inference.
- It generalizes the central limit theorem (CLT) to the setting of stochastic processes.
- Often referred to as the "Invariance Principle" or "functional central limit theorem".

Donsker's Theorem

Theorem 3.1 (Donsker's Invariance Principle)

Let $X_1, X_2, \dots$ be i.i.d. random variables with $\mathbb{E}[X_i] = 0$ and $Var(X_i) = 1$. Define the empirical process

$$S_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} X_i,$$

for $t \in [0,1]$. Then as $n \to \infty$, the process $S_n(t)$ converges in distribution in $D[0,1]$ to a standard Brownian motion $W(t)$.

The central limit theorem asserts that $S_n(1)$ converges in distribution to a standard Gaussian random variable $W(1)$ as $n \to \infty$. Donsker's invariance principle extends this convergence to the whole function $S_n(t)$.

Tightness

Here we define a concept of tightness for collections of measures and random variables. Intuitively this ensures that a collection of measures does not have mass that escapes to infinity. Tightness is often used to prove weak convergence.

Definition 3.2

Let $(S, \mathcal{S})$ be a measurable space. A collection of measures $\{\mu_i\}$ is tight if for all $\epsilon > 0$ there exists a compact set $K \in S$ such that $\sup_i \mu_i(K^c) < \epsilon$ for all i.

We say a random variable X is tight if for all $\epsilon > 0$ there is an M_ϵ such that

$$\mathbb{P}(\|X\| > M_\epsilon) < \epsilon$$

Tightness

Definition 3.3

A set is relatively compact if its closure is compact.

Let Π be a family of probability measures on $(S, \mathcal{S})$. We call Π relatively compact if every sequence of elements of Π contains a weakly convergent subsequence. Explicitly this means that if Π is relatively compact, then there exists a subsequence $\left(\mathbb{P}_{n_i}\right) \in \Pi$ and a probability measure Q, which need not be contained in $(S, \mathcal{S})$, such that $\mathbb{P}_{n_i} \Rightarrow_i Q$.

Theorem 3.4

If Π is tight, then it is relatively compact.

Corollary 3.5

If $(\mathbb{P}_n)$ is tight and each weakly convergent subsequence converges to $\mathbb{P}$, then the entire sequence converges weakly to $\mathbb{P}$.

Tightness

Definition 3.6

A modulus of continuity of an arbitrary function x is defined by

$$w(x,\delta) := \sup_{|s-t|\leq\delta} |x(s)-x(t)|$$

where $\delta \geq 0$.

Lemma 3.7

If

$$\left(X^n_{t_1},\ldots,X^n_{t_k}\right) \Rightarrow_n (X_{t_1},\ldots,X_{t_k})$$

holds for all $t_1,\ldots,t_k$, and if

$$\lim_{\delta\to 0}\limsup_{n\to\infty} \mathrm{P}\left[w\left(X^n,\delta\right) \geq \epsilon\right] = 0$$

for each positive ϵ, then $X^n \Rightarrow_n X$.

Tightness

Lemma 3.8

Suppose $0 = t_0 < t_1 < \ldots < t_k = 1$ *and*

$$\min_{1<i<k} (t_i - t_{i-1}) \geq \delta$$

If we define $I_i := [t_{i-1}, t_i]$, *then for arbitrary* x,

$$w(x,\delta) \leq 3 \max_{1\leq i\leq k} \sup_{s\in I_i} |x(s) - x(t_{i-1})|$$

and, for arbitrary $\mathbb{P}$,

$$\mathbb{P}[x : w(x,\delta) \geq 3\epsilon] \leq \sum_{i=1}^{k} \mathbb{P}\left[x : \sup_{s\in I_i} |x(s) - x(t_{i-1})| \geq \epsilon\right]$$

Proof Outline

- Show tightness of the sequence of processes.
- Demonstrate finite-dimensional convergence to those of the Brownian motion.
- Apply Prokhorov's theorem to conclude the weak convergence to Brownian motion.

Proof

Theorem 3.9

There exists on $(C, \mathcal{C})$ a probability measure, $\mathbb{W}$, with the finite dimensional distribution specified by Wiener measure.

Proof

Define

$$S_n(t) := \frac{1}{\sqrt{n}} S_{\lfloor nt \rfloor} + (nt - \lfloor nt \rfloor) \frac{1}{\sqrt{n}} X_{\lfloor nt \rfloor + 1}$$

at t. The function $X^n(w)$ is a linear interpolation (a linear mapping between points) between values at $S_i(t)/\sqrt{n}s$ at points i/n.

- The existence of the Wiener measure $\mathbb{W}$ is proven.
- Then

  $$(S_n(s), S_n(t) - S_n(s)) \underset{n}{\Rightarrow} (W_s, W_t - W_s)$$

  which implies

  $$(S_n(s), S_n(t)) \underset{n}{\Rightarrow} (W_s, W_t)$$

Proof

- Show the tightness

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}\left[w\left(S_n,\delta\right)\geq\epsilon\right]=0$$

to obtain

$$(S_n(t_1),\ldots,S_n(t_k))\underset{n}{\Rightarrow}\left(W_{t_1},\ldots,W_{t_k}\right)$$

- For lemma

$$\begin{aligned}\mathbb{P}\left[w\left(S_n,\delta\right)\geq 3\epsilon\right]&\leq\sum_{i=1}^{k}\mathbb{P}\left(\sup_{t_{i-1}\leq s\leq t_i}\left|S_n(s)-S_n(t_{i-1})\right|\geq\epsilon\right)\\&\leq\sum_{i=1}^{k}\mathbb{P}\left(\sup_{s<t_i-t_{i-1}}\left|S_s\right|\geq\epsilon\sqrt{n}\right)\\&\leq k\mathbb{P}\left(\max_{s\leq m}\left|S_s\right|\geq\epsilon\sqrt{n}\right).\end{aligned}$$

Proof

- By Etemadi's inequality, we then see that

$$\mathbb{P}\left[w\left(S_n,\delta\right)\geq 3\epsilon\right]\leq 3k\max_{s\leq m}\mathbb{P}\left[\left|S_s\right|\geq\frac{\epsilon\sqrt{n}}{3}\right]$$

- We can reformulate with Etemadi's inequality to be

$$\lim_{\lambda\to\infty}\limsup_{n\to\infty}\lambda^2\max_{s\leq n}\mathbb{P}\left[\left|S_s\right|\geq\lambda\sqrt{n}\right]=0.$$

Proof

- First, for large s, we use the central limit theorem to show that the partial sum converges to the standard normal distribution. So, by the central limit theorem, if s_λ in the maximum is large enough and $s_\lambda \le s \le n$, then

$$\mathbb{P}\left[|S_s| \ge \lambda\sqrt{n}\right] < \frac{3}{\lambda^4}$$

- In the second case, for small $s \le s_\lambda$, we use Chebyshev's inequality to show that

$$\mathbb{P}\left[|S_s| \ge \lambda\sqrt{n}\right] < \frac{s_\lambda}{\lambda^2 n}.$$

Donsker Theorem

Theorem 3.10 (Donsker (1952))

Let F be continuous distribution function. Define the empirical process:

$$\mathbb{G}_n(t) = \sqrt{n}(F_n(t) - F(t))$$

Then $\mathbb{G}_n$ converges weakly to a Brownian bridge G in the space $\mathcal{D}[0,1]$:

$$\mathbb{G}_n \rightsquigarrow G$$

where G is a Gaussian process with covariance function:

$$\mathbb{E}[G(s)G(t)] = F(s \wedge t) - F(s)F(t)$$

Definition and Properties

Definition

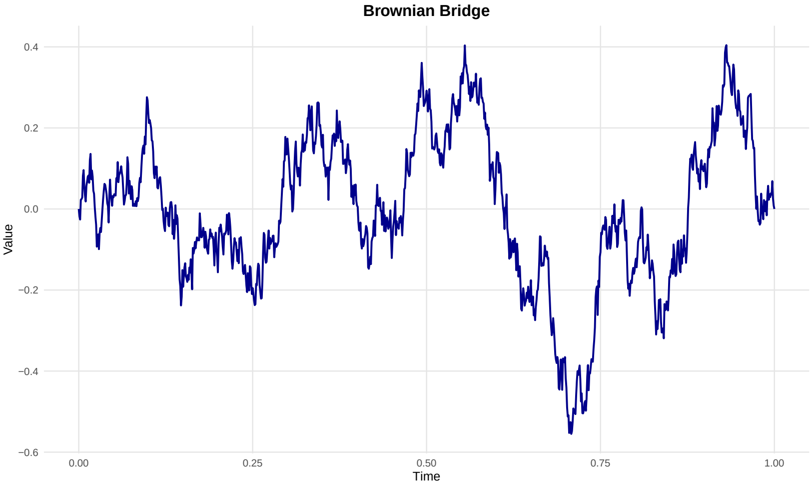
A **Brownian Bridge** is a stochastic process $B(t)$, for $t \in [0,1]$, defined by the conditional property that $B(0) = B(1) = 0$ given a standard Brownian motion $W(t)$. It can be expressed as:

$$B(t) = W(t) - tW(1)$$

Key Properties

- **Gaussian Process**: $B(t)$ has Gaussian increments with mean zero and covariance function given by $\min(s,t) - st$.
- **Continuity**: $B(t)$ enjoys the continuity properties of Brownian motion, but it is "pinned" at the endpoints 0 and 1 to be zero.

Brownian Bridge



Application: Extreme Process

Theorem 3.11 (Mapping Theorem)

If h is continuous on C, then $X^n \Rightarrow W$ implies $h(X^n) \Rightarrow h(W)$.

We can find the limiting distribution of $h(X^n)$ if we can find the distribution of $h(W)$, and we can in many cases find the distribution of $h(W)$ by finding the limiting distribution of $h(X^n)$ in some simple special case and then using $h(X^n) \Rightarrow h(W)$ in the other direction.

Application: Extreme Process

- Our goal is to derive the limiting distribution of

$$M_n = \max_{0 \le i \le n} S_i$$

- Since $h(x) = \sup_t x(t)$ is a continuous function on C, it follows from $X^n \Rightarrow W$ and the mapping theorem that $\sup_t X_t^n \Rightarrow \sup_t W_t$. Obviously, $\sup_t X_t^n = M_n / \sigma\sqrt{n}$, and so

$$\frac{M_n}{\sigma\sqrt{n}} \Rightarrow \sup_t W_t$$

Application: Extreme Process

- For the easy special case, assume that the independent ξ_i take the values ± 1 with probability $\frac{1}{2}$ each, so that $S_0, S_1, \ldots$ are the successive positions in a symmetric random walk starting from the origin.
- For each nonnegative integer a,

$$\mathrm{P}\left[M_n \geq a\right] = 2\mathrm{P}\left[S_n > a\right] + \mathrm{P}\left[S_n = a\right].$$

- Since

$$\mathrm{P}\left[M_n \geq a\right] - \mathrm{P}\left[S_n = a\right] = \mathrm{P}\left[M_n \geq a, S_n < a\right] + \mathrm{P}\left[M_n \geq a, S_n > a\right]$$

  The second term on the right is just $\mathrm{P}\left[S_n > a\right]$
- For reflection principle, we have

$$\mathrm{P}\left[M_n \geq a, S_n < a\right] = \mathrm{P}\left[M_n \geq a, S_n > a\right]$$

Application: Extreme Process

- Let $a_n = \lceil a n^{1/2} \rceil$, then

$$P\left[M_n/\sqrt{n} \geq a\right] = 2P\left[S_n > a_n\right] + P\left[S_n = a_n\right].$$

- The second term here goes to 0.
- $P\left[S_n > a_n\right] \to P[N > \alpha]$ by the central limit theorem, and so $P\left[M_n/\sqrt{n}\right] \to 2P[N > a]$ for $a \geq 0$.
- The limit distribution become

$$P\left[\sup_t W_t \leq a\right] = \frac{2}{\sqrt{2\pi}} \int_0^a e^{u^2/2} du, \quad a \geq 0$$

Application: Kolmogorov-Smirnov test

- The Kolmogorov-Smirnov (K-S) test is a nonparametric test used to determine whether two samples come from the same distribution.
- It compares the empirical distribution functions of two samples, or one sample with a theoretical distribution.
- It is particularly useful because it makes no assumption about the distribution of data.

The K-S Test Statistic

Definition

Given an empirical distribution function $F_n(x)$ for a sample and a theoretical distribution $F(x)$, the K-S test statistic is defined as:

$$D_n = \sup_x |F_n(x) - F(x)|$$

where sup denotes the supremum of the set of absolute differences.

Interpretation

D_n measures the maximum distance between the empirical distribution function of the sample and the theoretical distribution function.

Kolmogorov distribution

The Kolmogorov distribution is the distribution of the random variable

$$K = \sup_{t \in [0,1]} |B(t)|$$

where $B(t)$ is the Brownian bridge. The cumulative distribution function of K is given by

$$\Pr(K \leq x) = 1 - 2\sum_{k=1}^{\infty}(-1)^{k-1}e^{-2k^2x^2} = \frac{\sqrt{2\pi}}{x}\sum_{k=1}^{\infty}e^{-(2k-1)^2\pi^2/(8x^2)}.$$


Thank You