Probabilistic Analysis of Stochastic Processes

A Report submitted in partial fulfillment of requirements for the degree of

Master of Science (Research)

by

Mr. **Dakshesh Vasan**
Roll No: 18MS051
Integrated BS-MS Programme (2018-2023)
Indian Institute of Science Education and Research Kolkata

Under the supervision of
Dr. **Anirvan Chakraborty**
Assistant Professor, Department of Mathematics and Statistics
Indian Institute of Science Education and Research Kolkata

Dedicated to Anirvan sir,
Amma, Appa, Chandru, and Krithvi

अनन्याश्चिन्तयन्तो मां ये जनाः पर्युपासते।
तेषां नित्याभियुक्तानां योगक्षेमं वहाम्यहम्।।
~श्रीकृष्णः

Keep going, everything will be taken care of.

Author's Declaration

I hereby declare that this report is my own work, and to the best of my knowledge, contains neither material previously published or written by any other entity, nor substantial proportions of any material that has been accepted for the award of any other degree or diploma at IISER Kolkata or any other educational institution, except where due acknowledgement is made within this report. I certify that all copyrighted material incorporated into this report is in compliance with the Indian Copyright (Amendment) Act, 2012 and that I have received written permission from the copyright owners for my use of their work, which is beyond the scope of law. I agree to indemnify and save harmless IISER Kolkata from any and all claims that may be asserted or that may arise from any copyright violation.

19/05/2023

19 May 2023
Department of Mathematics and Statistics, IISER Kolkata

Dakshesh Vasan
18MS051

Certificate

This is to certify that the work entitled **'Probabilistic Analysis of Stochastic Processes'** embodies the study done by *Dakshesh Vasan* under my guidance and supervision for the partial fulfilment of the degree of *Master of Science (MS), Indian Institute of Science Education and Research Kolkata, India.* It has not previously formed the basis for the award of any Degree, Diploma, Associate-ship or Fellowship to him.

Anirvan Chakraborty. 19/05/23.

19 May 2023
Department of Mathematics and Statistics, IISER Kolkata

Dr. Anirvan Chakraborty
Assistant Professor

Abstract

Gaussian Processes are one of the most important family of stochastic processes given their wide range of use across scientific disciplines and profound implications in Statistics and Mathematics. This report, in three parts, covers some modern developments of tools that are applicable to the analysis of general Gaussian processes. The first part recollects basics, and provides motivation for a 'General Theory'. It also sets the background setup and broad goal of this report - to concern with sample path continuity of general Gaussian processes. The second part introduces a fundamental tool for the modern approach, based on which the two backbone concepts of the general theory - entropy and majorising measures are defined. Both these concepts attempt to measure the 'size' of a parameter space in different ways. The second part also sets the specific goal for the report - to understand and appreciate a 'Main Entropy Result' stated immediately after introducing entropy. Following this, essential inequalities of Gaussian process theory are covered and their implications in the analysis of boundedness and suprema distributions are discussed. Boundedness is related to continuity, and this leads to the third part which introduces majorising measures in the context of proving the Main Entropy Result. Other strong characterisations of continuity and boundedness of any general Gaussian process are also noted, following which the report covers specific examples of applications of the Main Entropy Result. The report concludes after having discussed and appreciated several motivations, applications and powerful implications of the entropy-based results while noting the existence of more powerful but harder results based on majorising measures.

Acknowledgement

I am extremely grateful to Dr. Anirvan Chakraborty, my mentor, for providing me a direction. He introduced me to Probability Theory and Stochastic Processes, and it is his guidance and support that has improved my confidence with which I approach new material or think of a problem. I am indebted to him for his advice throughout the past year. I will remain a fan of his problem-solving and try to emulate his way of attacking a problem fundamentally. I also acknowledge the contributions of Prof. Robert J. Adler, the probabilist behind the Notes I have studied for this Project. Deciphering his commentary with Dr. Chakraborty's guidance has introduced me to several topics in Gaussian Process Theory. I am grateful to Dr. Satyaki Mazumder for always being available when needed. He has looked out for me the past year. His words of encouragement have propelled me forward.

I also extend my sincere gratitude to the Dept. of Mathematics and Statistics, and the Dean of Academic Affairs, Prof. Asok K. Nanda, at IISER Kolkata; for providing me with the resources to study. I thank the Indian Nation and the KVPY-Scheme for the opportunity to study at low-cost without significant compromise in quality and facilities through the IISER-system, thereby enabling me to compete in the international domain.

All that went behind this project was possible only because of the support system I always have/had at some point: Ms. Annapoorani Vasan, Mr. Raman Vasan, Mr. Balachandar Vasan, Ms. Krithvi Vasan, Mr. Harsh Kumar, Mr. Praful Rahangdale, Ms. Kalpana Das, Ms. Ritika Sethi, Ms. Chinmayi Prabhu, Mr. Amardeep Kushwah, Mr. Abhiroop Chauhan, Mr. Soumajit Gayen, Mr. Siddikanta Mishra, Mr. Madhur Bhatt, Mr. Vivek Kumar Singh, and Mr. Ankit Raj. I am lucky to have or have had them in my life at any point of time.

I am grateful to Prof. Manjunath Krishnapur for his notes which helped me study comparison inequalities. I specially acknowledge Mr. A.B. Vishwanath who helped me with technical issues, and some of my friends with whom I share connections that give me joy, lessons, and support. These include Mr. Arkadeep Mitra, Mr. Aditya Dwarkesh, Mr. Shreyash Borkar, Area51, Lit-Quiz-Aarshi Gang, the Quantum Sociology Gang, Inquivesta-X (2020, 2023), Revival(2021), Change(2020), Change2.0(2023), SAC(2020-22), Mr. Anshuman Jaysingh, and Mr. Himanshu Upadhyay. These are some people who either look out for me, look up to me, or whom I look upto. I am grateful for their existence.

Contents

Part I

Preliminaries

Part I is both a recap of definitions and standard results in probability theory, and an introduction and motivation for a 'General Theory' of Gaussian processes. The terms and notations introduced here are what will be used in the entire report.

Part I begins from the concept of a 'random experiment' and takes it to the more precise measure-theoretic definitions in an attempt to establish a connection with the layman's idea of probability and actual probability theory. Measure-theoretic definitions and constructions of all fundamental concepts and entities are given a quick look. This is followed by a brief discussion on stochastic processes, and especially Gaussian processes. Real-life applications are mentioned along with a few mathematical implications, to provide motivation for the analysis of stochastic processes and particularly Gaussian processes.

Part I also gives an introduction to the dilemma of there being too many Gaussian processes of interest in existence. It then introduces a relatively modern approach in probabilistic analysis of stochastic processes that has been worked on by Dudley, Fernique, Talagrand, Marcus, and many other mathematicians since the later part of the last century. It stresses the need for a modern 'general' theory encompassing the concepts of entropy and majorising measures to study Gaussian processes, with the idea that the geometric structure of the parameter space has little to do with continuity and boundedness of a Gaussian process. Finally, Part I sets the background of the analysis that follows.

Chapter 1

The Probability Theory Starter Pack

1.1 Fundamental Notions of Probability Theory

Measure-Theoretic Definitions

Definition 1.1.1. A **Random Experiment** is an experiment that satisfies the following:

- The experiment is entirely *repeatable any number of times under identical conditions.*
- The experiment has *no deterministic outcome.* That is, it is possible for the experiment to produce a different outcome each time it is repeated.
- The exact outcome of the experiment *cannot be predicted in advance.*

Now, consider a random experiment. Let Ω be the sample space, ie, the set of all possible outcomes of said random experiment.

As elucidated by Stayer and Nagel[13], let $\mathfrak{F}$ be the set of possible events, ie, a set of some/all subsets of the sample space. The pair $(\Omega, \mathfrak{F})$ can be formalised by considering Ω as a non-empty set and $\mathfrak{F}$ as a σ-algebra on Ω. Then,

- Any $\omega \in \Omega$ is called an **Outcome**.
- Any $\{\omega\}$ is called an **elementary event**.
- Any $F \in \mathfrak{F}$ is called an **event**.

Further, a '*probability measure*' is defined as follows:

Definition 1.1.2. Let $\mathcal{P}$: $\mathfrak{F}$→[0,1] be a finite measure, such that $\mathcal{P}(\Omega) = 1$. Then $\mathcal{P}$ is called a **probability measure**.

For real world events, this probability measure acted on an event '*measures*' the probability of an event.

Definition 1.1.3. The measure space, $(\Omega, \mathfrak{F}, \mathcal{P})$ is called a **Probability Space**.

The conditional probability of an event E $\in \mathfrak{F}$ given F is given is given by a conditional probability measure acted on E.

Definition 1.1.4. A finite measure $\mathcal{P}^F$:$\Omega \to [0,1]$ defined by $\mathcal{P}^F(E) = P(E|F) \forall E \in \mathfrak{F}$ is called the **F-conditional probability measure** on $(\Omega, \mathfrak{F})$.

Definition 1.1.5. Let $(\Omega, \mathfrak{F}.\mathcal{P})$ be a probability space. If I is a non-empty index set and $\{F_i | i \in I\}$ is a family of events such that P$(\cap_{i \in I'} F_i) = \Pi_{i \in I'} P(F_i)$ $\forall$ $I' \subset$ I, then the said family of events is said to be '**independent**'.

Definition 1.1.6. Say τ is some appropriate topological space and $\mathfrak{B}_\tau$ be the Borel σ-algebra of τ. Then any measurable function X: $(\Omega, \mathfrak{F}, \mathcal{P}) \to (\tau, \mathfrak{B}_\tau)$ is said to be a τ-valued **random variable**.

Now, let X be a τ-valued random variable defined on $(\Omega, \mathfrak{F}, \mathcal{P})$. Since, there is now a measurable function, it is possible to define a push-forward measure in $(\tau, \mathfrak{B}_\tau)$. How can this be done?

Definition 1.1.7. The function $F_X : \mathfrak{B}_\tau \to [0,1]$ defined by $F_X(B) = \mathcal{P}(X^{-1}(B))$ $\forall$ B $\in \mathfrak{B}_\tau$ is called the **distribution function** of X.

This distribution function F_X is essentially the push-forward measure looked for in $(\tau, \mathfrak{B}_\tau)$. This function is also called the **cumulative distribution function**.

Definition 1.1.8. Note that if μ is a canonical Borel-measure in $(\tau, \mathfrak{B}_\tau, \mu)$ a non-negative measurable function $f : (\tau, \mathfrak{B}_\tau, \mu) \to (\bar{\mathbb{R}}, \mathfrak{B}_{\bar{\mathbb{R}}})$ can be obtained such that $F_X(B) = \int_B f d\mu$ $\forall$ B $\in \mathfrak{B}_{\bar{\mathbb{R}}}$. This function f is called the **probability density function** associated with X.

Further, by consequence of the Lebesgue Radon-Nikodym Theorem for signed measures, the probability density function, $f = \frac{dF_X}{d\mu}$.

Now, consider a random variable X such that $\exists$ a countable subspace $\tau' \subset \tau$ such that $F_X(\tau') = 1$ and $\{x\} \in \mathfrak{B}_\tau \forall x \in \tau'$. Then, X and its distribution F_X are called '***discrete***'. For this discrete case, the counting measure defined on $\mathfrak{B}_\tau$ can be used to obtain an appropriate density (mass) function, f.

Definition 1.1.9. If X is a τ-valued random variable and g: $\tau \to \mathbb{R}$ is a Borel function, then the **expectation** of g(X) is defined as the Lebesgue-Stieltjes integral E{g(X)} = $\int_\tau g(x) dF_X(x)$, provided the integral exists.

Definition 1.1.10. If X is a random variable, the function $\phi : \mathbb{R} \to \mathbb{C}$ defined by, $\phi(t) =$ E$[e^{itX}]$ is called it's **characteristic function**.

The characteristic function completely determines the behaviour and properties of the probability distribution of the random variable X. In fact, each characteristic function uniquely corresponds to a particular probability distribution and vice versa.

Modes of Convergence

Definition 1.1.11. Convergence in Probability
If (τ,d) is a metric space and $\{X_n\}_{n\in\mathbb{N}}$ is a sequence of τ-valued random variables and X is a τ-valued random variable, all defined on the same probability space such that:

For every $\epsilon > 0$, P(d$(X_n - X) > \epsilon) \to 0$ as n $\to \infty$, then $\{X_n\}_{n\in\mathbb{N}}$ is said to **converge in probability** to X.

This is denoted as $X_n \xrightarrow{\text{P}}$ X.

Definition 1.1.12. Convergence in Distribution/Weak Convergence
If $\{X_n\}_{n\in\mathbb{N}}$ is a sequence of random variables and X is a random variable, which may/may not be defined on the same probability space, such that:

$P(X_n \le x) \to P(X \le x)$ as n → ∞, then $\{X_n\}_{n\in\mathbb{N}}$ is said to **converge in distribution/weakly converge** to X.

This is denoted as $X_n \xrightarrow{\text{d}}$ X.

Definition 1.1.13. Convergence in L^p
If (τ,d) is a metric space and $\{X_n\}_{n\in\mathbb{N}}$ is a sequence of τ-valued random variables and X is a τ-valued random variable, all defined on the same probability space such that:

$E((d(X_n - X))^p) \to 0$ as n → ∞ for some p > 0, then $\{X_n\}_{n\in\mathbb{N}}$ is said to **converge in L^p** to X.

This is denoted as $X_n \xrightarrow{L^p}$ X.

Definition 1.1.14. Almost Sure Convergence
If (τ,d) is a metric space and $\{X_n\}_{n\in\mathbb{N}}$ is a sequence of τ-valued random variables and X is a τ-valued random variable, all defined on the same probability space such that:

$P\{\omega \mid d(X_n(\omega) - X(\omega)) \to 0 \text{ as n} \to \infty\} = 1$, then $\{X_n\}_{n\in\mathbb{N}}$ is said to **converge almost surely** to X.

This is denoted as $X_n \xrightarrow{\text{a.s}}$ X.

Remark 1.1.1. Relationships between Different Modes of Convergence

1. $X_n \xrightarrow{\text{a.s}} X \implies X_n \xrightarrow{\text{P}} X \implies X_n \xrightarrow{\text{d}} X$.
2. $X_n \xrightarrow{L^p}$ X for some p > 0 $\implies X_n \xrightarrow{\text{P}}$ X $\implies X_n \xrightarrow{\text{d}} X$.
3. If p ≥ q > 0, $X_n \xrightarrow{L^p} X \implies X_n \xrightarrow{L^q} X$.
4. If $X_n \xrightarrow{\text{P}} X$, then ∃ a sub-sequence $\{X_{n_k}\}_{k\in\mathbb{N}}$ of $\{X_n\}_{n\in\mathbb{N}}$ such that $X_{n_k} \xrightarrow{\text{a.s}} X$ as $k \to \infty$.
5. If $X_n \xrightarrow{\text{d}} X$ where X is some degenerate random variable, then $X_n \xrightarrow{\text{P}} X$.

1.2 Important and Useful Results

Results on Almost Sure Convergence

Result 1.2.1. If (τ,d) is a metric space and $\{X_n\}_{n\in\mathbb{N}}$ is a sequence of τ-valued random variables and X is a τ-valued random variable, all defined on the same probability space; the following are equivalent:

- $X_n \xrightarrow{\text{a.s}} X$
- P$\{\omega \mid d(X_n(\omega) - X(\omega)) > \frac{1}{k}$ infinitely often$\} = 0$ $\forall$ k $\in \mathbb{N}$.

Result 1.2.2. Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of (τ, d)-valued random variables such that $\forall\ \epsilon > 0, \sum_{n=1}^{\infty} P(d(X_n - X) > \epsilon) < \infty$ for some random variable X. Then, $X_n \xrightarrow{\text{a.s}}$ X.

Result 1.2.3. Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables. This sequence converges almost surely if either of the following hold:

1. $\sum_{n=1}^{\infty} sup_{m\geq n} P(d(X_n - X_m) \geq \epsilon) < \infty$ for every $\epsilon > 0$.
2. $\sum_{n=1}^{\infty} P(d(X_{n+1} - X_n) > \delta_n) < \infty$ where $\delta_n > 0$ and $\sum_{n=1}^{\infty} \delta_n < \infty$.

Other Important Results

Theorem 1.2.1. *(Slutsky's Theorem)*
If $g : \mathbb{R}\ x\ \mathbb{R} \to \mathbb{R}$ is continuous, and $\{X_n\}_{n\in\mathbb{N}}$, $\{Y_n\}_{n\in\mathbb{N}}$ are sequences of real-valued random variables such that $X_n \xrightarrow{d} X$ where X is a real valued random variable and $Y_n \xrightarrow{d} c$ where $c \in \mathbb{R}$; then $g(X_n, Y_n) \xrightarrow{d} g(X, c)$ as $n \to \infty$.

Theorem 1.2.2. *(Scheffe's Theorem)*
For each $n \in \mathbb{N}$, let X_n be a random variable having density f_n and X be a random variable having density f. Then, $f_n \to f$ as $n \to \infty \implies X_n \xrightarrow{d} X$.

Theorem 1.2.3. *(Lévy Continuity Theorem)*
Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables with the corresponding unique sequence of characteristic functions being $\{\phi_n\}$. Then:

1. *If X is a random variable X with characteristic function ϕ, then $X_n \xrightarrow{d} X \iff \phi_n \to \phi$ pointwise as $n \to \infty$.*
2. *If $\phi_n \to \psi$ pointwise as $n \to \infty$ for some function ψ which is continuous at 0; then*
   - *$\exists$ a probability measure/random variable X such that ψ is the unique characteristic function of X.*
   - *$X_n \xrightarrow{d} X$.*

Result 1.2.4. (Borel Cantelli Lemmas - 1st and 2nd respectively)

1. If $\{E_n\}_{n\in\mathbb{N}}$ is a sequence of events in a probability space such that $\sum_{n=1}^{\infty} P(E_n) < \infty$, then P($E_n$ infinitely often) = 0 $\iff$ P$(limsup_{n\to\infty} E_n) = 0$.
2. If $\{E_n\}_{n\in\mathbb{N}}$ is a sequence of pairwise independent/mutually independent events in a probability space such that $\sum_{n=1}^{\infty} P(E_n) = \infty$, then P($E_n$ infinitely often) = 1 $\iff$ P$(limsup_{n\to\infty} E_n) = 1$.

Theorem 1.2.4. *(Monotone Convergence Theorem)*
Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of non-negative real valued random variables for which $X_n(\omega) \leq X_{n+1}(\omega)$ for all n and ω, and suppose there exists a real valued random variable X such that $X_n \xrightarrow{a.s} X$. Then, $E\{X\} = lim_{n\to\infty} E\{X_n\}$.

Result 1.2.5. (Fatou's Lemma)
Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of real valued random variables and suppose there exists an integrable random variable X such that $X_n(\omega) \geq X_{n+1}(\omega)$ for all n and ω. Then, $liminf_{n\to\infty} E\{X_n\} \geq E\{liminf_{n\to\infty} X_n$.

Theorem 1.2.5. *(Dominated Convergence Theorem)*
Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of real valued random variables and X such that $X_n \xrightarrow{a.s} X$.

Suppose that there exists an integrable real-valued random variable Y *for which* $|X_n(\omega)| \leq Y(\omega)$ *for all* n *and* ω*. Then,* $lim_{n\to\infty} E\{X_n\} = E\{X\}$.

Chapter 2

Stochastic Processes

What are stochastic processes and why are they of interest?

2.1 What is a Stochastic Process?

Simply put, a stochastic process refers to any family of random variables. This could be anything, from a finite set of random variables up-to an uncountable collection of random variables defined on a probability space. Now, why call any such collection a 'process'?

Imagine a real-world phenomenon like rainfall at a given location. Suppose the occurrence of rainfall over time, at this location is to be studied. The simplest probabilistic model is obtained by assuming the daily observations of rainfall at the given location over a certain period of time, say a day, to be realisations of an independent and identically distributed sequence of random variables, each denoting the amount of rainfall at the given location on a particular day. Then, the observations are interpreted as a sample from a collection of random variables (the iid sequence), and each new day's observation is a realisation of the next random variable. Thus, we are studying a system's evolution over time, a 'process'. An attempt to solve any problem, or any investigation related to rainfall patterns at the given location over time, then boils down to an analysis of a stochastic process. Here, the process is a discrete time-indexed one, also called a 'time series'. In fact, any basic probabilistic approach to studying a real-world problem involves stochastic processes upon which further statistical/empirical analyses are done. The word 'stochastic' refers to the 'randomness' in the system and so, a stochastic process can ultimately be interpreted as the representation of the evolution of a system (or a 'process') comprising of random fluctuations over time - which may be continuous or discrete.

The calculation/estimation of one or more parameters of a stochastic process is enough to solve problems in many real-life applications, like in the studies of natural disasters, climate change, cellular processes, plant-germination, the behaviour of celestial objects beyond our skies, or even the random motion of molecules in specific states of matter. Hence, results from probabilistic analyses of stochastic processes are of great interest to scientists/statisticians/applied probabilists. However, there's much more to see than what meets the eye.

As is typical of an exploration by a mathematician, some abstraction into stochastic processes fuelled by a lot of curiosity has given way to many particularly interesting developments. In fact, theoretical ventures have given rise to greater avenues for applications in real-life, directly and indirectly; by helping solve major problems in several other mathematical domains of study. Quoting Robert J. Adler, "*These processes* that are used to provide mathematical models of real-life phenomena are not only the most useful but also generate the most interesting mathematics."[2] It is not difficult to note that just as it is when we move on from an element to a sequence of elements or from a function to a sequence of functions- moving from a random variable to investigating a family of random variables brings a voluminous amount of content to their analysis. The *randomness* further adds more to keep in mind while investigating the behaviour of these 'random functions'. For example, when we want to study convergence, we need to look at the different modes of convergence that we can arrive at. A similar situation then arises naturally for continuity. Then there is the behaviour of tails, the boundedness and behaviour of suprema/infima, and so on to look at. Unlike simple functions, here the distributions of the random variables will also play an additional role in determining the properties of the process. Hence, the *stochastic* nature of these processes adds to the awe and vastness of the subject.

Further, any stochastic process is also indexed by a non-empty set called the parameter space. Does the parameter space contribute to the nature of a process, and its applicability in any field of study? Let us first consider a simple time-indexed process $\{W_t | t \geq 0\}$. This process can also be visualised as $\{W((0,t]) | t \geq 0\}$. Now, we have a **set-indexed process**. From here, we can extend to a set-indexed process on the Borel $\sigma-$algebra: $\{\text{W(B)} \ | B \in \mathfrak{B}_{\mathbb{R}}\}$. Going in another direction, we can obtain higher-order Euclidean space-indexed processes called **Random Fields**. These random fields have tremendous applications in the modelling of surfaces that are generally rough and erratic, in areas as diverse as biology, geography, and turbulence studies.[2]. Studies of the ocean surface, metallic surfaces, image analysis, etc. have all been benefited by developments in the study of random fields. Yet another direction yields the function-indexed **Generalised Random Fields**, and **measure-indexed** processes. **Vector-valued** and **Banach space-valued processes** are other examples. Thus, the parameter space clearly brings more flavour into the study and analysis of stochastic processes.

The undoubted importance that stochastic analysis holds, given the vast implications of stochastic processes, serves as motivation to understand recent developments in the probabilistic analysis of an important family of stochastic processes. In fact, the diverse subject matter involved in the study of stochastic processes itself serves as motivation to invest into probability theory.

Let us now go through some fundamentals about stochastic process, beginning with a countably infinite sequence of random variables.

Infinite Sequence of Random Variables

By extending on the measure-theoretic construction of a random variable, we can also construct finite sequences of random variables (or random vectors in a finite dimensional vector space). However, the existence of an infinite sequence of random variables is not trivial because whenever we think of a sequence of random variables, we can at most be sure of the distribution of a finite sub-collection. The following result allows us to 'extend' on the same.

Result 2.1.1. Let $\{\mu_n\}_{n\geq 1}$ be a sequence of probability measures such that:
i) for each n $\in \mathbb{N}, \mu_n$ is a probability measure on $(\mathbb{R}^n, \mathfrak{B}_{\mathbb{R}^n})$,
ii) for each n $\in \mathbb{N}$, μ_{n+1}(B x $\mathbb{R}$) = $\mu_n \forall$ B $\in \mathfrak{B}_{\mathbb{R}^n}$.

Then, $\exists$ an infinite sequence $\{X_n : n \geq 1\}$ on a probability space $(\Omega, \mathfrak{F}, \mathcal{P})$ with $\Omega = \mathbb{R}^\infty, \mathfrak{F} = \mathfrak{B}_{\mathbb{R}^\infty}$ such that for each n $\in \mathbb{N}$, the distribution of $(X_{t_1}, X_{t_2}, ..., X_{t_n})$ is μ_n, for any $t_1, t_2, ..., t_n \in \mathbb{N}$, for each $n \in \mathbb{N}$.

Note: $\mathbb{R}^\infty$ is the space of all real sequences, ie, $\mathbb{R}^\infty = \{(x_1, x_2, ..., x_n, ...) | x_j \in \mathbb{R}$ $\forall j \in \mathbb{N}\}$.

Construction

By defining an appropriate sequence of measures that follow the conditions **(i) and (ii)** of the above result, any sequence of random variables can be constructed.

An independent and identically distributed (iid) sequence is the most frequently used concept in many applications. Lets look at a construction of the same.

Example 2.1.1. Construction of an iid Unif([0,1]) sequence:

- Define F: $([0,1], \mathfrak{B}_{[0,1]}) \to$ [0,1] such that F(B) $= \int_{B\cap[0,1]} d\nu(x)$ $\forall$ B $\in \mathfrak{B}_{[0,1]}$ where ν is the Borel measure on $\mathfrak{B}_{\mathbb{R}}$.
- Define a sequence of probability measures $\{\mu_n\}_{n\geq 1}$ such that for each n $\in \mathbb{N}$, μ_n is a probability measure on ($\mathbb{R}^n$, $\mathfrak{B}_{\mathbb{R}^n}$) defined by $\mu_n(A) = \Pi_{i=1}^n(\int_{A_i\cap[0,1]} \nu(dx)) = \Pi_{i=1}^n F(A_i)$ where A $= \Pi_{i=1}^n A_i \in \mathfrak{B}_{\mathbb{R}^n}$ and each $A_i \in \mathfrak{B}_{\mathbb{R}}$.
- Then, it is easily verified that the sequence $\{\mu_n\}_{n\in\mathbb{N}}$ obeys conditions (i) and (ii) in the above Result 2.2.1.

- Hence, $\exists$ an infinite sequence of random variables $\{X_n : n \geq 1\}$ on a probability space $(\mathbb{R}^\infty, \mathfrak{B}_{\mathbb{R}^\infty}, \mathcal{P})$ such that the distribution of $(X_{t_1}, X_{t_2}, ..., X_{t_n})$ is μ_n, for any $t_1, t_2, ..., t_n \in \mathbb{N}$, for each $n \in \mathbb{N}$.
- Now, observe that for each n $\in \mathbb{N}$, the distribution of X_i is μ_1. Further, for any n $\in \mathbb{N}$, for any $t_1, t_2, ..., t_n \in \mathbb{N}$, and for any $A_1, ..., A_n \in \mathfrak{B}_{\mathbb{R}}$;

$$P(X_{t_1} \in A_1 | (X_{t_2}, ..., X_{t_n}) \in \Pi_{i=2}^n A_i) = \frac{P((X_{t_1}, X_{t_2}, ..., X_{t_n}) \in \Pi_{i=1}^n A_i)}{P((X_{t_2}, ..., X_{t_n}) \in |\Pi_{i=2}^n A_i} = \frac{\mu_n(\Pi_{i=1}^n A_i)}{\mu_{n-1}(\Pi_{i=2}^n A_i)}$$
$$= \frac{\Pi_{i=1}^n F(A_i)}{\Pi_{i=2}^n F(A_i)} = F(A_1) = \mu_1(A_1) = P(X_{t_1} \in A_1).$$

Hence, we have an iid sequence of Unif([0,1]) random variables.

Standard Results on Sequences of Random Variables

Result 2.1.2. Strong Law of Large Numbers (SLLN)
Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of pairwise independent and identically distributed random random variables with E$|X_1| < \infty$. Then, $\frac{\sum_1^n X_i}{n} \xrightarrow{\text{a.s}} EX_1$ as n$\to \infty$. [This version of SLLN was proved first by Etemadi.]

Result 2.1.3. Weak Law of Large Numbers (WLLN)
Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of pairwise independent and identically distributed random random variables with E$|X_1| < \infty$. Then, $\frac{\sum_1^n X_i}{n} \xrightarrow{\text{P}} EX_1$ as n$\to \infty$.

Result 2.1.4. Central Limit Theorem (CLT)
Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of independent and identically distributed random variables. Then,

$$\sqrt{n}\frac{\frac{\sum_1^n X_i}{n} - EX_1}{\sqrt{(EX_1^2 - E^2X_1)}} \xrightarrow{\text{d}} N(0,1)$$

Standard Results on Series of Independent Random Variables

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of independent random variables on a probability space $(\Omega, \mathfrak{F}, \mathcal{P})$. Here listed are some interesting standard results that come handy, on the behaviour of the sequence of partial sums of these random variables $\{S_n\}_{n\in\mathbb{N}}$ where for each n $\in \mathbb{N}, S_n = \sum_{i=1}^{n} X_i$.

Result 2.1.5. (Lévy): If $S_n \xrightarrow{P} S$, then $S_n \xrightarrow{a.s} S$ as n $\to\infty$.

Result 2.1.6. Khinchin-Kolmogorov 1-Series Theorem
If E$X_n = 0$ $\forall$ n $\in \mathbb{N}$ and $\sum_{n=1}^{\infty} EX_n^2 < \infty$, then S_n converges almost surely and in L^2.

Result 2.1.7. Kolmogorov's 3-Series Theorem
S_n converges almost surely iff 1., 2., and 3. hold for some $0 < c < \infty$; where 1., 2., and 3. are as follows:

1. $\sum_{i=1}^{\infty} P(|X_i| > c) < \infty$
2. $\sum_{i=1}^{\infty} EY_i < \infty$
3. $\sum_{i=1}^{\infty} VarY_i < \infty$

where $Y_i = X_i \mathbb{1}_{(|X_i| \le c)}$; $\forall$ i ≥ 1.

Note: $\mathbb{1}$ is the indicator function on Ω, ie, For any A $\in \mathfrak{F}$, $\mathbb{1}_A(\omega) = 1$ when $\omega \in A$ and 0, otherwise.

Remark 2.1.1. S_n converges almost surely if 1., 2., 3. hold for some $0 < c < \infty$ and only if 1., 2., 3. hold $\forall$ c $\in (0, \infty)$.

From Sequence to Processes

Now, the notion of an infinite sequence of random variables can be generalised and extended to a collection of random variables indexed by an arbitrary non-empty set.

What makes a Stochastic Process?

Definition 2.1.1. A **stochastic process** with index set/parameter space, T is a family of random variables $\{X_t | t \in T\}$ defined on an appropriate probability space $(\Omega, \mathfrak{F}, \mathcal{P})$. [T can be any arbitrary well-defined set/family/collection of mathematical objects.]

An infinite sequence of random variables can model random phenomena occurring over a discrete set of time-points, etc. However, a collection of random variables indexed over say, [0,1], or $\mathbb{R}$, would be able to model random phenomena over a continuum.

Stochastic processes can also be perceived as *random functions.* $\{X_t(\omega) | t \in T\}$ can be equivalently written as $\{f(\omega, t) | t \in T\}$. Then,

- If we fix $\omega \in \Omega, \{f(\omega, t) | t \in T\}$ is a function. This is also called a *sample path.*
- If we fix $t \in T, \{f(\omega, t) | \omega \in \Omega\}$ is a random variable.

Just as we analyse real-valued functions, we can analyse these 'random' functions. Does the stochastic process have continuous sample paths? Are the sample paths differentiable/integrable? Such questions then arise and give rise to the analysis of stochastic processes.

While the existence and construction of an infinite sequence of random variables has been discussed appropriately, how can we be guaranteed the existence of any stochastic process? For this, we need a fundamental result credited to Kolmogorov.

Existence of a Stochastic Process

Definition 2.1.2. The Family of **finite dimensional distributions (fidis)** associated with a stochastic process $\{X_t | t \in T\}$ is the family of probability distributions,

$$\mu_{t_1, t_2, t_3, \ldots, t_k}(.) = P\{(X_{t_1}, X_{t_2}, \ldots, X_{t_k}) \in (.) | t_1, t_2, \ldots, t_k \in T, 1 \leq k < \infty\}.$$

Given a stochastic process exists, we would have a fidis associated with it. Furthermore, this fidis would satisfy the following consistency conditions:

For any $t_1, t_2, \ldots, t_k \in T, 2 \leq k < \infty$, and $B_1, B_2, \ldots, B_k \in \mathfrak{B}_{\mathbb{R}}$;

C.1 $\mu_{t_1, t_2, \ldots, t_k}(B_1 \times B_2 \times \ldots \times B_{k-1} \times \mathbb{R}) = \mu_{t_1, t_2, \ldots, t_{k-1}}(B_1 \times \ldots \times B_{k-1})$.

C.2 For any permutation $(i_1, i_2, ..., i_k)$ of (1, 2, ..., k) we have $\mu_{t_{i1},t_{i2},...,t_{ik}}(B_{i1} \times ... \times B_{ik}) = \mu_{t_1,t_2,...,t_k}(B_1 \times .... \times B_k)$.

Then, given any fidis, does there exist a stochastic process whose associated fidis is the same fidis given? And that is what the Kolmogorov Consistency/Existence Theorem tells us.

Theorem 2.1.1. *Kolmogorov Consistency/Existence Theorem*
Let T be a non empty set. Let $\mathcal{Q}_T = \{\mu_{t_1,t_2,...,t_k} |\ t_1, ..., t_k \in T; 1 \le k < \infty\}$ be a family of probability distributions such that for each $t_1, ..., t_k \in T; 1 \le k < \infty$, we have :

- *$\mu_{t_1,...,t_k}$ is a valid probability distribution on $(\mathbb{R}^k, \mathfrak{B}_{\mathbb{R}^k})$*
- *The Consistency Conditions (C1) and (C2) mentioned above are satisfied by $\mathcal{Q}_T$.*

Then, $\exists$ a probability space $(\Omega, \mathfrak{F}, \mathcal{P})$ and a stochastic process $\{X_t |\ t \in T\}$ such that $\mathcal{Q}_T$ is the collection of fidis of $\{X_t |\ t \in T\}$.

The Kolmogorov Consistency/Existence Theorem not only guarantees the existence of a stochastic process when given a certain family of probability distributions following some specifications. It also hints at how we can construct a stochastic process.

Continuity of a Stochastic process

Let X (or $\{X_t : (\Omega, \mathfrak{F}, \mathcal{P}) \to (\tau, \mathbb{B}_\tau) | t \in T\}$) be a τ-valued stochastic process defined on a parameter space T with metric d. Since there are different modes of convergence, the continuity of the sample paths of a stochastic process can also be of different kinds, which are:

- **Almost Sure Continuity**: Given t $\in T$, X is said to be almost sure (a.s) continuous at t if P$\{\omega \in \Omega \mid \lim_{d(s,t)\to 0} \tau(X_s(\omega), X_t(\omega)) = 0\} = 1$.
- **Mean Square Continuity**: Given t $\in T$, X is said to be continuous in mean-square at t if E$|X_t|^2 < \infty$ and $\lim_{d(s,t)\to 0}$ E$((\tau(X_s, X_t))^2) = 0$.
- **Continuity in Probability**: Given t $\in T$, X is said to be continuous in probability at t if $\forall\ \epsilon > 0$; $\lim_{d(s,t)\to 0}$ P$\{\omega \in \Omega \mid \tau(X_s(\omega), X_t(\omega)) \ge \epsilon\} = 0$.
- **Continuity in Distribution**: Given t $\in T$, if $\lim_{d(s,t)\to 0} F_{X_s}(x) = F_{X_t}(x)$ at each continuity point x of F_{X_t}; then X is said to be continuous in distribution at t.

2.2 Gaussian Processes

So far, we have looked at some common results, basics and provided ourselves some motivation to study stochastic processes. Now we move to the main essence of this report - recent developments in the Probabilistic Analysis of general Gaussian Processes.

What makes a Gaussian Process and what roles do Gaussian Processes play in the science of today? What makes Gaussian Processes as important a family of stochastic processes as they happen to be? Let us quickly go through why amongst the several stochastic processes available, Gaussian processes are of great interest, before embarking on the actual journey this report shall take us through.

What is a Gaussian Process?

Definition 2.2.1. Let T be a non-empty index set on which a stochastic process $\{X_t\}_{t\in T}$ is defined. We say that this stochastic process is a **Gaussian Process** if for any choice of $t_1, t_2, ..., t_k \in T; \alpha_1, \alpha_2, ..., \alpha_k \in \mathbb{R}$ and $k \geq 1$, the random variable $\sum_{i=1}^{k} \alpha_i X_{t_i}$ has a univariate normal distribution.

For such processes, we define:

- The mean function, $\nu : T \to \mathbb{R}$ by $\nu(t) := EX_t$
- The Covariance function/kernel, $\sigma : T \times T \to \mathbb{R}$ by $\sigma(s,t) = \text{Cov}(X_s, X_t)$. Alternatively, we could define the Correlation function appropriately.

Remark 2.2.1. The Covariance Function is positive semi-definite.
For any k $\in \mathbb{N}$, $t_1, t_2, ..., t_k \in \Lambda$, and $\alpha_1, \alpha_2, ..., \alpha_k \in \mathbb{R}$;
$\sum_{i=1}^{k} \alpha_i \sigma(t_i, t_j) \alpha_j = \sum_{i=1}^{k} \alpha_i E X_{t_i} X_{t_j} \alpha_j = \text{Var}(\sum_{i=1}^{k} \alpha_i X_{t_i}) \geq 0$.

Remark 2.2.2. Note that given a real-valued function defined on T and a positive semi-definite real valued function on $T \times T$, then a Gaussian process can be completely defined.

Thus, the mean and covariance function determine a Gaussian process completely and uniquely up to distribution.

Construction of a Gaussian Process

- Let $\nu : T \to \mathbb{R}$ be an arbitrary function. Let $\sigma : T \times T \to \mathbb{R}$ be an arbitrary positive semi-definite function. Then, for each each k ≥ 1, and $t_1, t_2, ..., t_k \in T$, define $\mu_{t_1,t_2,...,t_k}$ as the k-variate normal probability distribution with mean $\vec{m}_k = (\nu(t_1), \nu(t_2), ..., \nu(t_k))^T$ and covariance matrix $\Sigma_k = ((\sigma(t_i, t_j))_{i,j=1}^k$. Let $\mathcal{Q}_T = \{\mu_{t_1,t_2,...,t_k} \mid t_1, t_2, ..., t_k \in T;\ \text{k} \in \mathbb{N}\}$.

- Given the fidis $\mathcal{Q}_T$, we have that for each k $\in \mathbb{N}$ and for any choice of $t_1, t_2, ..., t_k \in T$; $\mu_{t_1,t_2,...,t_k}$ is a valid probability distribution on $(\mathbb{R}^k, \mathfrak{B}_{\mathbb{R}}^k)$.

- The Consistency Condition C.1 required by Theorem 2.1.1 is satisfied by $\mathcal{Q}_T$. This can be verified by an application of the Fubini-Tonelli Theorem for iterated integration.

- The Consistency Condition C.2 required by Theorem 2.1.1 is satisfied by $\mathcal{Q}_T$. This is a consequence of the form of the probability distribution of a multivariate Gaussian random variable(vector) and is easy to verify.

- Then, from Theorem 2.1.1, we obtain a stochastic process X or $\{X_t | t \in T\}$ whose fidis is $\mathcal{Q}_T$.

The process X obtained is such that for any k $\in \mathbb{N}$ and $t_1, t_2, ..., t_k \in T$, $(X_{t_1}, X_{t_2}, ..., X_{t_k})$ is distributed as a k-variate normal random variable(vector). Hence, for any $\alpha_1, ..., \alpha_k \in \mathbb{R}$, $\sum_{i=1}^k \alpha_i X_{t_i}$ is distributed normally, ie, the process obtained is a *Gaussian process.*

Importance of Gaussian Processes

"Gaussian processes can be viewed as a far-reaching infinite-dimensional extension of classical normal random variables."[9] This allows one to build up on the properties of symmetry, existence of moments, and other consequences of the relatively more convenient form that the multivariate normal density function takes. The resulting features and characteristics of Gaussian processes enable their use in probabilistic modelling across fields of study like Statistics, Climate and Weather Forecasting, Financial Mathematics, Information Theory, and even the largest goldmine in a world taken over by *'ChatGPT', 'Dall-E'* and the likes - Machine Learning & Artificial Intelligence.

Regarded as the easiest to use in computation and derivation of explicit solutions to the most difficult problems in mathematics and beyond, Gaussian processes are heavily studied. Yet, they still provide vast scope for exploration. Even in the case of random fields - stochastic processes whose parameter space is a Euclidean Space/lattice - many problems are solvable only if the field generated is a Gaussian random field.[2] While this can again be attributed to the convenient form that the multivariate Gaussian density function takes, another particular fact also adds weight to the significance of Gaussian Processes. This being: *Two Gaussian random variables are independent if and only if their covariance is zero.* This makes much of the theory and analysis of Gaussian processes simpler and more interesting compared to other processes. One of the key mathematical properties of Gaussian processes is their probabilistic interpretation, which allows for uncertainty quantification and probabilistic predictions. In addition, Gaussian processes have a rich mathematical structure, including kernel functions that specify the covariance between points, and they can be used for regression, classification, and other tasks in statistics. Gaussian processes also have important implications for statistical inference, particularly in the context of Bayesian methods. They can be used as prior distributions over functions, and posterior distributions can be obtained using Bayes' rule. This allows for flexible and interpretable models that can be updated with new data. In addition, Gaussian processes can be used for model selection, as different kernel functions can be compared based on their fit to the data.

The mathematical analysis of Gaussian processes is an active area of research, with many open questions and challenges. Some of the topics of interest include the asymptotic behaviour of Gaussian process models as the number of data points increases, the generalisation of properties of Gaussian process models, and the development of efficient computational methods for inference and prediction. There are also connections between Gaussian processes and other areas of mathematics, such as functional analysis, and harmonic analysis. This report will take us through some such interesting and deeply implicating results.

Beyond the mathematics, Gaussian processes find much importance across scientific disciplines. Gaussian Processes especially make attractive models when the requirement is to optimise prediction under the constraint of low sample size, a constraint common to many investigations under all disciplines of science. Some interesting examples are described below:

- *Gaussian Processes in Astrophysics*: There are particular astrophysical investigations into 'Active Galactic Nuclei'[single: Active Galactic Nucleus (AGN)] which refer to the compact regions at the centres of galaxies that have been observed to be unusually luminous with unusual spectral properties that are uncharacteristic of ordinary stars. The emission spectra from these AGNs cover multiple wavelengths. Physicists are especially curious about the variability of optical and UV emissions from an AGN. To interpret the 'stochasticity' involved, stationary Gaussian process models with the auto-correlation function adhering to a 'damped random walk' model are employed. Gaussian models apparently fit well with the data obtained from telescopes for the study of emission fluctuations. A very well collected comprehensive summary of these applications are available in Dr. Rhan-Rhys Griffith's PhD Dissertation[6]. The Gaussian Process model has also proved effective in the time-series (time-indexed processes) analysis done to analyse stellar activity signals and to infer stellar rotation periods. In fact, using Gaussian processes to model has at times proved better than standard Fourier transform techniques that are used for appropriate purposes otherwise (Wilkins, 2019).

- *Gaussian Processes in Chemistry*: In this age of machine learning and pattern recognition, chemists too are looking at computational techniques to model and predict the chemical behaviour and properties of molecules, as well as the outcome of their reactions with other molecules. These techniques give a lot of valuable information on what molecular structures are more stable and so on, allowing prioritisation of different aspects during any laboratory synthesis of chemicals. Computational techniques also allow prediction of the yield of a chemical reaction, and thus can be used to investigate ways to optimise chemical reactions. Since laboratory synthesis of molecules and chemical reactions can be time and money intensive, these techniques help researchers enhance their efficiencies by many times. Quoting Sir David MacKay (MacKay (2003)), "Gaussian processes are useful tools for automated tasks where fine tuning for each problem is not possible". Software like GAUCHE have modules specifically catering to Gaussian processes in Chemistry. They help simulate frequent scenarios in experimental chemistry and facilitate molecule discovery & chemical reaction optimisation through suitable techniques to train Gaussian Process models and obtain desirable results.[5]

Takeaway

Stochastic processes are instrumental in several investigations aiming to understand evolution of a system over time or make a prediction/forecast. They find use across domains and disciplines of study. Stochastic processes can be viewed as *random functions* upon which probabilistic analysis can be done similar to the analysis of functions in general. Among stochastic processes, Gaussian processes are of particular significance. The peculiar characteristics of a Gaussian distribution, along with the tendency of most random phenomena in nature to follow/approximately follow a Gaussian distribution; extend value to the Gaussian process appropriately as well.

The remainder of this report will now take us across several recent developments in probabilistic analysis of Gaussian processes specifically.

2.3 The Diversity in Gaussian Processes

Gaussian Processes find themselves in many mathematical, statistical and real-life applications. Apart from the consequent results being useful, their analysis is also intriguing, evoking a lot of study and exploration into different kinds of Gaussian processes.

While the layman would only be familiar with the usual discrete-time indexed process, mathematicians have defined much more in an attempt to satiate their unbounded curiosity. The upward climb is natural. We first begin at the base, with a Gaussian process defined on $\mathbb{Z}$ or any other countable set, the simplest and most easily visualised family of Gaussian processes. Any person into economics/finance understands a time series reasonably well. A short climb up would eventually lead us to a 'height that still lets us breathe comfortably', the real-indexed Gaussian process. Real-indexed Gaussian processes can be used to represent the evolution of any natural system over a continuum. Now, we have advanced from $\mathbb{Z}$ to $\mathbb{R}$. Naturally, a climb along number systems leads us to a $\mathbb{C}$-indexed Gaussian process. This can now be regarded as a $\mathbb{R} \times \mathbb{R}$-indexed process instead. At this point, we have already entered the domain of *Random Fields* which are just stochastic processes whose parameter space is a Euclidean space/lattice. Here, we get to see processes with parameter spaces that could be a circle in $\mathbb{R}^2$ or some non-trivial subset of $\mathbb{R}^n$ where n can be any natural number. This is where the climb becomes steep; and interesting.

Consider a particular centred Gaussian random field, say $\{X_\alpha | \alpha \in \mathbb{R}^k\}$. Note that any Gaussian random variable is characterised by its mean and variance, implying that a centred Gaussian process is uniquely characterised by its covariance function. Let σ: $\mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}$ be the covariance function (as defined in 2.2) of the Gaussian random field considered. Now, consider a special class of real-valued functions defined on $\mathbb{R}^k$, $\mathcal{F}$. Define for each $\phi \in \mathcal{F}$, a random variable X(ϕ) $= \int_{\mathbb{R}^k} \phi(t)X(t)dt$. Then, since arbitrary sums of Gaussian random variables are Gaussian, and an integral is fundamentally the limit of a sum, each $X(\phi)$ is a newly defined Gaussian random variable. The stochastic process defined now by $\{X(\phi) | \phi \in \mathcal{F}\}$, whose parameter space is a family of $\mathbb{R}$-valued functions, is called a *Generalised Random Field*. These processes add a whole new level of complexity compared to what we began with, don't they? Further, anyone familiar with measure theory would be aware of the one-one correspondence with measures and distribution functions. This means we can define *measure-indexed processes*. Gaussian processes with measures as index parameters have not been studied much and provide great scope for exploration.[**?**]

When we arrived at random fields, we could have chosen paths that are altogether different from the one taken above. One such trail takes us to *set-indexed processes*. Set-indexed Gaussian processes are widely known and are of great research interest too. These include everything from Brownian Processes to biophysicists' favourite problem of a drunkard walking one step forward and two steps backward. Extending on Euclidean spaces, Gaussian processes may be defined on other Metric Spaces and Banach spaces too. And another interesting family is that of Manifold-indexed Gaussian processes, naturally extending from the Euclidean spaces once again in another direction. They are apparently a fad now in Machine Learning discussions where they are trying to interpret data as a manifold.

We began with a simple discrete time-indexed Gaussian process. However, our climb is unending. Of course, for every hard problem, a mathematician can construct a harder problem for the world to go crazy with. If we now consider any general Gaussian process defined on a parameter space, we now know that it belongs to one of the many families that can be defined. Depending on the parameter space, what kind of space it is, and so on, we can get different Gaussian processes with different properties, some with established applications and others of great potential use.

One natural worry then appears. How are we going to study such a vast number of different kinds of Gaussian processes?

2.4 One Theory to Rule Them All

As discussed previously, our problem is no longer the analysis of some family of Gaussian processes. The problem now is the analysis of ALL Gaussian processes.

AIM: To study the sample path properties of a general Gaussian process.

For a long time now, literature has treated the different kinds of Gaussian processes separately. Necessary and sufficient conditions for sample path continuity, etc. have been obtained separately for $\mathbb{R}$-indexed Gaussian processes, multi-parameter processes, function-indexed processes, set-indexed processes and so on. These results have been derived with great importance awarded to the geometry of the parameter space of the Gaussian processes that they were derived for and are applicable to. The focus on the finer aspects of the parameter space and using them directly with the definitions of continuity, etc. to derive results thus makes the study of general Gaussian processes rather difficult.

Difficulty: The difficulties are two-fold, one being that the huge number of Gaussian processes that have been/can be defined on different parameter spaces with different geometrical properties would then have to be tediously analysed one by one separately each time. Two, many processes defined on parameter spaces with particularly complicated geometric structures might prove daunting, if not entirely impossible at times.

Fortunately, probabilists in the late 1960s began to notice that the precise geometric structure of the parameter space on which a Gaussian process is defined, had very little to do with the sample path properties of said process[1]. Probabilists then began working on the possibility of characterising sample path properties of all Gaussian processes with a single/few results under very minimal assumptions, if required. This soon led to the development of two key concepts: **entropy** and **majorising measures**, which form the entire basis for a "Modern Theory" of sample path properties of general Gaussian processes. This Modern Theory focuses on bypassing the geometric structure and finer details of the parameter space by only dealing with the 'size' of the parameter space, measured in terms of a metric that is defined appropriately so that it is applicable on any general parameter space. The concepts of metric entropy and majorising measures help measure this size and use it to obtain results that characterise some sample path properties.

Solution: To develop a unifying theory based on the modern attitude described above, that applies to the analysis of *all kinds of Gaussian processes at once.*

Setting: From hereon, the below is fixed and shall be taken for granted as the background setting for all the discussions that follow, unless otherwise stated.

1. All random variables, sequences of random variables or Gaussian processes considered in this report hereafter are **real-valued** and defined on a **complete probability space**, $(\Omega, \mathfrak{F}, \mathcal{P})$ appropriately.

2. If X denotes a Gaussian process, X refers to the Gaussian process as defined in 2.2.

X is said to be defined on T, the parameter space.

3. **All Gaussian processes considered hereafter are assumed to be centred**, that is, the mean function associated with them is a zero map.

Remark 2.4.1. Results obtained for a centred Gaussian process can be easily extended to non-centred Gaussian processes by addition of a constant. Hence, analysis of a general Gaussian process under the assumption that it is centred is equivalent to analysing any general Gaussian process. The ease of dealing with centred processes is thus helpful.

Remark 2.4.2. Note that for a centred Gaussian process, the covariance function determines the process completely and uniquely up to distribution.

Result 2.4.1. The covariance function of a centred Gaussian process is a positive semi-definite function as discussed in 2.2.1. The converse is also true. For any given real-valued positive semi-definite function σ on $T \times T$, there exists a centred Gaussian process parameterised by elements of T whose covariance function is σ. This converse is a direct consequence of Theorem 2.1.1: the construction is the same as in Section 2.2 with ν taken to be the zero map.

4. All Gaussian processes considered hereafter are assumed to be defined on any general **metric space as its parameter space**. The parameter space is also assumed to be **totally bounded in its metric.**

Remark 2.4.3. In the next chapter, a metric that can induce a metric structure even in a parameter space that does not have its own inherent metric structure, will be defined. Thus, assuming a metric structure does not affect the quest for 'generality'. Starting with any general metric space is a natural way to go.

Totally boundedness will allow for 'uniformly continuous functions map totally bounded sets to totally bounded sets'-type and other similar arguments, keeping the difficulties of an unbounded process away.

5. Hereafter, the parameter space of the Gaussian processes considered are assumed to be separable (has a countable dense subset). That is, the processes considered are assumed to be **separable stochastic process**.

Remark 2.4.4. The separability of the parameter space makes proofs of results easier. This is because, we can extend proofs from a finite set to a countable set, and then to the whole set since separability ensure the denseness of at least one countable set.

Hereafter, the goal of this report is to introduce and use ’modern’ techniques that help determine/characterise sample path continuity and later relate sample path continuity with the boundedness of the supremum, of any general Gaussian process. While doing so, we will also touch upon some interesting information about the distribution of suprema of any general Gaussian process over a fixed subset of the parameter space.

The Modern Approach in the context of analysing sample path continuity of general Gaussian processes is introduced next. Part II will take the discussion further.

Part II

Preparing for a Modern Attack

Now that Part I has fixed the background setup and a broad goal - to analyse sample path continuity of general Gaussian processes using modern developments with a more efficient attitude, Part II introduces the basic ideas behind a 'General Theory'. As it does so, it first motivates the definition of the most fundamental tool in the entirety of the content in this report - the canonical metric.

The canonical metric serves as a hack that removes reliance on specific geometry of parameter spaces in an ingenious manner. This leads to the definition of entropy and the introduction to the Main Result of this Report. The Main Result uses an entropy-based argument to establish almost sure sample path continuity of a Gaussian process. This Main Result is then left for later, so as to first cover the pre-requisites for being able to appreciate and understand the Main Theorem sufficiently. Essentially, Part II fixes the specific, ultimate goal of this report in the context of appreciating the Modern Theory: to understand and appreciate the Main Result and its implications.

Part II then goes ahead to introduce concepts and results that are essential in Gaussian Process Theory. Whole doing so, Part II will introduce the 'Distribution of suprema' problem, and cover essential inequalities including Borell's and Slepian's that give a lot of information useful for the analysis of the supremum of a general Gaussian process. These inequalities specifically concern with two lines of thought: one trying to figure out how the supremum of a Gaussian process is distributed (asymptotically), and the other, to compare the suprema distributions of two Gaussian processes defined on the same parameter space.

At the end of Part II, a connection is made between the almost sure boundedness of the supremum of a Gaussian process and the almost sure continuity of its sample paths. Some interesting results about continuity are then stated without proof to provide motivation for what follows in Part III.

Chapter 3

To Determine Sample Path Continuity

Approach

Let X be a centred Gaussian process with a parameter space (T, τ), which may be any arbitrary metric space. We want to determine the a.s. continuity of X on T.

Define

$$\rho_\tau^2(u) = \underset{\{(s,t)\in T|\tau(s,t)\le u\}}{Sup} E(X_s - X_t)^2 \tag{3.1}$$

.
Now, if X were a.s continuous, then

$$P(\omega| \underset{\tau(s,t)\to 0}{lim} |X_s(\omega) - X_t(\omega)| = 0) = 1.$$

Let $\omega \in \Omega$ be such that X(ω) is continuous. Then clearly,

$$\underset{\{(s,t)\in T|\tau(s,t)\le u\}}{Sup} |X_s - X_t| \xrightarrow{\text{a.s.}} 0, \text{as u} \to 0.$$

Since expectation is the value of an integral, we can use Dominated Convergence Theorem with appropriate boundedness assumptions on the sample paths of X, say $\underset{t\in T}{Sup} X_t \le$ Y a.s for some random variable Y with $EY^2 < \infty$, to argue that

$$\rho_\tau^2(u) \to 0, \text{as u} \to 0 \tag{3.2}$$

. Note that the limit (3.2) basically says that X is mean square continuous. Essentially, this means that a.s. continuity implies mean square continuity. Since a.s continuity requires X to be mean square continuous, questioning the a.s continuity of X makes little sense unless *we accept that X is at least be mean square continuous on its parameter space.*

Let us then take X to be mean square continuous, ie, 3.2 holds true throughout this report hereafter. To obtain full, almost sure sample path continuity, something more is needed on ρ_τ^2.

Given: X is mean square continuous.
To prove: X is a.s continuous, ie,

$$Sup_{\{(s,t)\in T|\tau(s,t)\leq u\}}|X_s - X_t| \xrightarrow{\text{a.s.}} 0, \text{as u} \to 0.$$

Idea: For each $k \in \mathbb{N}$, let

$$E_n(k) = \{\omega \in \Omega \mid Sup_{\{(s,t)\in T|\tau(s,t)\leq \frac{1}{n}\}}|X_s(\omega) - X_t(\omega)| > \frac{1}{k}\}.$$

If we prove that for each $k \in \mathbb{N}$,

$$\sum_{n=1}^{\infty} P(E_n(k)) < \infty \overset{\text{Result 1.2.4}}{\Longrightarrow} P(E_n(k) \text{ infinitely often}) = 0,$$

then from Result 1.2.1, we have a.s continuity.

Now, note that for each $k, n \in \mathbb{N}$, using Markov's Inequality, after noting that $P(X > c) = P(X^2 > c^2)$ when X is a non-negative random variable and c is a positive constant, implies

$$P(E_n(k)) \leq \frac{E(Sup_{\tau(s,t)\leq\frac{1}{n}}|X_s - X_t|)^2}{\frac{1}{k}} \leq k^2\rho_\tau^2(\frac{1}{n})$$

. Therefore for each $k \in \mathbb{N}$, we can bound $\sum_{n=1}^{\infty} P(E_n(k))$ by bounding the series $\sum_{n=1}^{\infty} k^2\rho_\tau^2(\frac{1}{n})$. Whether the latter can be bounded or not depends on the rate of convergence of $\rho_\tau^2(u) \to 0$ as $u \to 0$.

Thus, to move from mean square continuity to a.s continuity, we need to additionally focus on the **rate of convergence** of $\rho_\tau^2(u) \to 0$ or equivalently $\rho_\tau(u) \to 0$ as u $\to 0$. We essentially need to find some condition that sufficiently increases the rate of convergence to imply a.s continuity of sample paths.

3.1 The Canonical (pseudo-)Metric

Definition 3.1.1. Define $d : T \times T \to \mathbb{R}_+ \cup \{0\}$ such that

$$d(s,t) = \sqrt{E(X_s - X_t)^2}.$$

This function naturally induces a metric structure on T and is called the **canonical metric** for T and/or X.

Note that the canonical metric is actually a pseudo-metric because it satisfies all requirements of a metric except that $d(s,t) = 0$ does not imply $s = t$. However, this fact is insignificant compared to the greater implications that are arrived at using this (pseudo-)metric.

Now, to use this canonical metric in place of τ, we must first ensure that a.s τ-continuity and d-continuity of Gaussian processes are equivalent. This is achieved from the below result.

Lemma 3.1.1. *Given a centred Gaussian process X under the assumptions in 2.4, X is mean square continuous* $\iff$ *the covariance function of* σ *is continuous.*

Proof.
- *Forward Direction*: X is mean square continuous

  $$\implies \forall \epsilon > 0, \exists\ \delta > 0 \text{ such that } Sup_{\tau(s,t)\leq u} E(X_s - X_t)^2 < \epsilon \text{ whenever } |u| < \delta.$$

  Take an arbitrary $\epsilon > 0$. Take any $s, t \in T$ such that $\tau(s,t) \leq u$ for some $u > 0$. Now, $\exists\ \delta > 0$ such that whenever $u < \delta$,

  $$E(X_s - X_t)^2 < Sup_{\tau(s,t)\leq u} E(X_s - X_t)^2 < \epsilon. \tag{3.1.1}$$

  Now,
  $$E(X_s - X_t)^2 = (\sigma(s,s) - \sigma(s,t)) + (\sigma(t,t) - \sigma(s,t)) \tag{3.1.2}$$

  W.L.O.G assume $\sigma(s,s) \leq \sigma(t,t)$. Let $u < \delta$. Using this assumption with inequality 3.1.1 and equation 3.1.2,

  $$\frac{-\epsilon}{2} < \sigma(s,s) - \sigma(s,t) < \frac{\epsilon}{2} \text{ and } \frac{-\epsilon}{2} < \sigma(s,s) - \sigma(t,s) < \frac{\epsilon}{2}.$$

  This proves that $\epsilon - \delta$ definition of continuity is satisfied for $\sigma(.\ ,s)$ and $\sigma(s,\ .)$ when any arbitrary $s \in T$ is fixed. Therefore, σ is continuous on $T \times T$.

- *Backward Direction*: σ is continuous on $T \times T$. Take any $s, t \in T$. Now, since σ is continuous, taking limit $\tau(s,t) \to 0$ on both sides of equation 3.1.2 gives

  $$lim_{\tau(s,t)\to 0} E(X_s - X_t)^2 = 0.$$

  Hence, follows.

□

Result 3.1.1. If X is a centred Gaussian process defined on a probability space and parameter space as assumed in 2.4 with the addition that T is complete, and X is mean-square continuous with respect to τ; then τ-continuity and d-continuity of X are equivalent.

Proof. Firstly, assume X is τ-continuous, ie,

$$lim_{u\to 0}\ (Sup_{\tau(s,t)\leq u}\ |X_s - X_t|) = 0$$

For $\eta > 0$, let $A_\eta = \{(s,t) \in T \times T \mid d(s,t) \leq \eta\}$. Now, σ is continuous. Redefine d in terms of σ using equation 3.1.2. Clearly then, d is itself τ-continuous and each A_η is the pre-image of a closed set $(-\infty, \eta]$ under a continuous map, ie, each A_η is τ-closed in the metric topology of $(T,\tau) \times (T,\tau)$. Hence, $\cap_{\{\eta>0\}} A_\eta = A_0$ is closed.

Take $\epsilon > 0$. Now, T is complete and totally bounded, implies T is compact. Consider the open cover of A_0,

$$\{U_{(s',t')} = \{(s,t) \in T \times T \mid max(\tau(s,s'),(t,t') < \epsilon\}\}_{(s',t') \in A_0}$$

Then, $T \times T$ is compact as well and A_0 being a closed subset of a compact space is compact. Therefore, $\exists$ finite set B $\subset A_0$ such that $\{U_{(s',t')}\}_{(s',t') \in \text{B}}$ covers A_0 and

$$\cup_{\{(s',t') \in B\}} \{(s,t) \in T \times T \mid max(\tau(s,s'), \tau(t,t')) \leq \epsilon\}$$

covers A_{η_ϵ} for some $\eta_\epsilon > 0$. Then, for any $(s,t) \in A_{\eta_\epsilon}$; $\exists$ $(s',t') \in$ B such that $\tau(s,s'), \tau(t,t') \leq \epsilon$. and,

$$|X_s - X_t| \leq |X_s - X_{s'}| + |X_{s'} - X_{t'}| + |X_{t'} - X_t| \;\; \forall \;\; s,t \in T. \tag{3.1.3}$$

Note that $(s',t') \in$ B $\subset A_0 \implies d(s',t') \leq \eta \forall \eta > 0 \implies d(s',t') = 0 \implies X_{s'} = X_{t'}$ a.s. (3.1.4)

(This justifies the subscript 0 given to A_0.) Then, taking $Sup_{d(s,t) \leq \eta_\epsilon}$ both sides of inequality 3.1.3 and using equality 3.1.4,

$$Sup_{d(s,t) \leq \eta_\epsilon} |X_s - X_t| \leq Sup_{(s,t) \in A_{\eta_\epsilon}} [Sup_{\tau(s,t) \leq \epsilon}(|X_s - X_{s'}| + |X_{t'} - X_t|)] \tag{3.1.5}$$

Now, as $\epsilon \to 0$, the inequality 3.1.5 and τ-continuity implies the left hand side converges to 0 because the right hand side does. Also, note that as $\epsilon \to 0$, we have that $\eta_\epsilon \to 0$.

Therefore, limit $\epsilon \to 0$ both sides of 3.1.5 gives the d-continuity of X.

To prove the converse, now assume that X is d-continuous. Take any $\epsilon > 0$, $\exists$ $\delta_\epsilon > 0$ such that $\text{d}(s,t) < \delta \implies |X_s - X_t|$ a.s. Since X is given to be mean square continuous, lemma 3.1.1 and the definition of d imply d is continuous in (T,τ). This means that for δ_ϵ, $\exists$ $\eta_{\delta_\epsilon} > 0$ such that $\tau(s,t) < \eta_{\delta_\epsilon} \implies d(s,t) < \delta_\epsilon$. Therefore, for $\epsilon > 0$, $\exists$ $\eta(\epsilon) = \eta_{\delta_\epsilon} > 0$ such that $\tau(s,t) < \eta(\epsilon) \implies |X_s - X_t| < \epsilon$ a.s. Hence proved. □

Remark 3.1.1. We can assume T to be complete and totally bounded, or simply put, τ-compact hereafter whenever required. This assumption is again minimal since a sufficiently large number of general Gaussian processes fit into this assumption. By virtue of Result 3.1.1 then, d-continuity and τ-continuity of a Gaussian process are equivalent. The canonical metric is also very easy to work with. It removes direct dependence on the intricate details of the geometry of the parameter space by varying through the covariance function of the process. Hence, we can work with the canonical metric.

Working with the canonical metric

Now let us consider $\rho_d(u)$, just as we defined $\rho_\tau(u)$ earlier. Then,

$$\rho_d^2(u) = Sup_{\{s,t\in T|\ d(s,t)\leq u\}}\ E(X_s - X_t)^2 = Sup_{\{s,t\in T|\ d(s,t)\leq u\}}\ d^2(s,t)$$

Suppose that $\forall\ u \leq$ diam$(T), \exists\ s,t \in T$ such that $d(s,t) = u$. Then, $\rho_d(u) = u$ for all $u \leq diam(T)$.

Clearly, ρ_d does not hold information about the Gaussian process. The information which was contained in ρ_τ has now moved into the canonical metric itself in some sense. Since the canonical metric is defined through the covariance function of the process, it is likely that the information is now contained in how the metric looks at T. Better put, the information about the Gaussian process seems to be recorded in how the canonical metric measures the sizes of T and its subsets. This is the essence behind the concept of 'Entropy' and the Main Result in this report, which are introduced next.

3.2 Entropy the Magnificent

To study the relationship between the canonical metric d and the parameter space T, that appears to be encoded in the manner that d measures 'sizes' of T and its subsets, we define the following:

Definition 3.2.1. Metric Entropy Function
Let $N(\epsilon)$ be the smallest number of d-closed balls, of radius ϵ required to cover T. Then, H(ϵ)=log($N(\epsilon)$) is called the metric entropy function for T or X.

New Metric, New Assumption: Updating 2.4 after the introduction of the canonical metric, it is hereafter assumed that T **is totally bounded in the Canonical Metric.** Basically, this report will take for granted that for all $\epsilon > 0, N(\epsilon) < \infty$. There is a result that states that a Gaussian process can have a.s continuity of sample paths and be a.s bounded iff T is totally bounded in the canonical metric, and another condition (which will be mentioned in Part III) is satisfied. Hence, by imposing this assumption, we are not restricting ourselves from any cases that cater to the goal of this Report mentioned in 2.4.

3.2.1 The Main Result - Entropy and Continuity

Theorem 3.2.1. *Let X be a centred Gaussian Process with parameter space, T equipped with the canonical metric d. Let T be totally bounded in the canonical metric. Then,*

$$\int_0^\infty [log(N(\epsilon))]^{\frac{1}{2}} d\epsilon < \infty \implies X \text{ is a.s continuous on } T.$$

Remark 3.2.1. Define $w(\delta) = \int_0^\delta (log(N(\epsilon)))^{\frac{1}{2}} d\epsilon$. Then, w actually serves as a modulus of continuity for X. This means that $|X_s - X_t| \leq w(d(s,t))$.

Remark 3.2.2. Note that if $\epsilon > \text{diam}(T)$, then $N(\epsilon) = 1$ and $\text{H}(\epsilon) = 0$. Therefore to satisfy the sufficient condition in Theorem 3.2.1, we only need to prove $\int_0^{\text{diam}(T)} (log(N(\epsilon)))^{\frac{1}{2}} d\epsilon < \infty$.

Remark 3.2.3. Note that as ϵ increases, $N(\epsilon)$ decreases by definition. Therefore, the finiteness of the integral in Theorem 3.2.1 is completely determined by the behaviour of the integrand near and at zero.

Proof of Theorem 3.2.1

There are different proofs for Theorem 3.2.1 and its earlier versions.[1]

This report will roughly cover one proof. In fact, proving this Result is the ultimate goal of this entire report. However, the proof requires us to look at several other results that relate supremum-based results and a.s sample path continuity of Gaussian processes, without which we cannot move forward.

Hence, Part II hereafter will cover discussions on some essential inequalities first. The proof of Theorem 3.2.1 will be discussed in Part III.

Examples, Applications

Applications and implications of this result for different kinds of processes will be covered in Part III.

Chapter 4

The Behaviour of Suprema

In the study of a function in real analysis, continuity, differentiability, etc are some important 'functionals' that we analyse. So, it does makes sense to analyse sample path continuity of general Gaussian processes. Continuity happens to be a basic property of a function. Several results exist on continuous functions that can help extract more information about other operations or 'functionals' of a process. And hence, it made sense to introduce the canonical metric, metric entropy function and the Theorem 3.2.1 stated earlier that gives a sufficient condition for a.s sample path continuity. However, why worry about the distribution of the supremum of a Gaussian process?

Motivation

To begin with, analysing the behaviour of $\underset{t \in T}{Sup} X_t$ for a centred Gaussian process X on parameter space T, can indicate boundedness/unboundedness of the process.

Definition 4.0.1. X is said to be a.s bounded if $P\{\omega | \underset{t \in T}{Sup} |X_t(\omega)| < \infty\} = 1$. If $P\{\omega | \, |\underset{t \in T}{Sup} |X_t(\omega)| < \infty\} = 0$, X is a.s unbounded.

If a process is unbounded a.s, it is easy to decipher that X is a.s discontinuous.

Apart from this, there are the commonly studied probabilistic questions of 'hitting times'.

Definition 4.0.2. For any $A \subset \mathbb{R}$, the hitting time of A is defined as the random variable defined by $\lambda_A = inf\{t \in T : X_t \in A\}$.

Note that if X is any stochastic process on a parameter space T, then $\inf_{t \in T} X_t = \sup_{t \in T} -X_t$. Essentially, the analysis of infima and suprema are equivalent. Then, to analyse λ_A for any $A \in \mathbb{R}$ we need to understand the behaviour of supremum(infimum) of a process. Analysis of hitting times are useful in the estimation of survival times - to estimate the lifetimes of an object or even a human patient (some of these examples may be better modelled using non-Gaussian distributions).

A seemingly better justification to why we should be interested in the suprema of Gaussian distributions in particular is the fact that some statistical tests and approximations rely on certain limit distributions being the supremum of a Gaussian process.

For example, consider the Kolmogorov-Smirnov Test that is commonly used in non-parametric inference. It is used to check whether an obtained sample is drawn from a certain probability distribution. Let $\{X_i\}_{i=1}^n$ be an iid sample of observations. Let $F_n(x) = \mathbb{1}_{(-\infty, X_i}(x)$ be the empirical cdf. Let F be some well-defined and appropriate cdf.

- *Null Hypothesis*: The sample is drawn from F(x).
- *Alternate Hypothesis*: The sample is not drawn from F(x).

Define $D_n(F) = \sup_x |F_n(x) - F(x)|$. The, there is a result called Glivenko-Cantelli Lemma which is stated below.

Result 4.0.1. $D_n(F) \xrightarrow{a.s} 0$ as $n \to \infty$ under the null hypothesis. (Glivenko-Cantelli Lemma)

This result does not help much in approximation of the unknown F however. However, there is another result, as stated below which helps approximate the cdf of a given sample under the null hypothesis.

Result 4.0.2. Consider the Gaussian White Noise process, $\{W_t | t \in [0, T]\}$ (8.1.1) defined on $([0, T], \mathfrak{B}_{[0,T]}, m)$ where $T > 0$, $\mathfrak{B}_{[0,T]}$ is the Borel σ-algebra on $[0, T]$, and m is the Lebesgue Measure defined appropriately. Then, define the process B on $[0, T]$ such that B(t) $= W_t - \frac{t}{T} W_T$. Such a process B is called the Brownian Bridge on [0,T].

Then under the null hypothesis taken above, $\sqrt{n} D_n \xrightarrow{d} \sup_t |B(F(t))|$ as $n \to \infty$, where $\{B(t) | t \geq 0\}$ is the Brownian bridge on [0,1]. If the hypothesised cdf F is continuous, then $\sqrt{n} D_n \xrightarrow{d} \sup_t |B(t)|$ as $n \to \infty$.

Clearly then, if F is continuous,

$$F(x) \approx F_n(x) \pm \frac{\underset{t}{sup}|B(t)|}{\sqrt{n}}$$

for each appropriate x.

Thus, the behaviour and distribution of the supremum of a Gaussian process is of utmost interest and makes supremum of a process an important functional to analyse.

4.1 Borell's Inequality

Consider $P(\lambda) = P\{\underset{t\in T}{sup} X_t \geq \lambda\}$. As far as existing literature is concerned, the problem of finding the distribution of $\underset{t\in T}{sup} X_t$, is an impossible one to solve. To give an idea, the precise formula for $P(\lambda)$ is known only for few specific stationary Gaussian processes identified by specific covariance functions, defined on a finite interval subset of $\mathbb{R}$.[1]

However, there are many results that shed light on the behaviour of $\underset{n\to\infty}{lim} P(\lambda)$. Roughly, this behaviour should intuitively depend on two factors:

1. Lack of Homogenity of X on T: If EX_t^2 is not constant across T, then the Supremum(or Infimum) is likely to be near the point of maximal variance.
2. Local smoothness or lack thereof: If X is rougher near a point of maximal variance, it is likely that the Supremum/Infimum is larger/smaller.

The rigorous characterisation of these ideas is achievable by suitably bounding $P(\lambda)$. The theorem below gives such a bound.

Theorem 4.1.1. *(Borell's Inequality)*
Let X be a centred Gaussian process with sample paths bounded a.s. Let $\|X\| = \underset{t\in T}{Sup} X_t$, *and* $\sigma_T^2 = \underset{t\in T}{sup} EX_t^2$. *Then,* $\mathbb{E}\|X\| < \infty$ *and for all* $\lambda > 0$;

$$P\{|\ \|X\| - E\|X\|\ | > \lambda\} \leq 2e^{\frac{-\lambda^2}{2\sigma_T^2}}. \tag{4.1.1}$$

Hence, $\forall \lambda > E\|X\|$;

$$P\{\|X\| > \lambda\} \leq 2e^{\frac{-(\lambda - E\|X\|)^2}{2\sigma_T^2}}. \tag{4.1.2}$$

Note: The expectation of the supremum is no longer centred in general. Hence, it is natural to want to prove that the expectation is atleast finite before proving the inequality itself. In fact, the use of Borell's inequality itself requires information on the expectation of supremum. In a more general version of this theorem, the theorem also concludes that $\sigma_T^2 < \infty$. However, here we choose to take it as an extraneous assumption since this is sufficient for the discussion here and later in this report.

Proof of Borell's Inequality - I

To prove Theorem 4.1.1, we will need the following lemmas. Before that, let us look at a growth condition that allows the use fo Result 1.2.5 appropriately and prove the result that follows.

Definition 4.1.1 (Subgaussian Growth)**.** A function f : $\mathbb{R}^k \to \mathbb{R}$ is said to have subgaussian growth if for each $\epsilon > 0, \exists\ C(\epsilon) > 0$ such that $|f(x)| \leq C(\epsilon)\exp(-\epsilon\|x\|^2_{\mathbb{R}^k})\ \forall\ x \in \mathbb{R}^k$.

The following result is very useful in proving many results involving Gaussian processes.

Result 4.1.1 (Gaussian Integration by Parts)**.** If X is a centred Gaussian random vector in $\mathbb{R}^n$, and f is a continuously differentiable real-valued function on $\mathbb{R}^n$ such that ∇f has subgaussian growth; then $E(X_i f(X)) = \sum_{i=1}^{n} E(X_i X_j) E\big[\frac{\partial f}{\partial x_j}(X)\big]$.

Proof. • Let n = 1. Note that p'(x) = $\frac{-1}{EX^2} x p(x)$ where $p(x)$ is the pdf of X. Then, use product rule to find that

$$E(Xf(X)) = (-\sigma^2)\int_{-\infty}^{\infty} (f(x)p(x))'\,dx + \sigma^2 E(f'(X)).$$

- Claim: The first term in the above equation is zero. Now, since n = 1, ∇f is the same as f', and has subgaussian growth. Then, for a fixed $\epsilon > 0$, $\exists\ C(\epsilon^2)$ such that;

$$-C(\epsilon^2)\exp(-\epsilon^2|x|^2) \le f'(x) \le C(\epsilon^2)\exp(-\epsilon^2|x|^2)$$

Integrating on [-a,a] for some a > 0;

$$-C(\epsilon^2)\int_{-a}^{a}\exp(-\epsilon^2|x|^2)\ dx \le f(a) - f(-a) \le C(\epsilon^2)\int_{-a}^{a}\exp(-\epsilon^2|x|^2)\ dx$$

$$\implies -2C(\epsilon^2)\int_{0}^{a}\exp(-\epsilon^2|x|^2)\ dx \le f(a) - f(-a) \le 2C(\epsilon^2)\int_{0}^{a}\exp(-\epsilon^2|x|^2)\ dx$$

Note that $p(a) = p(-a)$ since $X \sim N(0, EX^2)$ and $p(a) = \dfrac{1}{\sqrt{2\pi EX^2}}\ \exp(\dfrac{-x^2}{2EX^2})$;

$$\implies -2C(\epsilon^2)\left(\frac{e^{\frac{-x^2}{2EX^2}}}{\sqrt{2\pi EX^2}}\right)\int_{0}^{a} e^{-\epsilon^2|x|^2}\ dx \le f(a)p(a) - f(-a)p(-a)$$

$$\le 2C(\epsilon^2)\left(\frac{e^{\frac{-x^2}{2EX^2}}}{\sqrt{2\pi EX^2}}\right)\int_{0}^{a} e^{-\epsilon^2|x|^2}\ dx$$

Applying $lim_{a\to\infty}$; we get $lim_{a\to\infty}\int_{-a}^{a}(f(x)p(x))'\ dx = 0$. Therefore when n = 1, $E(Xf(X)) = EX^2E(f'(X))$.

- Now, fix i. Define Z_j's such that $X_j = \frac{EX_iX_j}{EX_i^2}X_i + Z_j\ \forall j \neq i$. Then, $Z \equiv (Z_1, ..., Z_n)$ is clearly independent of X_i by definition and $f(X)$ can be written as some function of $X_i, Z_1, ..., Z_{i-1}, Z_{i+1}, ..., Z_n$; say $g(X_i, Z_1, ..., Z_n)$. Then,

$$E[X_if(X)] = E[X_ig(X_i, Z_1, ..., Z_n)] = E_Z\big[E_{X_i|Z}[X_ig(X_i, Z)\ |Z]\big].$$

- Note that the conditional expectation $E_{X_i|Z}[X_ig(X_i, Z)\ |Z]$ deals with a 1-variable case, and hence we can use the result obtained for the n = 1 case above.

$$\begin{aligned}
\implies E[X_if(X)] &= E_Z\left[(X_i^2)E_{X_i|Z}\left[\frac{\partial g}{\partial x_i}(X_i, Z)\right]\right]\\
&= (EX_i^2)E_Z\left[E_{X_i|Z}\left[\frac{\partial g}{\partial x_i}(X_i, Z)\right]\right]\\
&= (EX_i^2)E_{X_i}\left[\frac{\partial g}{\partial x_i}(X_i, Z)\right]\\
&= (EX_i^2)E\left[\frac{\partial f}{\partial x_i}(X_1, ..., X_n)\right]\\
&= (EX_i^2)E\left[\sum_{j=1}^{n}\frac{\partial f}{\partial x_j}(X)\frac{\partial}{\partial x_i}\Big(\frac{EX_iX_j}{EX_i^2}X_i + Z_j\Big)\right]\\
&= E\left[\sum_{j=1}^{n}E(X_iX_j)E\left[\frac{\partial f}{\partial x_j}(X)\right]\right].
\end{aligned}$$

Hence Proved.

□

Corollary 4.1.2. *Let X, Y be two independent k-dimensional centred Gaussian vectors, each with covariance matrix equal to the appropriate Identity matrix. Let* $f, g \in C^1(\mathbb{R}^k)$ *be two bounded functions such that their partial derivatives of up-to 2nd order have subgaussian growth. Then,*

$$Cov(f(X), g(X)) = \int_0^1 E(\nabla f(X).\nabla g(\alpha X + \sqrt{1-\alpha^2}Y))d\alpha.$$

where $\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(X) \\ \vdots \\ \frac{\partial f}{\partial x_k}(X) \end{pmatrix}$.

Proof. Interpolate between X and Y through $\theta \in [0, \frac{\pi}{2}]$ as $Z(\theta) = (cos\theta)X + (sin\theta)Y$. Let $\Psi(\theta) = E[f(X)g(Z(\theta))]$. Now, f,g are bounded. Then, $\exists M > 0$ such that for any $h > 0, x, y \in \mathbb{R}^k, \theta \in [0, \frac{\pi}{2}]$,

$$\left| f(x)\frac{1}{h}(g(z(\theta+h)) - g(z(\theta)) \right| \leq M;$$

where $z(\theta) = (cos\theta)x + (sin\theta)y$. This means the limit as $h \to 0$ on the above inequality exists, and

$$\begin{aligned}
&\lim_{h\to 0}\left| f(x)\frac{1}{h}(g(z(\theta+h)) - g(z(\theta)) \right| \\
&= \lim_{h\to 0}\left| f(x)\frac{\partial}{\partial h}(g(cos(\theta+h)x + sin(\theta+h)y)) \right| \\
&\leq \lim_{h\to 0}\left\{ \left| f(x)\sum_{i=1}^{k}(-sin(\theta+h))\frac{\partial}{\partial x_i}(g(cos(\theta+h)x + sin(\theta+h)y)) \right| \right. \\
&\quad \left. + \left| f(x)\sum_{i=1}^{k}(cos(\theta+h))\frac{\partial}{\partial y_i}(g(cos(\theta+h)x + sin(\theta+h)y)) \right| \right\} \\
&= M\left\{ \left| f(x)\sum_{i=1}^{k}(-sin\theta)\frac{\partial}{\partial x_i}(g(cos\theta x + sin\theta y)) \right| \right. \\
&\quad \left. + \left| f(x)\sum_{i=1}^{k}(cos\theta)\frac{\partial}{\partial y_i}(g(cos\theta x + sin\theta y)) \right| \right\}.
\end{aligned}$$

The last expression is clearly bounded by an integrable function owing to the subgaussian growth conditions assumed of f, g. Hence, by Result 1.2.5,

$$\Psi'(\theta) = E\left[f(X)\frac{d}{d\theta}(g(Z(\theta)))\right] = \sum_{j=1}^{k}E\left[f(X)\left[\frac{\partial g}{\partial z_j}(Z(\theta))(-sin\theta X_j + cos\theta Y_j)\right]\right]$$

$$= \sum_{j=1}^{k}\left\{E_Y\left\{(-sin\theta)E_{X|Y}\left[f(X)\frac{\partial g}{\partial z_j}(cos\theta X + sin\theta Y)X_j\right]\right\} + E\left\{(cos\theta)Y_j f(X)\frac{\partial g}{\partial z_j}(Z(\theta))\right\}\right\}$$

Using Gaussian Integration by parts on both terms separately, and noting the independence of X_i, X_j, Y_i, Y_j $\forall i \neq j$ and

$$= \sum_{j=1}^{k} \Bigg\{ E_Y \Bigg\{ (-sin\theta) \Bigg[\sum_{i=1}^{k} (EX_iX_j) E_{X|Y} \Bigg(\frac{\partial}{\partial x_i} (f(X)) \frac{\partial g}{\partial z_j} (Z(\theta)) \Bigg) \Bigg] \Bigg\}$$

$$+ cos\theta E_X \Bigg[f(X) E_{Y|X} \Bigg[Y_j \frac{\partial g}{\partial z_j} (cos\theta X + sin\theta Y) \Bigg] \Bigg] \Bigg\}$$

$$= \sum_{j=1}^{k} \Bigg\{ (-sin\theta) (\sum_{i=1}^{k} EX_iX_j) \Bigg[E \Bigg(\frac{\partial f}{\partial x_i} (X) \frac{\partial g}{\partial z_j} (Z(\theta)) \Bigg) + E \Bigg((cos\theta) f(X) \frac{\partial^2 g}{\partial z_i z_j} (Z(\theta)) \Bigg) \Bigg]$$

$$+ (cos\theta) f(X) (\sum_{i=1}^{k} EY_iY_j) E \Bigg[(sin\theta) \frac{\partial^2 g}{\partial z_i \partial z_j} (Z(\theta)) \Bigg] \Bigg\}$$

$$= \sum_{j=1}^{k} \Bigg\{ (-sin\theta) E \Bigg[\frac{\partial f}{\partial x_j} (X) \frac{\partial g}{\partial z_j} (Z(\theta)) \Bigg] - (sin\theta cos\theta) E \Bigg[f(X) \frac{\partial^2 g}{\partial z_j^2} (Z(\theta)) \Bigg]$$

$$+ (sin\theta cos\theta) E \Bigg[f(X) \frac{\partial^2 g}{\partial z_j^2} (Z(\theta)) \Bigg] \Bigg\} = (-sin\theta) E\{\nabla f(X).\nabla g(Z(\theta))\}.$$

Now, a simple application of the Fundamental Theorem of Calculus leads to the result as follows:

$$\begin{aligned}
\mathrm{Cov}(f(X), g(X)) &= E[f(X)g(X)] - E[f(X)]E[g(X)] \\
&= E[f(X)g(X)] - E[f(X)]E[g(Y)] = E[f(X)g(Z(0)] - E[f(X)g(Z(\pi/2))] \\
&= -(\Psi(\pi/2) - \Psi(0)) = -\int_0^{\frac{\pi}{2}} \Psi'(\theta) d\theta = -\int_0^{\frac{\pi}{2}} (-sin\theta) E\{\nabla f(X).\nabla g(Z(\theta))\} d\theta \\
&= \int_0^1 E\{\nabla f(X).\nabla g(\alpha X + \sqrt{1-\alpha^2} Y)\} d\alpha
\end{aligned}$$

where the last line follows from a change of variables, from $cos\theta$ to α. □

Now, we prove the lemma that is key to proving Theorem 4.1.1 below:

Lemma 4.1.3. *Let X be a k-dimensional random vector of centred, unit-variance, independent Gaussian random variables. If $f : \mathbb{R}^k \to \mathbb{R}$ is Lipschitz continuous with Lipschitz constant $\sigma \in (0, \infty)$, then for all $\lambda > 0$,*

$$P\{f(X) - Ef(X) > \lambda\} \leq e^{\frac{-\lambda^2}{2\sigma^2}}.$$

Proof. Take f with the assumptions mentioned in the Lemma. Since f is continuous, it can be approximated by a sequence of polynomials, each defined on a compact support. Hence, it is sufficient to prove the lemma for some function, $f \in C^1(\mathbb{R}^k)$, having a compact support, with ∇f having Euclidean norm at most σ, the Lipschitz Constant.

Considering such a function f, by exponential Chebyshev's Inequality, we have that for any t ≥ 0

$$P(f(X) - Ef(X)) > \lambda) \leq e^{-t\lambda} E(e^{t(f(X)-Ef(X))}).$$

Take $g(x) := e^{tf(x)}$ for some $t \geq 0$. Then the partial derivatives of f and g upto 2nd order have subgaussian growth since their supports are compact by assumption. Applying the previous corollary,

$$\begin{aligned}
Cov(f(X), g(X)) &= \int_0^1 tE(e^{tf(Z_\alpha)} \nabla f(X).\nabla f(Z_\alpha))d\alpha \\
&= \int_0^1 tE(e^{tf(Z_\alpha)} max\{\|\nabla f(X)\|, \|\nabla f(Z_\alpha)\|\})d\alpha \\
&\leq t\sigma^2 \int_0^1 E(e^{tf(Z_\alpha)})d\alpha = t\sigma^2 \int_0^1 E(e^{tf(X)})d\alpha = t\sigma^2 E(e^{tf(X)}).
\end{aligned}$$

Let $h(t) = E(e^{t(f(X)-Ef(X))})$. Then, by DCT,

$$h'(t) = E[\frac{d}{dt}(e^{t(f(X)-Ef(X))})] = \frac{1}{e^{tEf(X)}} Cov(f(X), g(X)) \leq t\sigma^2 E(e^{t(f(X)-Ef(X))}) = t\sigma^2 h(t).$$

$$\implies h'(t) \leq t\sigma^2 h(t) \implies \int \frac{h'(t)}{h(t)} \leq \int t\sigma^2 + C$$

Since $h(0) = 1$, we have

$$log(h(t)) \leq \frac{1}{2}t^2\sigma^2 \implies E(e^{t(f(X)-Ef(X))}) \leq e^{\frac{1}{2}t^2\sigma^2}.$$

Then,

$$P(f(X) - Ef(X)) > \lambda) \leq e^{-t\lambda} e^{\frac{1}{2}t^2\sigma^2}, t \geq 0.$$

Optimizing this upper bound over $t \geq 0$, we get the minimum upper bound for $t = \frac{\lambda}{\sigma^2}$. Substituting the same, we obtain the desired result:

$$P\{f(X) - Ef(X) > \lambda\} \leq e^{\frac{-\lambda^2}{2\sigma^2}}.$$

□

Borell's Inequality Proof - II

Proof of Theorem 4.1.1 for finite T.

- Assume X is an a.s bounded centred Gaussian process on a finite parameter space T. Let T = $\{t_1, t_2, ..., t_k\}$. It is then trivial that $\sigma_T^2 = max(EX_{t_1}^2, EX_{t_2}^2, ..., EX_{t_k}^2) < \infty$. On the other hand, a bound on the expectation of the maximum order statistic can be obtained as follows:

$$\begin{aligned}
e^{\frac{\sqrt{logk}}{\sigma_T}E\|X\|} &\underset{inequality}{\overset{Jensen's}{\leq}} Ee^{\frac{\sqrt{logk}}{\sigma_T}\|X\|} = E\|e^{\frac{\sqrt{logk}}{\sigma_T}X}\| \\
&\underset{of\ max}{\overset{Property}{\leq}} \sum_{i \in T} E[e^{\frac{\sqrt{logk}}{\sigma_T}X_i}] \\
&\overset{M.G.F}{=} \sum_{i \in T} e^{\frac{1}{2}EX_i^2 \frac{logk}{\sigma_T^2}} \leq ke^{\frac{1}{2}logk}
\end{aligned}$$

where the last equality is from the definition of σ_T^2. Then, $E\|X\| \leq \sigma_T \frac{3}{2}\sqrt{logk} < \infty$.

- Now, we have to prove that (4.1.1) holds. Say $X_{t_1},...,X_{t_k}$ are independent (if they are not already independent, take copies that are independent). Let $Y_i = \frac{1}{\sqrt{EX_i^2}} X_i \forall i \in T$. Now, for any $x, y \in \mathbb{R}^k$, W.L.O.G. say $max(x) = x_i \geq max(y) = y_j$ (for some i, j $\in T$). Then,

$$|max(x) - max(y)| = |x_i - y_j| \leq max(x-y) \leq \sqrt{\sum_{i=1}^{k} |x_i - y_i|^2} = ||x-y||_{\mathbb{R}^k}.$$

Also for any $x, y \in \mathbb{R}^k$,

$$min(x) = -max(-x) \implies |min(x) - min(y)| = |-max(-x) + max(-y)| \leq ||-y+x||_{\mathbb{R}^k} = ||x-y||_{\mathbb{R}^k}$$

And $-Y_i = -\frac{1}{\sqrt{EX_i^2}} X_i \forall i \in T$.

- For any $\lambda > 0$ then, by Lemma 4.1.3,

$$P\{\max_{i\in T}(Y_i) - E\max_{i\in T}(Y_i) > \frac{\lambda}{\sigma_T}\}, P\{\min_{i\in T}(-(Y_i)) - E\min_{i\in T}(-(Y_i)) > \frac{\lambda}{\sigma_T}\} \leq e^{\frac{-\lambda^2}{2\sigma_T^2}} \quad (4.1.3)$$

. Now, note that

$$\sigma_T\{\max_{i\in T}(\frac{1}{\sqrt{EX_i^2}}X_i) - E\max_{i\in T}(\frac{1}{\sqrt{EX_i^2}}X_i)\} = \{\max_{i\in T}(\frac{\sigma_T}{\sqrt{EX_i^2}}X_i) - E\max_{i\in T}(\frac{\sigma_T}{\sqrt{EX_i^2}}X_i)\}$$
$$\geq \{\max_{i\in T}(X_i) - E\max_{i\in T}(X_i)\} \quad (4.1.4)$$

and

$$\sigma_T\{\min_{i\in T}(\frac{-1}{\sqrt{EX_i^2}}X_i) - E\min_{i\in T}(\frac{-1}{\sqrt{EX_i^2}}X_i)\} = \{\min_{i\in T}(\frac{-\sigma_T}{\sqrt{EX_i^2}}X_i) - E\min_{i\in T}(\frac{-\sigma_T}{\sqrt{EX_i^2}}X_i)\}$$
$$\geq \{\min_{i\in T}(-X_i) - E\min_{i\in T}(-X_i)\} \quad (4.1.5)$$

where the first equality and subsequent inequality in both (4.1.4) and (4.1.5) follow because $\sigma_T \geq \sqrt{EX_i^2} > 0 \ \forall i \in T$ by definition.

- Then by (4.1.3),(4.1.4),

$$P\{\max_{i\in T}(X_i) - E\max_{i\in T}(X_i) > \lambda\} \leq P\{\sigma_T(\max_{i\in T}(Y_i) - E\max_{i\in T}(Y_i)) > \lambda\} \leq e^{\frac{-\lambda^2}{2\sigma_T^2}}$$

and similarly by (4.1.3),(4.1.5),

$$P\{\min_{i\in T}(-X_i) - E\min_{i\in T}(-X_i) > \lambda\} \leq P\{\sigma_T(\min_{i\in T}(-Y_i) - E\min_{i\in T}(-Y_i)) > \lambda\} \leq e^{\frac{-\lambda^2}{2\sigma_T^2}}.$$

Then,

$$P\{|\ ||X|| - E||X||\ | > \lambda\} = P\{\ ||X|| - E||X||\ > \lambda\} + P\{-||X|| - E(-||X||)\ > \lambda\}$$
$$= P\{\ ||X|| - E||X||\ > \lambda\} + P\{\min_{i\in T}(-X_i) - E\min_{i\in T}(-X_i)\ > \lambda\} \leq 2e^{-\frac{\lambda^2}{\sigma_T^2}}$$

- Hence proved (4.1.1) holds for finite T.

- For any $\lambda > E\|X\|$,

$$P\{\|X\| > \lambda\} = P\{\|X\| > \lambda - E\|X\| + E\|X\|\} = P\{\|X\| > \lambda_0 + E\|X\|\} + P\{\|X\| < E\|X\| - \lambda_0\}$$
$$= P\{|\ \|X\| - E\|X\|\ | > \lambda_0\} \leq 2e^{-\frac{\lambda_0^2}{\sigma_T^2}} = 2e^{-\frac{(\lambda - E\|X\|)^2}{\sigma_T^2}}$$

where $\lambda_0 = (\lambda - E\|X\|) > 0$. and the last inequality follows from (4.1.1). Hence proved (4.1.2) holds for finite T.

□

Borell's Inequality Proof - III

Proof of Theorem 4.1.1 in general.

- Assume X is an a.s bounded centred Gaussian process on a parameter space T with assumptions taken since 2.4. T is separable $\implies \exists$ countable dense subset $T' \subset T$. Let $T' = \{t_1, t_2, ...\}$ and $T_n = \{t_1, ..., t_n\}$ for each $n \in \mathbb{N}$.

- Claim 1: $\sup_{t\in T_n} X_t \xrightarrow{a.s} \sup_{t \in T'} X_t$ as $n \to \infty$.
For each $\epsilon > 0$; by definition of supremum, $\exists\ i \in \mathbb{N}$ such that

$$X_{t_i} > \sup_{t\in T'} X_t - \epsilon \implies \sup_{t\in T_n} X_{t_i} > \sup_{t\in T'} X_t - \epsilon \text{ a.s } \forall n \geq i.$$

This proves Claim 1.

- Now, fix $\epsilon > 0$. By definition of supremum and dense property of T', $\exists\ t_0 \in T'$ such that $X_{t_0} > \|X\| - \epsilon$ a.s. Then, $\sup_{t\in T'} X_t \geq \|X\| - \epsilon$ a.s. $\implies 0 \leq \|X\| - \sup_{t\in T'} X_t \leq \epsilon$ a.s.
Taking $\epsilon \downarrow 0, \|X\| = \sup_{t\in T'} X_t$ a.s.
Claim 1 then implies that $\sup_{t\in T_n} X_t \xrightarrow{a.s} \sup_{t\in T} X_t = \|X\|$ as $n \to \infty$ and consequently by Theorem 1.2.4 (since $\{\sup_{t\in T_n} X_t\}_{n\in\mathbb{N}}$, $\{\sup_{t\in T_n} EX_t^2\}_{n\in\mathbb{N}}$ are increasing sequences by definition of T_n, T'), $E\sup_{t\in T_n} X_t \to E\|X\|$ and $\sup_{t\in T_n} EX_t^2 \to \sigma_T^2$ as $n \to \infty$.

- $\sigma_T^2 < \infty$ is taken for granted as noted earlier.

- Claim 2: $E\|X\| < \infty$
Assume $E\|X\| = \infty$. Choose a $\lambda_0 > 0$ large enough such that P$\{\|X\| < \lambda_0\} \geq \frac{3}{4}$ and $e^{\frac{-\lambda_0^2}{\sigma_T^2}} \leq \frac{1}{4}$.
Now, $E\sup_{t\in T_n} X_t \to E\|X\|$ as $n \to \infty \implies \exists N \in \mathbb{N}$ such that $\forall n \geq N, E\sup_{t\in T_n} X_t > 2\lambda_0$.

Then, we can show that $P\{|\underset{t\epsilon T_n}{sup}X_t - E\underset{t\epsilon T_n}{sup}X_t| > \lambda_0\} \leq 2e^{\frac{-\lambda^2}{2\sigma_T^2}} \leq \frac{1}{2}$ and;

$$P\{|\underset{t\epsilon T_n}{sup}X_t - E\underset{t\epsilon T_n}{sup}X_t| > \lambda_0\} \geq P\{E\underset{t\epsilon T_n}{sup}X_t - \underset{t\epsilon T_n}{sup}X_t > \lambda_0\} \geq$$
$$P(\|X\| < E\underset{t\epsilon T_n}{sup}X_t) - \lambda_0) \geq P(\|X\| < \lambda_0) \geq \frac{3}{4}, (\rightarrow\leftarrow)$$

$\implies E\|X\| < \infty$.

- Now, we have to prove that (4.1.1) holds. Define $K_n = |\underset{t\epsilon T_n}{sup}X_t - E\underset{t\epsilon T_n}{sup}X_t|$ and $K = | \|X\| - E\|X\| |$. From Claim 1 and it's consequences, $K_n \overset{a.s}{\to} K$ as $n \to \infty$. By Fatou's Lemma, for any $\lambda > 0$;

$$P\{\omega|K(\omega) > \lambda\} = \int_\Omega \mathbb{1}\{K(\omega) > \lambda\}d\mathcal{P} \leq \underset{n\to\infty}{liminf} \int_\Omega \mathbb{1}\{K_n(\omega) > \lambda\}d\mathcal{P}$$
$$\leq \underset{n\to\infty}{liminf} P(\omega|K_n(\omega) > \lambda) \leq 2e^{\frac{-\lambda^2}{2\sigma_T^2}}$$

where the last inequality follows from Borell's inequality for finite parameter space proved earlier. Hence proved that (4.1.1) holds for T.

- For any $\lambda > E\|X\|$,

$$P\{\|X\| > \lambda\} = P\{\|X\| > \lambda - E\|X\| + E\|X\|\} = P\{\|X\| > \lambda_0 + E\|X\|\} + P\{\|X\| < E\|X\| - \lambda_0\}$$
$$= P\{| \|X\| - E\|X\| | > \lambda_0\} \leq 2e^{-\frac{\lambda_0^2}{\sigma_T^2}} = 2e^{-\frac{(\lambda - E\|X\|)^2}{\sigma_T^2}}$$

where $\lambda_0 = (\lambda - E\|X\|) > 0$. and the last inequality follows from (4.1.1). Hence proved (4.1.2) holds for general separable T.

□

4.2 Distribution of $\underset{t\epsilon T}{Sup}X_t$

Now that Theorem 4.1.1 has been proved, let us make a few observations and remarks.

Take (4.1.2). Apply log on both sides. Divide by λ^2 and apply $\lim_{\lambda\to\infty}$ both sides to get the following result due to Landau and Shepp(1970) and Marcus and Shepp(1971):

Result 4.2.1. If X is a centred Gaussian process with sample paths bounded a.s, then

$$\underset{\lambda\to\infty}{lim} \frac{log(P\{\underset{t\epsilon T}{sup}X_t > \lambda\})}{\lambda^2} = \frac{-1}{2\sigma_T^2} \text{ where } \sigma_T^2 = \underset{t\epsilon T}{sup}EX_t^2 \quad (4.2.1)$$

Note: The original proof of this result may not have been done using Borell's inequality/Theorem 4.1.1.

Now, for any $X \sim N(0, \sigma^2)$, the following can be proven simply by a clever use integration by parts. If X is a centered Gaussian rv with variance σ^2, a clever integration by parts yields

$$(1 - \frac{\sigma^2}{\lambda^2})(\frac{\sigma}{\sqrt{2\pi}})\frac{1}{\lambda}e^{-\frac{\lambda^2}{2\sigma^2}} \leq P\{X > \lambda\} \leq (\frac{\sigma}{\sqrt{2\pi}})\frac{1}{\lambda}e^{-\frac{\lambda^2}{2\sigma^2}} \tag{4.2.2}$$

Quick Proof.

$$\frac{1}{\sigma\sqrt{2\pi}}\int_\lambda^\infty \frac{1}{x}xe^{\frac{-x^2}{2\sigma^2}}dx = (\frac{\sigma}{\sqrt{2\pi}})(\frac{1}{\lambda})e^{\frac{-\lambda^2}{2\sigma^2}} - \frac{\sigma}{\sqrt{2\pi}}\int_\lambda^\infty \frac{1}{x^3}xe^{\frac{-x^2}{2\sigma^2}}dx \tag{4.2.3}$$

$$= (\frac{\sigma}{\sqrt{2\pi}})(\frac{1}{\lambda})e^{\frac{-\lambda^2}{2\sigma^2}} - (\frac{\sigma}{\sqrt{2\pi}})(\frac{\sigma^2}{\lambda^3})e^{\frac{-\lambda^2}{2\sigma^2}} + \frac{3\sigma^3}{\sqrt{2\pi}}\int_\lambda^\infty \frac{1}{x^4}e^{\frac{-x^2}{2\sigma^2}}dx$$

$$= (1 - \frac{\sigma^2}{\lambda^2})(\frac{\sigma}{\sqrt{2\pi}})(\frac{1}{\lambda})e^{\frac{-\lambda^2}{2\sigma^2}} + \frac{3\sigma^3}{\sqrt{2\pi}}\int_\lambda^\infty \frac{1}{x^4}e^{\frac{-x^2}{2\sigma^2}}dx \tag{4.2.4}$$

(4.2.3), (4.2.4) give the upper and lower bounds in (4.2.2) respectively. □

Now, apply log and then $\lim_{\lambda\to\infty}$ on (4.2.2). This gives

$$\lim_{\lambda\to\infty}\frac{log(P\{X > \lambda\})}{\lambda^2} = \frac{-1}{2\sigma^2} \tag{4.2.5}$$

.

Remark 4.2.1. Note that a comparison of (4.2.5) and (4.2.1) shows that **if X is a centred Gaussian process on T with a.s bounded sample paths, then** $\|X\|$ **behaves like a centred Gaussian random variable with variance** $\sigma_T^2 = \sup_{t\in T} EX_t^2$ **asymptotically**.

Borell's inequality/Theorem 4.1.1 not only gives a bound on the tail probabilities of the supremum of a Gaussian distribution, but also gives some information about its (asymptotic) distribution. The remarks below end this chapter:

Remark 4.2.2. Borell's inequality also helps find a bound on the tail probabilities of $\sup_{t\in T}|X_t|$. This is because the symmetry of a centred Gaussian process implies

$$P\{\sup_{t\in T}|X_t| > \lambda\} = P\{\sup_{t\in T}X_t > \lambda\} + P\{\sup_{t\in T}X_t < -\lambda\} = 2P\{\|X\| > \lambda\}$$

for any $\lambda > E\|X\|$, for which Theorem 4.1.1 gives a bound.

Remark 4.2.3. Note that $E\|X\|$ is required to make use of Theorem 4.1.1's inequalities.

Chapter 5

Comparison of Suprema

Apart from the distribution of the supremum, another line of investigating suprema is between processes. There are many 'comparison inequalities' that allow comparison between two or more processes in Gaussian process theory. An illustration of how they work: Suppose X, Y are two centred processes defined on T with $EX_t^2 = EY_t^2 \forall t$. Say X is more correlated than Y. Then, Y is in some sense 'rougher' than X. This makes it more likely that $||Y||$ is larger than $||X||$ where $||.|| = \underset{t}{sup}(.)$. Comparison inequalities make such arguments rigorous.

5.1 Kahane's Inequality

An important comparison inequality, from which many other inequalities can be derived directly or with some tweaking, is the Kahane's inequality which we will look at first.

Theorem 5.1.1. *[Kahane's inequality] Let X and Y be a.s bounded centred Gaussian random vectors on $\mathbb{R}^n$. Let $f \in C^2(\mathbb{R}^n)$ be such that its partial derivatives up to 2nd order have subgaussian growth and $EY_iY_j \lesseqgtr EX_iX_j \implies \frac{\partial^2 f}{\partial x_i \partial x_j} \lesseqgtr 0$. Then, $Ef(X) \leq Ef(Y)$.*

Proof.

- Assume X, Y are independent (otherwise take independent copies on the same probability space).
- Interpolate between X and Y by defining $Z(\theta) = cos\theta X + sin\theta Y$; $\theta \in [0, \frac{\pi}{2}]$.
- Define $\Psi(\theta) = Ef(Z(\theta))$ on $[0, \frac{\pi}{2}]$.
- Claim: $\Psi'(\theta) \geq 0$

- Take any small $h > 0$. Then take $\epsilon = h^2$. By Definition of subgaussian growth,

$$\left|\frac{f(Z(\theta+h)) - f(Z(\theta))}{h}\right| \leq \left|\frac{f(Z(\theta+h))}{h}\right| + \left|\frac{f(Z(\theta))}{h}\right|$$
$$\leq \frac{C(h^2)}{h}\big[e^{-h^2\|Z(\theta+h)\|^2} - e^{-h^2\|Z(\theta)\|^2}\big] \leq \frac{C(h^2)}{h}\big[e^{-h^2\|Z(\theta+h)\|^2}\big]$$

Note that $\|Z(\theta+h)\|^2$ is a sum of squares of n Gaussian random variables, and thus has a χ^2_n-distribution. Then, $E\big[e^{-h^2\|Z(\theta+h)\|^2}\big] = E\big[e^{i(ih^2)\|Z(\theta+h)\|^2}\big] = (1-2i(ih^2))^{-n/2} < \infty$. Therefore, we can use Theorem 1.2.5 to to interchange derivative and expectation as follows:

$$\Psi'(\theta) = \frac{d}{d\theta}Ef(Z(\theta)) = E\Big[\frac{d}{d\theta}f(Z(\theta))\Big] = E\left[\sum_{i=1}^{n}\Big(\frac{dZ_i(\theta)}{d\theta}\Big)\Big(\frac{\partial f(Z(\theta))}{\partial z_i}\Big)\right]$$
$$= \sum_{i=1}^{n}E\Big[(-sin\theta X_i + cos\theta Y_i)\frac{\partial f(Z(\theta))}{\partial z_i}\Big] = (-sin\theta)\sum_{i=1}^{n}E\left[X_i\frac{\partial f(Z_i(\theta))}{\partial z_i}\right] + (cos\theta)\sum_{i=1}^{n}E\left[Y_i\frac{\partial f(Z_i(\theta))}{\partial z_i}\right]$$

(5.1.1)

- Since the partial derivatives of f up to 2nd order also have subgaussian growth, we can use Gaussian Integration by parts,

$$E\left[X_i\frac{\partial f(Z(\theta))}{\partial z_i}\right] = E_Y\left[E_{X|Y}\left[(X_i)\frac{\partial f}{\partial z_i}(cos\theta X + sin\theta y)\big|Y = y\right]\right] =$$
$$E_Y\left[\sum_{j=1}^{n}(E_{X|Y}X_iX_j)E_{X|Y}\Big[\frac{\partial f}{\partial x_j\partial z_i}(cos\theta X + sin\theta y)\big|Y = y\Big]\right]$$
$$= E_Y\left[\sum_{j=1}^{n}(E_{X|Y}X_iX_j)E_{X|Y}\Big[\Big(\frac{\partial Z_j}{\partial x_j}\Big)\frac{\partial f}{\partial z_j\partial z_i}(cos\theta X + sin\theta y)\big|Y = y\Big]\right]$$
$$= \sum_{j=1}^{n}(EX_iX_j)(cos\theta)E\left[\frac{\partial f}{\partial z_i\partial z_j}(Z(\theta))\right]$$

Similarly,

$$E\left[Y_i\frac{\partial f(Z(\theta))}{\partial z_i}\right] = \sum_{j=1}^{n}(EX_iX_j)(sin\theta)E\left[\frac{\partial f}{\partial z_i\partial z_j}(Z(\theta))\right]$$

- Substituting the above in (5.1.1), we get

$$\Psi'(\theta) = (cos\theta sin\theta)\sum_{j=1}^{n}E\left[\frac{\partial f}{\partial z_i z_j}(Z(\theta))\right](EY_iY_j - EX_iX_j)$$

- Under the assumptions of the theorem statement, it is now easy to see that $\Psi'(\theta) \geq 0$. Hence the claim is proved.

- This means Ψ is an increasing function on $[0, \frac{\pi}{2}]$.
Then, $\Psi(0) \leq \Psi(\frac{\pi}{2}) \implies Ef(X) \leq Ef(Y)$. Hence proved.

□

Note that Kahane's Inequality here is given for centred Gaussian vectors, and not processes. We will use this to prove some other important inequalities for processes in the sections that follow.

5.2 Slepian's Inequality

Slepian's Inequality gives a rigorous statement for the idea that the supremum of the process whose covariance/correlation is smaller (rougher process) should dominate over the supremum of the other process (smoother process), given two processes defined on the same parameter space whose variances at each parameter are equal.

Theorem 5.2.1. *[Slepian's Inequality] If X, Y are a.s bounded centred Gaussian processes on T such that $EX_t^2 = EY_t^2$ for all $t \in T$ and $E(X_t - X_s)^2 \leq E(Y_t - Y_s)^2$ for all $s, t \in T$; then for all real λ, $P\{\|X\| > \lambda\} \leq P\{\|Y\| > \lambda\}$.*

See that the condition $E(X_t - X_s)^2 \leq E(Y_t - Y_s)^2$ is the same as saying $EX_sX_t \geq EY_sY_t \forall T$ because $EX_t^2 = EY_t^2$. This exactly captures what is said before the theorem statement.

Slepian's Inequality Proof - I

Proof.

- Let the finite parameter space T have $|T| = k < \infty$ elements. Fix $\lambda \in \mathbb{R}$.
- Consider the function $h(x) = \prod_{i=1}^{k} \mathbb{1}_{(-\infty,\lambda]}(x_i)$. Clearly each $\mathbb{1}_{(-\infty,\lambda]} \in L^1(\mathbb{R})$.
- Now, it is a well-known that the Schwartz space of $\mathbb{R}$ is dense in $L^1(\mathbb{R})$. This implies that we can construct a sequence of smooth, bounded, non-negative, non-increasing real-valued functions on $\mathbb{R}$ that converges to $\mathbb{1}_{(-\infty,\lambda]}$. Say $\{f_i^{(m)}\}_{m=1}^{\infty}$ is such a sequence for each $i \in \mathbb{N}_k$. Then, for each $m \in \mathbb{N}$, define $h^{(m)}(x) = \prod_{i=1}^{k} f_i^{(m)}(x_i)$.
- By definition, $h^{(m)}$ is smooth. Also, since each $f_i^{(m)}$ is a Schwartz function, $h^{(m)}(x)$ vanishes as $\|x\|_{\mathbb{R}^k}$ goes to ∞. Then for each $\epsilon > 0, \exists N_\epsilon > 0$ such that

  $$\|x\|_{\mathbb{R}^k} > N_\epsilon \implies |h^{(m)}(x)| < e^{-\epsilon\|x\|^2}$$

  . Let M= $\sup_{\|x\|_{\mathbb{R}^k} \leq N_\epsilon} |h^{(m)}(x)| < \infty$ because $h^{(m)}$ is bounded. Then, define $C(\epsilon)$ =M$e^{\epsilon N_\epsilon^2}$. Clearly, $h^{(m)}$ has subgaussian growth with this $C(\epsilon)$. Similarly, it can be shown that partial derivatives up to 2nd order also have subgaussian growth.

- Also by definition, for any $1 \leq i, j \leq k, i \neq j$,

$$\frac{\partial^2 h^{(m)}}{\partial x_i \partial x_j}(x) = (\prod_{l \neq i,j} f_l^{(m)}(x_l)) f_i^{(m)\prime}(x_i) f_j^{(m)\prime}(x_j).$$

Since $f_i^{(m)}$'s are non-negative and non-increasing,

$$\frac{\partial^2 h^{(m)}}{\partial x_i \partial x_j}(x) \geq 0$$

. Further,

$$E(X_t - X_s)^2 \leq E(Y_t - Y_s)^2, EX_t^2 = EY_t^2 \forall s, t \in T \implies EX_sX_t \geq EY_sY_t$$

.

- Then, we have the conditions of Kahane's inequality satisfied with

$$\frac{\partial^2 h^{(m)}}{\partial x_i \partial x_j}(x) \geq 0$$

when $EX_iX_j \geq EY_iY_j$. By Kahane's inequality (Theorem 5.1.1), $Eh^{(m)}(Y) \leq Eh^{(m)}(X)$. This is true for all m. Since each $h^{(m)}$ is bounded by the integrable function $h^{(1)}$ by construction of these functions, Theorem 1.2.5 is applicable and thus,

$$\lim_{m \downarrow \infty} Eh^{(m)}(Y) \leq \lim_{m \downarrow \infty} Eh^{(m)}(X) \implies E\prod_{i=1}^{k} \mathbb{1}_{(-\infty,\lambda]}(Y) \leq E\prod_{i=1}^{k} \mathbb{1}_{(-\infty,\lambda]}(X)$$
$$\implies P\{\|Y\| < \lambda\} \leq P\{\|X\| < \lambda\} \implies P\{\|X\| > \lambda\} \leq P\{\|Y\| > \lambda\}.$$

- Since λ was chosen arbitrarily, this is true for all $\lambda \in \mathbb{R}$. Hence proved Slepian's inequality for finite parameter space case.

□

Slepian's Inequality Proof - II

Proof.
- Let X, Y be the centred Gaussian processes satisfying the conditions given in the statement of the Theorem 5.2.1. As assumed since 2.4, let T be separable. This means that ∃ countable dense subset $T' \subset T$. Let $T' = \{t_1, t_2, ...\}$ and $T_n = \{t_1, ..., t_n\}$ for each $n \in \mathbb{N}$.

- Claim 1: $\sup_{t \in T_n} X_t \xrightarrow{a.s} \sup_{t \in T'} X_t$ as $n \to \infty$.
For each $\epsilon > 0$; by definition of supremum, $\exists\ i \in \mathbb{N}$ such that

$$X_{t_i} > \sup_{t \in T'} X_t - \epsilon \implies \sup_{t \in T_n} X_{t_i} > \sup_{t \in T'} X_t - \epsilon \text{ a.s } \forall n \geq i.$$

This proves Claim 1.

- Now, fix $\epsilon > 0$. By definition of supremum and dense property of T', $\exists\ t_0 \in T'$ such that $X_{t_0} > \|X\| - \epsilon$ a.s. Then, $\sup_{t\in T'} X_t \geq \|X\| - \epsilon$ a.s. $\implies 0 \leq \|X\| - \sup_{t\in T'} X_t \leq \epsilon$ a.s. Taking $\epsilon \downarrow 0, \|X\| = \sup_{t\in T'} X_t$ a.s.

$$\text{Claim 1 then implies that } \sup_{t\in T_n} X_t \xrightarrow{a.s} \sup_{t\in T} X_t = \|X\| \text{ as } n \to \infty \tag{5.2.1}$$

.

- Fix any $\lambda \in \mathbb{R}$. Define $G_n(\omega) = \mathbb{1}_{\{\omega | \sup_{t\in T_n} X_t(\omega) > \lambda\}}(\omega)$ and $G(\omega) = \mathbb{1}_{\{\omega | \sup_{t\in T} X_t(\omega) \geq \lambda\}}(\omega)$. Similarly, define $H_n(\omega) = \mathbb{1}_{\{\omega | \sup_{t\in T_n} Y_t(\omega) > \lambda\}}(\omega)$ and $H(\omega) = \mathbb{1}_{\{\omega | \sup_{t\in T} Y_t(\omega) \geq \lambda\}}(\omega)$. Then, (5.2.1) $\implies G_n \xrightarrow{a.s} G, H_n \xrightarrow{a.s} H$ as $n \to \infty$.
- Now,

$$P\{\sup_{t\in T} X_t > \lambda\} = P\{\sup_{t\in T} X_t \geq \lambda\} = E[G] \overset{Fatou's}{\underset{lemma}{\leq}} \liminf_{n\to\infty} E[G_n] = \liminf_{n\to\infty} P\{\sup_{t\in T_n} X_t > \lambda\}$$

- Apply Slepian's Inequality for finite parameter space to the last term. Then, the above continues as:

$$\implies P\{\sup_{t\in T} X_t > \lambda\} \leq \liminf_{n\to\infty} P\{\sup_{t\in T_n} Y_t > \lambda\} = \liminf_{n\to\infty} E[H_n] \overset{DCT}{=} E[H] = P\{\sup_{t\in T_n} Y_t > \lambda\}.$$

  where we are able to apply Theorem 1.2.5 for the marked equality because the probability measure is a finite measure.

- Hence proved.

□

Slepian's Inequality - Remarks

Under the given conditions in the statement of Theorem 5.2.1, from symmetry of the centred Gaussian processes, $\inf_T X_t \overset{d}{=} -\sup_T(-X_t) \overset{d}{=} -\sup_T(X_t)$. Therefore, we can state the following:

Corollary 5.2.2. *If X, Y are a.s bounded centred Gaussian processes on T such that $EX_t^2 = EY_t^2$ for all $t \in T$ and $E(X_t - X_s)^2 \leq E(Y_t - Y_s)^2$ for all s,t $\in T$; then for all real λ, $P\{\inf_{t\in T} X_t > \lambda\} \geq P\{\inf_{t\in T} Y_t > \lambda\}$.*

Proof. Note that $\inf_T X_t \overset{d}{=} -\sup_T(-X_t) \overset{d}{=} -\sup_T(X_t)$ due to symmetry arguments. Then, apply Slepian's inequality. □

Corollary 5.2.3 (Gordon's Inequality)**.** *Let* $(X_{ij})_I, (Y_{ij})_I$ *be two collections of centred Gaussian variables defined on* $I = \{(i,j) : 1 \leq i \leq n, 1 \leq j \leq m\}$ *such that*

- $EX_{ij}^2 = EY_{ij}^2$ $\quad (i,j) \in I,$
- $EX_{ij}X_{ik} \leq EY_{ij}Y_{ik}$ $\quad (i,j),(i,k) \in I,$
- $EX_{ij}X_{lk} \geq EY_{ij}Y_{lk}$ $\quad (i,j),(l,k) \in I, i \neq l.$

$$\textit{Then, for all real } \lambda_{ij}, P\left\{ \bigcap_{i=1}^{n} \bigcup_{j=1}^{m} [X_{ij}\lambda_{ij}] \right\} \geq P\left\{ \bigcap_{i=1}^{n} \bigcup_{j=1}^{m} [Y_{ij}\lambda_{ij}] \right\}$$

Idea of Proof. Just like Slepian's inequality (Theorem 5.2.1), this is also a consequence of Kahane's Inequality (Theorem 5.1.1). An appropriate function is taken, and the method that follows is similar to the proof of Slepian's. □

Remark 5.2.1.

1. Gordon's inequality is a more general version. Take $I = \{(i,j) | i = 1, 1 \leq j \leq m\}$ and the specific case that corresponds to Slepian's inequality 5.2.1 is obtained.
2. Gordon's inequality sheds light on the min-max of a rectangular array of Gaussian variables. For all $\lambda > 0, P\{\underset{1\leq i\leq n 1\leq j\leq m}{min\,max} X_{ij} \geq \lambda\} \geq P\{\underset{1\leq i\leq n 1\leq j\leq m}{min\,max} Y_{ij} \geq \lambda\}$.
3. For an increasing function g on $\mathbb{R}$, Gordon's inequality implies that $E\{\underset{1\leq i\leq n 1\leq j\leq m}{min\,max}\, g(X_{ij})\} \geq E\{\underset{1\leq i\leq n 1\leq j\leq m}{min\,max}\, g(Y_{ij})\}$. This means a similar result is true from Slepian's Inequality Theorem 5.2.1 as well.

Remark 5.2.2. Borell's inequality upon a little tweaking does give something for $\underset{T}{sup}|X_t|$. Slepian's inequality upon a little tweaking also does give a result for the infimum of a process. However, Slepian's inequality does not give any comparison related information for $\underset{T}{sup}|X_t|$.

Example 5.2.1. Take T={1,2}. Let $X_1, X_2 \sim N(0,1)$. Define $P_\rho(\lambda) = P_\rho(max(X_1, X_2) > \lambda)$ and $\hat{P}_\rho(\lambda) = P_\rho(max(|X_1|, |X_2|) > \lambda)$ where ρ is the correlation coefficient between X_1 and X_2. Let $\Phi(x) = P(Y > x)$ where $Y \sim N(0,1)$. Now,
$P_{-1}(\lambda) = P(|X_1| > \lambda) = P(X_1 > \lambda) + P(-X_1 > \lambda) = 2\Phi(\lambda)$.
$P_0(\lambda) = 1 - P(X_1, X_2 \leq \lambda) = 1 - (1 - \Phi(\lambda)^2) = 2\Phi(\lambda) - \Phi(\lambda)^2$
$P_1(\lambda) = \Phi(\lambda)$
$\hat{P_{-1}}(\lambda) = P(|X_1| > \lambda) = 2\Phi(\lambda)$
$\hat{P}_0(\lambda) = 1 - P^2(|X_1| < \lambda) = 1 - [(\frac{1}{2} - \Phi(\lambda))^2]^2 = 4(\Phi(\lambda) - \Phi^2(\lambda))$
$\hat{P}_1(\lambda) = P(|X_1| > \lambda) = 2\Phi(\lambda)$

Define process X' on T with $X'_i = X_i \forall i \in T$ The two variables are correlated by 1

Define process Y' on T with $Y'_i = X_i \forall i \in T$ The two variables are correlated by 0

Define process Z' on T with $Z'_i = X_i \forall i \in T$ The two variables are correlated by -1 $EX_i'^2 = EY_i'^2 = EZ_i'^2 \forall i \in T$. And $E(X'_1 - X'_2)^2 = 0, E(Y'_1 - Y'_2)^2 = 2, E(Z'_1 - Z'_2)^2 = 4$. So we have three Gaussian processes satisfying all requirements of Theorem 5.2.1. Now,

$E(X'_i - X'_j)^2 \leq E(Y'_i - Y'_j)^2 \leq E(Z'_i - Z'_j)^2 \forall i, j \in T$ and
$P_1(\lambda) = P(\|X'\| > \lambda) \leq P_0(\lambda) = P(\|Y'\| > \lambda) \leq P_{-1}(\lambda) = P(\|Z'\| > \lambda)$. This verifies Slepian's Inequality.

But, $\hat{P_{-1}}(\lambda) < \hat{P}_0(\lambda)$ and $\hat{P}_1(\lambda) < \hat{P}_0(\lambda)$ which is clearly not following Slepian's inequality. Clearly, the conditions of Theorem 5.2.1 do not provide an inequality for comparing $\sup_T |X_t|$ along the lines of the inequality that the theorem does give for $\sup_T X_t$.

Now we move on to a corollary of Slepian's inequality and open a new discussion building on this corollary, in the next section.

5.3 Sudakov-Fernique Inequality

Let us begin with a corollary of Slepian's Inequality.

Corollary 5.3.1. *If X, Y are a.s bounded centred Gaussian processes on T such that $EX_t^2 = EY_t^2$ for all $t \in T$ and $E(X_t - X_s)^2 \leq E(Y_t - Y_s)^2$ for all $s,t \in T$; then $E\|X\| \leq E\|Y\|$.*

Proof.

$$E\|X\| = \int_0^\infty P\{\|X\| > \lambda\}d\lambda - \int_{-\infty}^0 P\{\|X\| < \lambda\}d\lambda$$
$$\leq \int_0^\infty P\{\|Y\| > \lambda\}d\lambda - \int_{-\infty}^0 P\{\|Y\| < \lambda\}d\lambda = E\|Y\|$$

where the inequality follows from Slepian's Inequality in Theorem 5.2.1. □

Now, Slepian's Inequality and its corollary require that $EX_t^2 = EY_t^2 \forall t \in T$. The natural question then is whether we can do away with this requirement and find a comparison inequality without it.

Intuitively, a comparison between tail probabilities does not make much sense without assuming constant variances across the parameter space. Suppose that we have two centred Gaussian processes and the covariance is smoother for one process. Let us not assume the condition of identical variances. For the sake of visualisation, consider one process to have very high variances. Now, it becomes possible that this be the smoother process. The rougher process then although rougher, has very low variances relatively. An inequality along the lines of Theorem 5.2.1 would predict that the probability of the supremum of the latter process being larger is greater than the probability of the supremum of the former being larger. However, this need not be true. The smoother process with higher process would actually have the better chance of having a dominating supremum among the two. This makes the condition of identical variances important for a comparison between tail probabilities of suprema.

However, in spite of variances being identical or not, both the processes are centred. The expectations of suprema is likely to depend only on the relation between different random variables which are part of the same process, ie, the correlations/covariances. It does seem possible then, that a comparison of expectations may do away with the dependence on constant variances. Indeed, this happens to be true.

Theorem 5.3.2 (Sudakov-Fernique Inequality)**.** *If X, Y are a.s bounded, centred Gaussian processes on T such that $E(X_s - X_t)^2 \leq E(Y_s - Y_t)^2$; then $E\|X\| \leq E\|Y\|$.*

Sudakov-Fernique Inequality - Remarks

Remark 5.3.1. According to [12], there are many extensions and variations of the Sudakov-Fernique inequality that have implications themselves.

Remark 5.3.2. Like Slepian's Inequality, Sudakov-Fernique Inequality also does not hold for $\underset{T}{sup}|X_t|$. Note the following counter example.

Example 5.3.1. Let $Z \sim N(0,\sigma^2), \xi > 0$. Let X be a centred Gaussian process with sample paths bounded a.s. Let $Y_t^{\xi} = \xi Z + X_t \forall t \in T$.
Then, $E(Y_s^{\xi} - Y_t^{\xi})^2 = E(X_s - X_t)^2 \forall s,t \in T$. Y also has a.s bounded sample paths, by definition. Then,

$$E\underset{T}{sup}|Y_t^{\xi}| \geq E\underset{T}{sup}||\xi Z| - |X_t||.$$

Now $\lim_{\xi\to\infty} E||Y_t^{\xi}|| = \infty$. But, $E\underset{T}{sup}X_t < \infty$. This means $\exists \xi_0$ such that for all $\xi > \xi_0, E||Y_t^{\xi}{}_0|| > E||X||$. This clearly contradicts 5.3.2.

Now we move to the proof of Sudakov-Fernique Inequality staring with the finite parameter space case, as usual by now.

Sudakov-Fernique Inequality Proof - I

Proof of Theorem 5.3.2 for finite T. Assume T is finite.

- Let the finite parameter space T have $|T| = k < \infty$ elements.
- Idea: We can take max function, Theorem 5.1.1, and apply Kahane's inequality to be done. This is similar to the approach in the proof of Theorem 4.1.1. However, we cannot use max function directly here as it does not satisfy the requirements for Theorem 5.1.1. Hence, we resort to a smooth approximation of the same and use the methods in the proof of Theorem 5.1.1 to arrive at the conclusion.
- Define $f_\beta : \mathbb{R}^k \to \mathbb{R}$ such that $f_\beta(x) = \frac{1}{\beta} log(\sum_{i=1}^{k} e^{\beta x_i})$ for some $\beta > 0$. Clearly $f_\beta \in C^2(\mathbb{R})$.
- Define

$$p_i(x) = \frac{e^{\beta x_i}}{\sum_{i=1}^{k} e^{\beta x_i}}$$

. Note that :

$$\frac{\partial f_\beta}{\partial x_i}(x) = p_i(x) \ \& \ \frac{\partial f_\beta}{\partial x_i \partial x_j}(x) = \beta \delta_{ij} p_i(x) - \beta p_i(x) p_j(x)$$

- Now, take X and Y. Since T is finite, X, Y can be considered as two Gaussian k-dimensional random vectors. Assume X and Y are independent (if they actually aren't, take independent copies on the same probability space). Interpolate through $\theta \in [0, \frac{\pi}{2}]$ by defining $Z(\theta) = cos\theta X + sin\theta Y$.

- Define $\Psi(\theta) = E[f_\beta(Z(\theta))]$.

$$\frac{d}{d\theta}E[f_\beta(Z(\theta))] = \lim_{h\to 0}\Big\{\frac{1}{h}E[f_\beta(Z(\theta+h)) - f_\beta(Z(\theta))]\Big\}$$

  Now, $\max_i Z_i(\alpha) = \frac{1}{\beta}log(e^{\beta \max_i Z_i(\alpha)}) \leq f_\beta(Z_i(\alpha)) \leq \frac{1}{\beta}logk + \max_i Z_i(\alpha)$ for any $\alpha \in [0, \frac{\pi}{2}]$. Take $h << 1$. Since X, Y are given to have a.s bounded sample paths, any particular Gaussian vector obtained by interpolation between them is also bounded a.s. Also, there is no point in going forward if we are not prepared to accept that $E\max_i Z_i(\alpha) < \infty \forall \alpha$ because then there will be no useful meaning to the Theorem. Hence, we accept that $E\max_i Z_i(\alpha) < \infty \forall \alpha$. Then,

$$\frac{1}{h}\Big|f_\beta(Z(\theta+h)) - f_\beta(Z(\theta))\Big| \leq \frac{1}{h}|f_\beta(Z(\theta+h))| + \frac{1}{h}|f_\beta(Z(\theta))|$$

  Each term after the last inequality can be bounded by $\frac{1}{h}\Big|\frac{1}{\beta}logk + \max_i Z_i(\alpha)\Big|$ or $\frac{1}{h}\Big|\max_i Z_i(\alpha)\Big|$ for appropriate α which are integrable since $E\max_i Z_i(\alpha) < \infty \forall \alpha$. ($\frac{1}{\beta}logk + \max_i Z_i(\alpha)$ is just $Z(\alpha)$ shifted).

- Claim: Gaussian Integration by Parts is applicable for the first partial derivatives of f_β. It is not trivial to check for subgaussian growth for f_β. Hence, we will have to directly prove Result **??** holds. However, we only need to prove again the part where subgaussian growth property of the function was required for Result **??** to hold. The rest of the steps follow exactly as they do in the proof of Result **??**.

  Given $X(0, EX^2)$, to prove: $E[Xf'_\beta(X)] = EX^2E[f''_\beta(X)]$ Starting from the left hand side, and applying two facts, namely: $p'(x) = \frac{-1}{EX^2}xp(x)$ where p(x) is the pdf of X and $f \in C^2(\mathbb{R})$,

$$E(Xf(X)) = (-\sigma^2)\int_{-\infty}^{\infty}(f(x)p(x))'dx + \sigma^2E[f'(X)].$$

  obtained by product rule of differentiation,
  $E\Big[Xp_1(x)\Big] = -EX^2\Big[\lim_{n\to\infty}\big[p_1(n)p(n) - p_1(-n)p(-n)\big]\Big] + EX^2E[p'_1(x)] = EX^2E[f''_\beta(x)]$
  where $p_1(x)$ is as per the definition of p_i's above (here there is only one component). Thus, we are done proving this claim.

- This implies we can use Theorem 1.2.5 to take the derivative inside expectation.

Then,

$$\Psi'(\theta) = E[\frac{d}{d\theta}f_\beta(Z(\theta))] = E\Big[\sum_{i=1}^{k}\frac{d}{d\theta}(Z_j(\theta))\frac{\partial f_\beta}{\partial z_j}(Z(\theta))\Big]$$
$$= (-sin\theta)\sum_{j=1}^{k}E\Big[X_j\frac{\partial f_\beta}{\partial z_j}(Z(\theta))\Big] + (cos\theta)\sum_{j=1}^{k}E\Big[Y_j\frac{\partial f_\beta}{\partial z_j}(Z(\theta))\Big]$$
$$= (cos\theta)(sin\theta)\sum_{i,j=1}^{k}E\Big[\frac{\partial^2 f_\beta}{\partial z_i z_j}(Z(\theta))\Big](EY_iY_j - EX_iX_j)$$
$$= (cos\theta)(sin\theta)\sum_{i,j=1}^{k}(EY_iY_j - EX_iX_j)(E[\beta p_i(Z(\theta))\delta_{ij}] - E[p_i(Z(\theta))p_j(Z(\theta))])$$
$$= (cos\theta)(sin\theta)\{\sum_{i=1}^{k}(EY_i^2 - EX_i^2)E[p_i(Z(\theta))] - \sum_{i,j=1}^{n}(EY_iY_j - EX_iX_j)E[p_i(Z(\theta)p_j(Z(\theta))]\}$$
$$= (cos\theta)(sin\theta)\Bigg\{\sum_{i,j=1}^{k}(EX_i^2 - EY_i^2)E[p_i(Z(\theta))p_j(Z(\theta))] - \sum_{i,j=1}^{k}(EY_iY_j - EX_iX_j)E[p_i(Z(\theta))p_j(Z(\theta))]\Bigg\}$$
$$= (cos\theta)(sin\theta)\frac{1}{2}[\sum_{i,j=1}^{k}[(E(Y_i - Y_j)^2) - E(X_i - X_j)^2]E[p_i(Z(\theta))p_j(Z(\theta))]]$$

where the 4th equality comes from Gaussian integration by parts. Now, it is given that $E(Y_i - Y_j)^2 \geq E(X_i - X_j)^2 \forall i, j \in T$ and $E[p_i(Z(\theta)p_j(Z\theta)] \geq 0 \implies \Psi'(\theta) > 0$.

- $\implies Ef_\beta(X) \leq Ef_\beta(Y)$ We chose an arbitrary $\beta > 0$ to define f_β. Therefore this result is true for all $\beta > 0$. With the assumptions taken above that made DCT applicable for $E[f_\beta(X)]$ for any centred Gaussian random vector X representing a process that has a.s bounded sample paths, we can take $\lim_{\beta\to\infty}$ both sides and use DCT to get
$$E\|X\| \leq E\|Y\|$$
. Hence proved.

□

Sudakov-Fernique Inequality Proof - II

Proof of Theorem 5.3.2 for general T.

- Let X, Y be the centred Gaussian processes satisfying the conditions given in the statement of the Theorem 5.3.2. As assumed since 2.4, let T be separable. This means that ∃ countable dense subset $T' \subset T$. Let $T' = \{t_1, t_2, ...\}$ and $T_n = \{t_1, ..., t_n\}$ for each $n \in \mathbb{N}$.
- Claim 1: $\sup_{t\in T_n} X_t \xrightarrow{a.s} \sup_{t\in T'} X_t$ as $n \to \infty$.
For each $\epsilon > 0$; by definition of supremum, $\exists\ i \in \mathbb{N}$ such that
$$X_{t_i} > \sup_{t\in T'} X_t - \epsilon \implies \sup_{t\in T_n} X_{t_i} > \sup_{t\in T'} X_t - \epsilon \text{ a.s } \forall n \geq i.$$

This proves Claim 1.

- Now, fix $\epsilon > 0$. By definition of supremum and dense property of T', $\exists\ t_0 \in T'$ such that $X_{t_0} > \|X\| - \epsilon$ a.s. Then, $\sup_{t\in T'} X_t \geq \|X\| - \epsilon$ a.s $\implies 0 \leq \|X\| - \sup_{t\in T'} X_t \leq \epsilon$ a.s Taking $\epsilon \downarrow 0, \|X\| = \sup_{t\in T'} X_t$ a.s

$$\text{Claim 1 then implies that } \sup_{t\in T_n} X_t \xrightarrow{a.s} \sup_{t\in T} X_t = \|X\| \text{ as } n \to \infty \tag{7.3.1}$$

.

- Now, for each $n, n+1 \in \mathbb{N}, E[\sup_{t\in T_n} X_t] \leq E[\sup_{t\in T_{n+1}} X_t] \leq E[\sup_{t\in T} X_t]$.

- For each $n \in \mathbb{N}, E[\sup_{t\in T} X_t] \leq E[\sup_{t\in T} Y_t]$. Applying $\lim_{n\to\infty}$ both sides, 1.2.4, and 7.3.1;

$$E\|X\| \leq E\|Y\|.$$

$\square$

Chapter 6

Boundedness and Continuity

In real analysis, it is well-established how a continuous function defined on a compact space is always bounded. The same is of course reflected in sample path properties of Gaussian Processes. If X is a centred Gaussian process defined on a compact parameter space T, and X has a.s sample path continuity; then X has bounded sample paths a.s.

If at all possible, under what conditions would the converse be true? This is the question discussed in this chapter.

Lemma 6.0.1. *Let X be a centred Gaussian on parameter space T, and $t_0 \in T$, then;*

$$E\underset{t\in T}{sup}X_t \leq E\underset{t\in T}{sup}|X_t| \leq E|X_{t_0}| + 2E\underset{t\in T}{sup}X_t$$

Proof. First Inequality:

$$E\underset{t\in T}{sup}X_t = E[max(\underset{t|X_t\geq 0}{sup}X_t - \underset{t|X_t<0}{inf}X_t)] \geq E\underset{t\in T}{sup}X_t.$$

Second Inequality:

$$E\underset{t\in T}{sup}|X_t| = E\underset{t\in T}{sup}|X_t - X_{t_0} + X_{t_0}| \leq E\underset{t\in T}{sup}|X_t - X_{t_0}| + E|X_{t_0}|$$

Now, $|X_t - X_{t_0}| = max(X_t - X_{t_0}, X_{t_0} - X_t) \leq max(\underset{t\in T}{sup}(X_t - X_{t_0}), -\underset{t\in T}{inf}(X_t - X_{t_0}))$.

Note that at $t = t_0, (X_t - X_{t_0}) = 0 \implies -\underset{t\in T}{inf}(X_t - X_{t_0}), \underset{t\in T}{sup}(X_t - X_{t_0}) \geq 0$.

Then, applying the property that max(a,b)≤a+b if a,b ≥ 0;

$$|X_t - X_{t_0}| \leq max(\underset{t\in T}{sup}(X_t - X_{t_0}), -\underset{t\in T}{inf}(X_t - X_{t_0})) \leq |X_{t_0}| + \underset{t\in T}{sup}(X_t - X_{t_0}) - \underset{t\in T}{inf}(X_t - X_{t_0})$$

$$\implies E\underset{t\in T}{sup}|X_t| \leq E|X_{t_0}| + E\underset{t\in T}{sup}(X_t - X_{t_0}) - E\underset{t\in T}{inf}(X_t - X_{t_0})$$

$$= E\sup_{t\epsilon T}|X_t| \leq E|X_{t_0}| + E\sup_{t\epsilon T}X_t - E\inf_{t\epsilon T}X_t = E|X_{t_0}| + 2E\sup_{t\epsilon T}X_t$$

where the last equality follows because $X_t \overset{d}{=} -X_t \implies E\sup_{t\epsilon T}X_t = E\sup_{t\epsilon T}(-X_t) = -E\inf_{t\epsilon T}$. ☐

Theorem 6.0.2. *If X is a centred Gaussian process defined on parameter space T, 1) $P\{Sup_{t\epsilon T}X_t < \infty\} = 1 \iff$ 2) $Esup_{t\epsilon T}X_t < \infty \iff$ 3) $Ee^{\alpha\|X\|^2} < \infty$ for small $\alpha > 0$.*

Proof.

$1 \implies 2$: $P\{Sup_{t\epsilon T}X_t < \infty\} = 1 \implies Esup_{t\epsilon T}X_t < \infty$ by Theorem 4.1.1.

$2 \implies 1$: $Esup_{t\epsilon T}X_t < \infty \implies P\{Sup_{t\epsilon T}X_t < \infty\} = 1$ trivially.

$2 \implies 3$:

$$Ee^{\alpha\|X\|^2} = \int_0^\infty P(e^{\alpha\|X\|^2} > \lambda)d\lambda \leq \int_0^{E\|X\|} P(e^{\alpha\|X\|^2} > \lambda)d\lambda + \int_{E\|X\|}^\infty P(e^{\alpha\|X\|^2} > \lambda)d\lambda$$

$$\leq \int_0^{E\|X\|} d\lambda + \int_{E\|X\|}^\infty P\Big(\|X\| > \sqrt{\frac{log(\lambda)}{\alpha}}\Big)d\lambda \overset{Theorem}{\underset{4.1.1}{\leq}} E\|X\| + \int_{E\|X\|}^\infty 2exp\Big(-\frac{1}{2\sigma_T^2}(\sqrt{log(\lambda)^{\alpha^{-1}}} - E\|X\|)^2\Big)d\lambda$$

$$= E\|X\| + 4\alpha \int_{\sqrt{logE\|X\|^{\alpha^{-1}}}}^\infty ue^{-\frac{1}{2\sigma_T^2}(u-E\|X\|)^2} e^{(\alpha u^2)}du \leq E\|X\| + 4\alpha\int_0^\infty ue^{\alpha u^2 - \frac{(u-E\|X\|)^2}{2\sigma_T^2}}du.$$

where Theorem 4.1.1 is applicable because $2 \iff 1$ is assumed to be true. Note that $\int_0^\infty ue^{\alpha u^2 - \frac{(u-E\|X\|)^2}{2\sigma_T^2}}du < \infty \implies Ee^{\alpha\|X\|^2} < \infty$. The last integral indeed is finite for $\alpha > 0$ sufficiently small such that $\alpha - \frac{1}{2\sigma_T^2} << 0$ so that the last integral obtained can be restated as the expectation of an appropriate Gaussian variable. Hence, proved.

$3 \implies 2$: Check that as x goes to ∞, $e^{\alpha x^2} \geq x$ for any $\alpha > 0$. This is trivially true for $x < 0$ as well. Let f be the pdf of ||X||. Then,

$$E[\|X\|] \leq \int_{-\infty}^0 xf(x)dx + \int_0^N xf(x)dx + \int_N^\infty xf(x)dx \leq \int_{-\infty}^0 e^{\alpha x^2}f(x)dx + \int_0^N xf(x)dx + \int_N^\infty e^{\alpha x^2}f(x)dx$$

This 1st and 3rd terms are bounded because of 3 holding, and the 2nd term is trivially finite. Hence, proved. ☐

Remark 6.0.1. From Lemma 6.0.1 and Theorem 6.0.2, note that:
$\sup_{t\epsilon T}|X_t| < \infty$ a.s $\implies E\sup_{t\epsilon T}|X_t| \iff E\|X\| < \infty \iff \sup_{t\epsilon T}X_t < \infty$ a.s. Hence, boundedness of the mod process is equivalent to boundedness of the process itself.

Before moving to the next section, recall the canonical metric defined earlier in Definition 3.1.1.

6.1 Relationship between Boundedness and Continuity

The following result describes not only answers the question of 'what conditions can imply continuity given boundedness, of sample paths a.s. ?' but also provides more specific information which will be discussed after the result is stated.

Theorem 6.1.1. *If X is a.s bounded on T and τ is a metric on T such that the canonical metric d is τ-uniformly continuous. Then, X is τ-uniformly continuous a.s iff* $\lim\limits_{\eta\to 0}\phi_\tau(\eta) = 0$ *where* $\phi_\tau(\eta) = E\sup\limits_{\tau(s,t)<\eta}(X_s - X_t)$.

Proof. $\Longrightarrow$)
Say X is τ-uniformly continuous a.s.

$$\Longrightarrow P\{\omega | \lim_{\eta\to 0}\sup_{\tau(s,t)<\eta}|X_s(\omega) - X_t(\omega)| = 0\} = 1.$$

$\therefore$ For each $\omega \in \Omega' \subset \Omega$ such that $\Omega' = \{\omega | \lim\limits_{\eta\to 0}\sup\limits_{\tau(s,t)<\eta}|X_s(\omega) - X_t(\omega)| = 0\}$,

$$\lim_{\eta\to 0}\sup_{\tau(s,t)<\eta}(X_s(\omega) - X_t(\omega)) = 0 \tag{6.1.1}$$

Now, X is a.s bounded. By Theorem 6.1.1, $E\|X\| < \infty$. And for any $\eta > 0$,

$$E\sup_{\tau(s,t)<\eta}(X_s - X_t) \le E\sup_{s,t\in T}(X_s - X_t) \le 2E\|X\| < \infty \tag{6.1.2}$$

Then,

$$\lim_{\eta\to 0}\phi_\tau(\eta) = \lim_{\eta\to 0}E\sup_{\tau(s,t)<\eta}(X_s - X_t) = E\big[\lim_{\eta\to 0}[\sup_{\tau(s,t)<\eta}(X_s - X_t)]\big] = 0$$

since (6.1.2) allows the use of DCT to take limit inside expectation, and (6.1.2) implies the last equality to zero.

$\Longleftarrow$)
Now say, d is τ-uniformly continuous and $\lim\limits_{\eta\to 0}\phi_\tau(\eta) = 0$. Then, for each n $\in \mathbb{N}$, $\exists\ \eta_n > 0$, such that $\{\eta_n\}_{n\in\mathbb{N}}$ is a decreasing sequence,

$$\tau(s,t) < \eta_n \Longrightarrow d(s,t) < 2^{-n} \text{ for any } s,t \in T \text{ and for each } n \in \mathbb{N}, \tag{6.1.3}$$

and

$$|\phi_\tau(\eta_n)| < 2^{-n} \text{ and for each } n \in \mathbb{N}. \tag{6.1.4}$$

Now, consider $A_n = \{\omega | \sup_{\tau(s,t)<\eta_n} |X_s(\omega) - X_t(\omega)| > 2^{-\frac{n}{2}}\}$ for each $n \in \mathbb{N}$. X is a.s bounded on T. Then, by Theorem 4.1.1, for any $n \geq 3$;

$$P\{A_n\} \leq 2exp\{\frac{-1}{2\sigma_T^2}(\frac{1}{2^{\frac{n}{2}}} - \phi_\tau(\eta_n))^2\}$$

where

$$\sigma_T^2 = \sup_{\tau(s,t)<\eta_n} E(X_s - X_t)^2 \leq \sup_{d(s,t)<2^{-n}} E(X_s - X_t)^2 = \sup_{d(s,t)<2^{-n}} d^2(s,t) < 4^{-n}. \quad (6.1.5)$$

(by using (6.1.3)). Then, continuing on the same, for $n \geq 3$,

$$P\{A_n\} \leq 2exp\{\frac{-1}{2\sigma_T^2}(2^{-\frac{n}{2}} - \phi_\tau(\eta_n))^2\} \overset{(6.1.5)}{\leq} 2exp\{\frac{-4^n}{2}(2^{-\frac{n}{2}} - \phi_\tau(\eta_n))^2\} \overset{(6.1.4)}{\leq} 2exp\{\frac{-4^n}{2}(2^{-n} - 2^{\frac{n}{2}})^2\}$$
$$= 2e^{-2^{n-1} - 2^{-1} + 2^{\frac{-3n}{2}}} \leq 2e^{-2^{n-1} - 2^{-1} + 1} = 2e^{-2^{n-1} + 2^{-1}}.$$

Then,

$$\sum_{n=3}^{\infty} P(A_n) \leq 2e^{0.5} \sum_{n=3}^{\infty} e^{-2^{n-1}} \leq 2e^{0.5} \sum_{n=3}^{\infty} \frac{1}{2^{n-1}} < \infty \underset{1.2.4}{\overset{Result}{\Longrightarrow}} P(\limsup_{n\to\infty} A_n) = 0$$

Now, as $n \to \infty, \eta_n \to 0$.

$$\therefore, P(\limsup_{n\to\infty} A_n) = 0 \implies P(\{\omega | \lim_{\eta\to 0} \sup_{\tau(s,t)<\eta} |X_s(\omega) - X_t(\omega)| > 0\}) = 0$$

$\implies$ X is τ-uniformly continuous a.s. □

6.2 Implications on Continuity

Theorem 6.1.1 gives an 'if and only if' condition for continuity. Note that it does not only tell about d-metric but also about any metric with respect to which d is uniformly continuous. Further, it tells us about 'uniform continuity' and not just continuity.

Corollary 6.2.1. *Let X be a.s bounded on T and τ is a metric on T such that the canonical metric d is τ-uniformly continuous, and $\lim_{\eta\to 0}\phi_\tau(\eta) = 0$. Then, for all $\epsilon > 0, \exists$ an a.s finite random variable $\delta = \delta(\omega)$ such that, for almost all ω,*

$$W_\tau(\eta) = \sup_{\tau(s,t)<\eta} |X_s - X_t| \leq \phi_\tau(\eta)|log\ \phi_\tau(\eta)|^\epsilon,$$

for all $\eta \leq \delta(\omega)$. That is, $\phi_\tau(.)|log\ \phi_\tau(.)|^\epsilon$ is a uniform sample modulus for X in the metric τ.

The proof of this corollary is similar to the proof of the Theorem 6.1.1 and is not required to appreciate the result. This essentially gives a bound on the τ-modulus of continuity of X. The theorem is thus not only giving a relationship between boundedness and continuity but also providing finer details about the modulus of uniform continuity.

Gaussian process theory is also very rich in 'zero-one' laws. Though the proofs are highly involved, here are some interesting results stated without proof that give a lot of information about the continuity question for any general Gaussian process.

Theorem 6.2.2. *For a Gaussian process X on T,* $P\{\lim_{s\to t} X_s = X_t \text{ for all } t \in T\} = 1 \iff P\{\lim_{s\to t} X_s = X_t\} = 1$*, for each* $t \in T$*.*

This theorem is analogous to saying a function in real analysis is uniformly continuous iff it is continuous at each point in the domain. Uniform continuity implying continuity at each point is trivial. However, the converse implication is what is remarkably interesting.

Theorem 6.2.3. *For a Gaussian process X on T,* $P\{X \text{ is continuous for all } t \in T\} = 0$ *or 1.*

Coupled with the previously stated theorem, this result says that a Gaussian process has either uniformly continuous sample paths a.s or is discontinuous a.s. There is no other possibility.

While the analysis of continuity might be the most routine and mundane, it still appears interesting in the context of Gaussian processes. Part II concludes on this note.

Part III

Attacking the Sample Path Continuity Question

Part II began with an introduction to the modern theory, especially the concept of 'Entropy' and Theorem 3.2.1. So far, this report has emphasised on several results useful in analysing the distribution of suprema,and boundedness of Gaussian processes. Part II also shed light on the apparent relationship between a.s boundedness and a.s continuity of sample paths of Gaussian processes. All of this makes us ready for a detailed discussion on 3.2.1. However, before that, the other half of the backbone of the 'General Theory' is required.

Part III introduces 'Majorising Measures', which along with entropy forms the basis of the modern approach to analysing continuity and suprema of Gaussian processes. Majorising measures, much stronger than entropy in the way they measure size, give very strong results and implications for continuity and boundedness. Some of these are also stated in Part II. The major highlight with majorising measures here is a result that validates our background setup in this report. However, the focus of this report is still the Main Entropy Result and this introduction to majorising measures is made only in the context of proving Theorem 3.2.1. Since majorising measures are harder to visualise, construct and handle, the entropy result makes up for not being as strong as majorising measures.

The relative convenience of the entropy argument to establish almost sure continuity whilst providing several direct/indirect applications and implications is best felt by noting some actual examples. For this very reason, Part III then covers specific examples of the consequence(s) of applying the Main Entropy Result to analyse Brownian Sheet Processes on different parameter spaces, the Gaussian process on R, the set-indexed processes, random fields, and function-indexed processes. The focus in these examples is on the development of suitable bounds, the simplicity of the entropy integral argument, the applicability on processes with very different geometry, the mathematical implications, and the power of a modern attitude.

This part shows that several complicated processes defined on complicated parameter spaces can be treated with the same tools without any discrimination. This proves the might of the modern theory.

Chapter 7

Revisiting the Main Result

Coming back to the introduction of the Modern Approach, and the Main Result of this report (Theorem 3.2.1), there is another tool apart from 'entropy' that is a huge deal for the 'General Theory'. In fact, it is this tool that is far more powerful. This tool is called 'Majorising Measure'. Like entropy, it also is a way to measure the 'size' of the parameter space. Here, majorising measures will only be used to prove the Main Result featuring Entropy.

7.1 Majorising Measures, the Mighty

What are they?

Consider a centred Gaussian process on parameter space T in line with the assumptions in 2.4, and 3.2 which refers to the assumption taken that T is totally bounded in the canonical metric, d.

Define a probability measure $m : \mathfrak{B}_T \to [0,1]$ where $\mathfrak{B}_T$ denotes the Borel σ-algebra of T.

Define a function $g : (0,1] \to \mathbb{R}^+ \cup \{0\}$ such that

$$g(t) = \sqrt{log(\frac{1}{t})}, 0 < t \leq 1.$$

For any $s \in T, \epsilon > 0$;

$$B_d(t,\epsilon) = \{s \in T | d(s,t) < \epsilon\} \text{ (Open d-ball of radius } \epsilon \text{ around t).}$$

Definition 7.1.1. Majorising Measure
If

$$\underset{t \in T}{sup} \int_0^\infty g(m(B_d(s, \epsilon)))d\epsilon < \infty, \tag{7.1.1}$$

then m is called a **majorising measure** for (T, d).

Remark 7.1.1. For the entire discussion on majorising measures, assume that the diam(T) in the d-metric is strictly positive. This is because, if diam(T)=0, then for any $\epsilon > 0$;

$$T \subset B_d(t, \epsilon) \implies m(B_d(t, \epsilon)) = 0$$

and the integral in (7.1.1) vanishes, thus making every probability measure a valid majorising measure trivially anyways.

Remark 7.1.2. Note that for any $\epsilon > diam(T), t \in T; T \subset B_d(t, \epsilon) \implies m(B_d(t, \epsilon)) = 0$.
$\therefore$ it suffices to check that

$$\underset{t \in T}{sup} \int_0^{diam(T)} g(m(B_d(s, \epsilon)))d\epsilon < \infty$$

holds to conclude that m is a majorising measure. This boils down every problem of verifying whether a measure is a majorising measure to just having to check the finiteness $\underset{t \in T}{sup} \int_0^\delta g(m(B_d(s, \epsilon)))d\epsilon$ for any $\delta > 0$ close to 0.

Majorising measures are not as intuitive as entropy. Even their construction is very difficult. In fact, there has been no generally applicable method found to construct a majorising measure.[1]

However, majorising measures have been found to be an important and strong tool to study boundedness and especially continuity of general Gaussian processes. To appreciate this fact, note that majorising measure-based arguments give necessary and sufficient conditions for a.s sample path continuity of any general Gaussian process defined on a totally bounded(in d) metric space.

How do they work?

To get a rough idea of how they work, focus on the integral in (7.1.1). If m is a majorising measure, then every such integral defined for any $t \in T$ must be finite.

As $\epsilon \downarrow 0, g(m(B_d(t,\epsilon))) \uparrow \infty$ and $m(B_d(t,\epsilon)) \downarrow 0$. For the integral to be finite, this means that the rate of decrease in the measure of d-balls must be more than the rate of divergence of $g(x)$ as x goes to 0, so as to control the integral. And since (7.1.1) has a supremum involved, the integral must be controlled at all points of T for a valid majorising measure.

Roughly, this implies that a majorising measure puts as much 'mass' as possible wherever d-balls are small. Then m can vaguely be thought to assign 'point'-masses to each point in the parameter space, to an extent that a larger set in the σ-algebra only has a slight increase in its measure when compared to the measure of a much smaller ball. This will control the growth of $g(m(B_d(t,\epsilon)))$ as $\epsilon \downarrow 0$.

Now, if we have a process with the covariance function having some rough patches here and there, the small d-balls in those regions are going to show more irregularities than larger d-balls. This is because roughness is observed locally and hence, the larger the d-ball, the lesser the effect of roughness. This idea gives a greater meaning to majorising measures. Since majorising measures give as much weight as possible to small d-balls to such an extent that they ensure (7.1.1) holds; this also means that they are able to control the effect of irregularities in the smallest d-balls across the entire parameter space. This includes regions where the covariance function is rough(er). To clarify again, if (T,d) admits a majorising measure, the finiteness in (7.1.1) implies controlled irregularities even at the roughest regions of the parameter space. And this should go on to imply that X is a.s bounded over T.

This rough attempt at understanding majorising measures leads us to an important result which more strongly and rigorously establishes the idea discussed here.

Boundedness

Theorem 7.1.1. *Let m be a probability measure on (T,d). Then,*

$$E\|X\| \leq K \sup_{t \in T} \int_0^\infty g(m(B_d(t,\epsilon)))d\epsilon$$

for some universal constant $K \in (1,\infty)$.

Proof. The proof of this Theorem is too involved and is irrelevant to the goal of this Report. Hence, only the idea of the proof is stated here.

Instead of directly working with X, a process Y is defined on the same parameter space such that $E(X_s - X_t)^2 \leq E(Y_s - Y_t)^2$ and Y has a simpler structure. A suitable bound that lets us prove the result is obtained on $E\|Y\|$. Sudakov-Fernique Inequality (Theorem 5.3.2) is applied to argue that the same bound works for $E\|X\|$. □

Remark 7.1.3. Again, the upper limit of the integral in the bound of Theorem 7.1.1 can be taken as diam(T) for the same reason given in Remark 7.1.2.

Remark 7.1.4. Note that if (T,d) admits a majorising measure, then $E\|X\| < \infty$. Theorem 6.0.2 further implies that X is a.s bounded on T. This is in line with the idea discussed earlier that the existence of a majorising measure must somehow imply boundedness.

Remark 7.1.5. A converse result of Theorem 7.1.1 also exists. If X is bounded with probability one, then ∃ a probability measure m on (T,d) satisfying

$$K^{-1} \sup_{t \in T} \int_0^\infty g(m(B_d(t,\epsilon)))d\epsilon \leq E\|X\|.$$

for an appropriately chosen K.

Remark 7.1.6. With Theorem 7.1.1 and the previous remark, we can conclude that X is a.s bounded on T if and only if (T,d) admits a majorising measure.

Remark 7.1.7. Yes, the goal is to prove a sufficient condition for continuity of general Gaussian processes and Theorem 7.1.1 seems to be more centred around 'boundedness' rather than continuity. However, the intricate relationship between boundedness and continuity of a Gaussian process has been established already. We will see how Theorem 7.1.1 helps with continuity next.

Continuity

The following result provides a nice sufficient condition for a.s boundedness and continuity of X on T. It requires use of the Theorems 7.1.1 and 6.1.1.

This theorem is going to be used directly in the proof of the Main Entropy Result 3.2.1.

Theorem 7.1.2. *Let where* $\gamma_m(\eta) = \sup_{t\epsilon T} \int_0^\eta g(m(B_d(t,\epsilon)))d\epsilon$.
Let X be a centred Gaussian process on T. If there exists a probability measure m on (T,d) such that $\lim_{\eta\to 0}\gamma_m(\eta) = 0$, *then X is a.s bounded and uniformly continuous.*

Remark 7.1.8. The converse also holds, giving rise to an 'if and only if' result. So, the following holds.

A centred Gaussian process X on T is a.s bounded and uniformly continuous if and only if there exists a probability measure m on (T,d) such that

$$\lim_{\eta\to 0}\gamma_m(\eta) = 0.$$

Since the other direction is not required for the goal of this report, it shall not be proved.

Remark 7.1.9. There is a stronger version of the 'if and only if' result mentioned in Remark 7.1.8 which states that X will be almost surely bounded and uniformly continuous on an arbitrary metric space (T,d) if and only if T is totally bounded in the d-metric and $\lim_{\eta\to 0}\gamma_m(\eta) = 0$.

This is interesting because this implies that the assumption we took in Section 3.2 that T is totally bounded in the d-metric does not deny us of any Gaussian processes that we may seek to analyse sample path continuity or boundedness of. And any totally bounded metric space is separable. Hence, this stronger version validates the 'generality' of whatever is covered in this report, in spite of taking the assumptions we took in 2.4 and 3.2.

Here, it is clear that majorising measures strongly and efficiently characterise the a.s sample path continuity of Gaussian processes.

When Theorem 3.2.1 was introduced, a remark also introduced a function based on the entropy result that serves as a modulus of continuity. Then, for a tool as strong as majorising measures, one should expect it to provide a modulus of continuity too. This is indeed the case.

Proof of Theorem 7.1.2

Proof. Given a centred Gaussian process X on T under the assumptions in 2.4 and 3.2. Given there exists a probability measure m on (T,d) such that $\lim_{\eta\to 0}\gamma_m(\eta) = 0$.

- To Prove: X is a.s bounded.

  Given $\lim_{\eta\to 0}\gamma_m(\eta) = 0. \implies \exists\eta > 0$ such that $\gamma_m(\eta) < \infty$.
  Then, by Theorem 7.1.1, $\exists K > 1$ such that

$$
\begin{aligned}
E\|X\| &\le K\sup_{t\in T}\int_0^\infty g(m(B_d(t,\epsilon)))d\epsilon \\
&= K\gamma_m(\eta) + K\sup_{t\in T}\int_\eta^{diam(T)} g(m(B_d(t,\epsilon)))d\epsilon + K\sup_{t\in T}\int_{diam(T)}^\infty g(m(B_d(t,\epsilon)))d\epsilon
\end{aligned}
$$

  where the second term is trivially finite and the third term vanishes. Hence, $E\|X\| < \infty$. By Theorem 6.0.2, X is a.s bounded.

- To Prove: X is a.s uniformly continuous. Define $\phi_d(\eta) = E\sup_{d(s,t)<\eta} E(X_s - X_t)$.

  By Theorem 6.1.1, since X is proven to be a.s bounded and d is trivially d-uniformly continuous, it suffices to prove that $\lim_{\eta\to 0}\phi_d(\eta) = 0$.

  Define U $= \{(s,t) \in T\times T | d(s,t) < \eta\}$.
  Define a two-parameter process on $T\times T$, Y as Y$(s,t) = X_s - X_t$. Then, Y is a centred Gaussian process.
  Define a metric d' on $T\times T \to \mathbb{R}^+ \cup \{0\}$ as

$$d'((s,t),(s',t')) = \sqrt{E[(X_s - X_t) - (X_{s'} - X_{t'})]^2} = \sqrt{E[Y(s,t) - Y(s',t')]^2}$$

  which is essentially the canonical metric on $T\times T$.

  Observation 1: Now, note that for $(s,t) \in T\times T$, the open ball in d'-metric is defined as

$$B_{d'}((s,t),\epsilon) = \{(s',t') | d'((s,t),(s',t')) < \epsilon\}$$

  If $s' \in B_d(s,\frac{\epsilon}{2})$ and $t' \in B_d(t,\frac{\epsilon}{2})$, then by the triangle inequality property of d',

$$
\begin{aligned}
d'((s,t),(s',t')) &\le d'((s,t),(s,s)) + d'((s,s),(s',t')) \\
&= \sqrt{E[(X_s - X_t) - 0]^2} + \sqrt{E[(X_{s'} - X_{t'}) - 0]^2} = d(s,t) + d(s',t') < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.
\end{aligned}
$$

  $\implies B_d(s,\frac{\epsilon}{2}) \times B_d(t,\frac{\epsilon}{2}) \subset B_{d'}((s,t),\epsilon)$

  Observation 2: Let

$$d'((s,t),U) = \inf_{(s',t')\in U} d'((s,t),(s',t'))\forall(s,t) \in T\times T.$$

  Since U is closed, the infimum in the definition of $d'((s,t),U)$ is achieved. That is, for every $(s,t) \in T\times T, \exists(s',t') \in U$ such that $d'((s,t),(s',t')) = d'((s,t),U)$. Define

ϕ on $T \times T$ such that $\phi((s,t)) = (s',t') \in U$ such that $d'((s,t),(s',t')) = d'((s,t),U)$. Now note that for any $(s,t) \in U, (s',t') \in T \times T$;

$$d'((s,t),\phi(s',t')) \leq d'((s,t),(s',t'))+d'((s',t'),\phi(s',t')) = d'((s,t),(s',t'))+d'((s',t'),U)$$

Since $(s,t) \in U, d'((s',t'),U) \leq d'((s',t'),(s,t)) \implies d'((s,t),\phi(s',t')) \leq d'((s,t),(s',t'))$. Then, for any $(s,t) \in U$, if $(s',t') = \phi((s'',t''))$ for some $(s'',t'') \in T \times T$, then;

$$d'((s,t),(s't')) = d'((s,t),(s'',t'')) \leq 2d'((s,t),(s'',t'')) < 2\epsilon$$

$\therefore, \forall (s,t) \in U, \phi(B_{d'}((s,t),\epsilon)) \subseteq U \cap B_{d'}((s,t),2\epsilon)$.

Observation 3: Consider g as defined for the definition of a majorising measure. Then, $g(xy) = \sqrt{-logx - logy}$
Case 1: If $x, y \leq 1$, then $-logx, -logy \geq 0 \implies g(xy) \leq g(x) + g(y)$. Case 2: If $x \leq 1, y > 1$, then $-logy < 0 \implies g(xy) \leq g(x) < 2g(x)$.

Now, define set-function μ on the Borel σ-algebra of U taken as a subspace of T×T, by $\mu(A) = (m \otimes m) \circ \phi^{-1}(A)$ where $m \otimes m$ is the product measure based on the probability measure, m on T.
$\mu(U) = (m \otimes m) \circ \phi^{-1}(U) = (m \otimes m)(T \times T) = 1$
$\mu(\bigcup_{j=1}^{\infty} E_j) = (m \otimes m)(\bigcup_{j=1}^{\infty} \phi^{-1}(E_j)) = \sum_{j=1}^{\infty}(m \otimes m)(\phi^{-1}(E_j)) = \sum_{j=1}^{\infty}\mu(E_j)$ for any countable disjoint collection of subsets $\{E_i\}_{i=1}^{\infty}$ of U. The second-last equality follows because $\{\phi^{-1}(E_j)\}_{j=1}^{\infty}$ is a countable disjoint collection of subsets in $T \times T$. Hence, μ is a valid probability measure on the subsets of U.

And, for all $(s,t) \in U$;

$$\begin{aligned} m(B_d(s,\frac{\epsilon}{2}))m(B_d(t,\frac{\epsilon}{2})) = (m \otimes m)(B(s,\frac{\epsilon}{2}) \times B(t,\frac{\epsilon}{2})) \leq (m \otimes m)(B_{d'}((s,t),\epsilon)) \\ \leq (m \otimes m)(\phi^{-1}(U \cap B_{d'}((s,t),2\epsilon)) = \mu(U \cap B_{d'}((s,t),2\epsilon)). \end{aligned}$$

where the two inequalities follow from the Observations 1 and 2.

Then, by Theorem 7.1.1, noting that by definition of U, diam$(U) \leq 2\eta$; $\exists K > 1$ such that

$$\begin{aligned} \phi_d(\eta) &\leq K \sup_{(s,t)\in U} \int_0^{2\eta} g(\mu(B_{d'}((s,t),\epsilon)))d\epsilon \\ &\leq K \sup_{(s,t)\in U} \int_0^{2\eta} g(m(B_d(s,\frac{\epsilon}{4}))m(B_d(t,\frac{\epsilon}{4})))d\epsilon \\ &\leq 4 \sup_{(s,t)\in T\times T} \int_0^{\frac{\eta}{2}} g(m(B_d(s,\epsilon))m(B_d(t,\epsilon)))d\epsilon \\ &\leq 8 \sup_{t\in T} \int_0^{\frac{\eta}{2}} g(m(B_d(t,\epsilon)))d\epsilon = 8K\gamma_m(\frac{\eta}{2}). \end{aligned}$$

where the inequalities follow from Observations 1,2, and 3.
Then, as $\eta \to 0$, $\gamma_m(\eta) \to 0 \implies \phi_d(\eta) = 0$. Hence, by Theorem 6.1.1, X is a.s uniformly continuous.

□

7.2 Proof of The Main Entropy Result

It is finally time to discuss the Main Result of this Report and it's proof.

In the last section, enough information has been provided to justify that majorising measures are an efficient tool to characterise a.s sample path continuity of general Gaussian processes. At the same time, majorising measures are also difficult to work with. It is not a trivial task to construct a majorising measure on a general metric space, and there is no known method to do so.

However, the concept of entropy- though not as efficient as majorising measures- can be used to characterise a.s continuity of sample paths. Entropy is a much simpler concept compared to majorising measures and is thus, easier to deal with. In the following chapters of Part III, some applications of the entropy arguments-based Theorem 3.2.1 are discussed. These applications will show how the entropy tool can be used to attack the continuity problem in very differently indexed Gaussian processes. The simplicity of the specific results that can be derived for a specific family of processes from the 'General' Main Result prove the 'ease' of using the entropy-based argument.

Before moving on to the Proof of the Result, a Lemma that is required to prove the Result is stated and proved.

The Theorem 3.2.1 is restated below for convenience.

Theorem 3.2.1 Statement

Let X be a centred Gaussian process on an arbitrary metric space T. Recall the assumptions taken in 2.4 and 3.2.
Define $N(\epsilon)$ be the smallest number of closed d-balls of radius ϵ required to cover T.
Define $H(\epsilon) = log(N(\epsilon))$, this is called the metric entropy function for T. Then,

$$\int_0^\infty (H(\epsilon))^{\frac{1}{2}} d\epsilon = \int_0^\infty (log(N(\epsilon)))^{\frac{1}{2}} d\epsilon < \infty \implies \text{X has a.s sample path continuity.}$$

To prove, we need Theorem 7.1.2 and a Lemma that is stated below.

Lemma 7.2.1. *If $\int_0^\infty (H(\epsilon))^{\frac{1}{2}} d\epsilon < \infty$, then $\exists$ a majorising measure m and a universal constant $K \in (1, \infty)$ such that*

$$\sup_{t \in T} \int_0^\eta g(m(B_d(t,\epsilon)))d\epsilon < K\Big(\eta|log\ \eta| + \int_0^\eta (H(\epsilon))^{\frac{1}{2}} d\epsilon\Big),$$

for all $\eta > 0$.

Idea of Proof of Lemma 7.2.1: Let diam(T) = 1. For each n ≥ 0, let $\{A_1^{(n)}, A_2^{(n)}, ..., A_{N(2^{-n})}^{(n)}\}$ be a minimal family of d-balls of radius 2^{-n} that covers T.

Let

$$B_k^{(n)} = A_k^{(n)} \setminus \bigcup_{j<k} A_j^{(n)} \text{ for each } n \in \mathbb{N}, k \in \mathbb{N}_{N(2^{-n})}.$$

Note that for each k, $B_k^{(n)} \neq \phi$ by the 'minimality' in the construction of $\{A_k^{(n)}\}_{k=1}^{2^{-n}}$. Then, for each n, we have a partition, $\{B_k^{(n)}\}_{k=1}^{2^{-n}}$ of T. For each n, k; choose a sequence of points, $\{t_{n,k}\} \subset T$ such that each $t_{n,k} \in B_k^{(n)}$.

Define a set-function on the power set of T as

$$m(E) = \frac{1}{2}\sum_{n=0}^{\infty} \frac{1}{2^n N(\frac{1}{2^n})} \sum_{k=1}^{N(\frac{1}{2^n})} \chi_E(t_{n,k}),$$

where for any set $A \subset T$, for any $t \in T$; $\chi_A(t) = 1$ if $t \in A$ and $\chi_A(t) = 0$ otherwise. Then,

1. $m(\phi) = 0$ trivially
2. $m(T) = \frac{1}{2}\sum_{n=0}^{\infty} \frac{1}{2^n N(\frac{1}{2^n})} \sum_{k=1}^{N(\frac{1}{2^n})} \chi_T(t_{n,k}) = \frac{1}{2}\sum_{n=0}^{\infty} \frac{1}{2^n N(\frac{1}{2^n})} (N(\frac{1}{2^n})) = \frac{1}{2}\sum_{n=0}^{\infty} \frac{1}{2^n} = 1.$
3. Let $\{E_i\}_{i=1}^{\infty}$ be any countable collection of disjoint subsets of T. Then,

$$\begin{aligned}
m(\bigcup_{i=1}^{\infty} E_i) &= \frac{1}{2}\sum_{n=0}^{\infty} \frac{1}{2^n N(\frac{1}{2^n})} \sum_{k=1}^{N(\frac{1}{2^n})} \chi_{\bigcup_{i=1}^{\infty} E_i}(t_{n,k}) \\
&= \frac{1}{2}\sum_{n=0}^{\infty} \frac{1}{2^n N(\frac{1}{2^n})} \sum_{k=1}^{N(\frac{1}{2^n})} \sum_{i=1}^{\infty} \chi_{E_i}(t_{n,k}) \\
&= \frac{1}{2}\sum_{n=0}^{\infty} \frac{1}{2^n N(\frac{1}{2^n})} \sum_{i=1}^{\infty} \sum_{k=1}^{N(\frac{1}{2^n})} \chi_{E_i}(t_{n,k}) \\
&= \frac{1}{2}\sum_{n=0}^{\infty}\sum_{i=1}^{\infty} \frac{1}{2^n N(\frac{1}{2^n})} \sum_{k=1}^{N(\frac{1}{2^n})} \chi_{E_i}(t_{n,k}) \\
&= \sum_{i=1}^{\infty}\left[\frac{1}{2}\sum_{n=0}^{\infty} \frac{1}{2^n N(\frac{1}{2^n})} \sum_{k=1}^{N(\frac{1}{2^n})} \chi_{E_i}(t_{n,k})\right] = \sum_{i=1}^{\infty} m(E_i)
\end{aligned}$$

where the third and fourth equalities follows from the Fubini's Theorem for Infinite Series which is applicable twice because for each $n \in \mathbb{N}$,

$$\sum_{(k,i)\in\{1,2,...,N(\frac{1}{2^n})\}\times\mathbb{N}} \chi_{E_i}(t_{n,k}) \le N(\frac{1}{2^n}) < \infty$$

and,

$$\sum_{(n,i)\in(\mathbb{N}\cup 0)\times\mathbb{N}} \frac{1}{2^n N(\frac{1}{2^n})} \sum_{k=1}^{N(\frac{1}{2^n})} \chi_{E_i}(t_{n,k}) \le \sum_{(n,i)\in(\mathbb{N}\cup 0)\times\mathbb{N}} \frac{1}{2^n N(\frac{1}{2^n})} N(\frac{1}{2^n}) = 2 < \infty.$$

. Hence, the set-function m is a valid probability measure on (T, d).

Now, note that for each $t \in T$, if $\epsilon \in (2^{-(n+1)}, 2^{-n}]$, then

$$m(B_d(t,\epsilon)) \geq (2^{n+1}N(2^{-(n+1)})^{-1}.$$

This implies that $\forall t \in T, n \geq 0$;

$$\int_0^{2^{-n}} \sqrt{(log(1/m(B(t,\epsilon))))}d\epsilon \leq \sum_{k=n+1}^{\infty} 2^{-k}\sqrt{log(2^k N(2^{-k}))}$$
$$\leq \sum_{k=n+1}^{\infty} 2^{-k}\sqrt{klog(2)} + 2\int_0^{2^{-n}} \sqrt{log(N(\epsilon))}d\epsilon$$
$$\leq (n+2)2^{-n}\sqrt{log2} + 2\int_0^{2^{-n}} \sqrt{log(N(\epsilon))}d\epsilon.$$

We know that for any non-dyadic number, we can find a dyadic sequence that converges to it. This fact can be used to conclude the proof. □

Remark 7.2.1. The above Lemma is very important because it relates entropy to majorising measures. Using this lemma, not only Theorem 3.2.1, but many other entropy arguments-based results can be implied by results involving majorising measures.

Proof of Theorem 3.2.1:

Now we have everything required to prove Theorem 3.2.1.

Proof of Theorem 3.2.1. Let m be a probability measure on (T,d). Let $\gamma_m(\eta) = \sup_{t \in T} \int_0^{\eta} g(m(B_d(t,\epsilon)))d\epsilon$.
By Theorem 7.1.2, it is sufficient to prove that $\gamma_m(\eta) \to 0$ as $\eta \to 0$.

Now, by Lemma 7.2.1, $\exists$ a majorising measure μ and a universal constant K such that

$$\gamma_\mu(\eta) < K\Big(\eta|log\ \eta| + \int_0^{\eta}(H(\epsilon))^{\frac{1}{2}}d\epsilon\Big), \tag{7.2.1}$$

for all $\eta > 0$.

Given, $\int_0^{\infty}(H(\epsilon))^{\frac{1}{2}}d\epsilon < \infty$.
Then, applying $\lim_{\eta\to\infty}$ to both sides of the inequality (7.2.1) causes the right hand side to converge to 0. Hence, $\lim_{\eta\to\infty}\gamma_m(\eta) = 0$.

Hence, proved. □

Remark 7.2.2. There is more that can be said from the entropy condition, $\int_0^{\infty}(log(H(\epsilon)))^{\frac{1}{2}}d\epsilon < \infty$: that $\exists$ a universal constant K such that

$$E\|X\| \le K\int_0^{\infty}(log(H(\epsilon)))^{\frac{1}{2}}d\epsilon.$$

Note that if we take $\eta = diam(T)$ in lemma 7.2.1, we get that

$$\sup_{t\epsilon T}\int_0^{diam(T)} g(m(B_d(t,\epsilon)))d\epsilon < K\left(diam(T)\ |log(diam(T))| + \int_0^{diam(T)}(H(\epsilon))^{\frac{1}{2}}d\epsilon\right)$$
$$\le M\int_0^{diam(T)}(H(\epsilon))^{\frac{1}{2}}d\epsilon$$

for some M such that

$$\frac{M}{2} > max\left(\frac{\text{diam(T) log(diam(T))}}{\left(\int_0^{diam(T)}(H(\epsilon))^{\frac{1}{2}}d\epsilon\right)}, K\right)$$

.

Remark 7.2.3. While Remark 7.2.2 gives an upper bound for $E\|X\|$ using an entropy condition, there is also a result due to Sudakov (1971) which gives a lower bound for $E\|X\|$ that involves entropy conditions,

$$K \sup_{\epsilon}\ \epsilon\ (log(N(\epsilon)))^{\frac{1}{2}}$$

. The proof of this result involves a clever application of Theorem 5.3.2.

Theorem 3.2.1 gives only one direction. Is the converse true? The result below tells that it is true for stationary Gaussian processes.

Result 7.2.1. Let X be a stationary Gaussian process on a compact subset of an infinite group or a compact group. Then,

$$\text{X is a.s bounded on T} \iff \text{X is. a.s continuous on T} \iff \int_0^{\infty}\sqrt{log(H(\epsilon))} < \infty.$$

So, for stationary processes we have a strong result using an easier tool entropy, whereas in general, we still have to depend on majorising measures for the most strongest results.

This concludes the discussion on the Main Result. It is now time to check out the modern theory in play. The first applications we see are on Gaussian White Noise processes, in the next chapter.

Chapter 8

The Brownian Family of Processes

The Brownian Family is of great interest in several disciplines. From modelling the random motion of particles, to statistical tests in inference, Brownian processes find importance in several investigations of science. Let us see how the Entropy Result deals with Brownian Processes.

8.1 Gaussian White Noise Processes

Definition 8.1.1. Let (E, ε, ν) be a σ-finite measure space. A **Gaussian White Noise** based on ν is a random set function W on the sets A $\in \varepsilon$ of finite ν-measure such that:
i) W(A) is a centred Gaussian random variable and $EW^2(A) = \nu(A) < \infty$
ii) A $\cap$ B = $\phi \implies$ W(A $\cap$ B) = W(A)+W(B) a.s.
iii) If A $\cap$ B = ϕ, then W(A) and W(B) are independent.

Note that here, we are focusing on a set-indexed process, $\{W(A) | A \in \varepsilon\}$. To be convinced of whether such a process exists, let us look at the covariance function.

Remark 8.1.1. Define $R_\nu : \varepsilon \times \varepsilon \to \mathbb{R}^+ \cup \{0\}$ as

$$R_\nu(AxB) = EW(A)W(B) = \nu(A \cap B)$$

For any A_i's $\in \varepsilon$ and α_i's $\in \mathbb{R}$;

$$\sum_{i,j} \alpha_i R_\nu(A_i, A_j)\alpha_j = \sum_{i,j} \alpha_i \alpha_j \nu(A_i \cap A_j) = \sum_{i,j} \alpha_i \alpha_j \int_E \chi_{A_i}\chi_{A_j} d\nu = \int_E (\sum_i \alpha_i \chi_{A_i})(\sum_i \alpha_j \chi_{A_j}) d\nu$$
$$= \int_E (\sum_i \alpha_i \chi_{A_i})^2 d\nu \geq 0$$

where $\chi_A(x) = 1$ if $x \in A$ and 0 otherwise. Clearly R_ν is positive definite and is hence a well-defined covariance function. Therefore, it determines a well-defined centred Gaussian process, W on ε.

Remark 8.1.2. From (ii) in Definition 8.1.1, it seems like the Gaussian White Noise, W behaves as a random signed measure.

Remark 8.1.3. Because of the property (iii) in Definition 8.1.1, W is said to have 'independent increments'.

The Canonical Metric: Now that the Gaussian White Noise Process has been defined, the canonical metric in ε here,

$$d(A,B) = \sqrt{E\{W(A) - W(B)\}^2} = \sqrt{\nu(A) + \nu(B) - 2\nu(A \cap B)} = \sqrt{\nu(A \Delta B)}$$

Now, we see some specific examples of Gaussian White Noise Processes.

The Brownian Sheet

What is it?

Take $E = \mathbb{R}^k_+ = \{t \equiv (t_1, t_2, ..., t_k) |\ t_i's \geq 0\}, \varepsilon = \mathfrak{B}_{\mathbb{R}^k_+}, \nu = \lambda$ where λ is the Lebesgue measure associated with $\mathbb{R}^k_+$. Consider the Gaussian White Noise on $(\mathbb{R}^k_+, \mathfrak{B}_{\mathbb{R}^k_+}, \lambda)$.

<u>Convention</u>: Define an analogue of 'intervals' in $\mathbb{R}^k_+$ in the following manner:

-(a,b] = $\prod_{i=1}^{k} (a_i, b_i]$ where for each i, $(a_i, b_i]$ is an interval in $\mathbb{R}$.

-[a,b] = $\prod_{i=1}^{k} [a_i, b_i]$ where for each i, $[a_i, b_i]$ is an interval in $\mathbb{R}$.

-(a,b) = $\prod_{i=1}^{k} [a_i, b_i)$ where for each i, $[a_i, b_i)$ is an interval in $\mathbb{R}$.

-(a,b) = $\prod_{i=1}^{k} (a_i, b_i)$ where for each i, (a_i, b_i) is an interval in $\mathbb{R}$.

for a, b $\in \mathbb{R}^+_k$, a$_i$'s, b$_i$'s $\in \mathbb{R}_+$.

Definition 8.1.2 (Brownian Sheet Process)**.** The Process $\{W_t = W((0,t])|t_i's \geq 0\}$ is called the **Brownian Sheet Process**.
Covariance function $R(s,t) = EW_sW_t = \lambda((0,s] \cap (0,t]) = \prod_{i=1}^{k} min\{s_i, t_i\}$

Definition 8.1.3 (Pinned Brownian Sheet Process)**.** Further, the process defined by $\mathring{W}_t = W_t - |t|W_1$, where $|t| = \Pi_{i=1}^{i=k} t_i)$, is called the **pinned Brownian sheet process**.

Observations and Remarks

1. Now, take k = 1. Then,
   - $EW_0^2 = \lambda(\{0\}) = 0 \implies W_0 = 0$.
   - W has independent increments by definition 8.1.1 and 8.1.2.
   - $W_t \sim N(0, ct)$ where c = 1. Thus, the following remark is made.

Remark 8.1.4. When $k = 1, \{W_t : t \geq 0\}$ is the standard Brownian Process.

2. For any $k > 1$, if we think of W as a signed random measure (as remarked in 8.1.2), then we can visualise W_t as a distribution function,

$$W_t = W((0,t]) = \int_0^t W(dx).$$

   Note that $W(dx)$ and integration of a random variable do not make sense without studying of stochastic calculus. This is just an interesting way to look at the Brownian Sheet since a distribution function essentially 'covers' the domain, and is not exactly accurate. This visualisation does justice to calling these processes 'Sheets'.

3. When $k > 1$, if we fix the indices so that we consider the Brownian Sheet process only on one axis, say (t, 0, 0, ..., 0); then

$$EW_t^2 = \prod_{i=1}^{k} t_i = 0 \implies W_t = 0$$

   since $t_2, ..., t_k = 0$. Again, W here is behaving like a measure.

4. For any k>1, if we fix k-1 indices, then the process reduces to a 1-D Brownian Motion. To verify this, say we fix $t_2, ..., t_k$. Then, the process is defined on the first coordinate as $\{W_t = W((0, t_2, .., t_k), (t, t_2, ..., t_k))|t \geq 0\}$.

   Then, $EW_t^2 = t$ x $\prod_{i=2}^{k}(t_i)$. Therefore, $W_t \sim N(0, ct)$ where c = $\prod_{i=2}^{k} t_i$, thereby giving a scaled 1-D Brownian motion (standard if c =1).

5. Consider a 1-D pinned Brownian Sheet process, $\mathring{W}_t = W_t - tW_1$. Notice that $\mathring{W}_1 = W_1 - W_1 = 0$. And $\mathring{W}_0 = 0$. This means that the process fluctuates across (0,1) but is pinned to 0 at 0 and 1. This is why the process is called a pinned Brownian Motion. Extending to $k > 1$, the process is pinned along the boundary of $[0,1]^k$. (Easy to check)

Motivation

Brownian Processes appear in multivariate statistics, non-parametric inference, empirical process theory and so on. This makes their analysis, especially of sample path continuity, important. Some interesting results where they appear and are useful are described below.

Result 8.1.1. Let $\mathbb{Z}_+^k$ be the k-dimensional integer lattice in $\mathbb{R}_+^k$, and $\{X_i\}_{i\in\mathbb{Z}_+^k}$ be an iid sequence of random variables with $EX_i^2 = 1$. Let $n(t) = \{i \in \mathbb{Z}_+^k : 1 \le i_j \le [nt_j]; j = 1, 2, 3, ..., k\}$ for each $t \in [0,1]^k$ where [x] refers to the integer part of x. Then,

$$\frac{1}{n^{\frac{k}{2}}} \sum_{i\in n(t)} X_i \xrightarrow{d} W_t; t \in [0,1]^k$$

where W_t is the Brownian Motion.

Proof. Now, for any k, fix $t \in [0,1]^k$. Then, applying Theorems 2.1.4 and 1.2.1, we get

$$\frac{1}{n^{\frac{k}{2}}} \sum_{i_1=1}^{[nt_1]} \sum_{i_2=1}^{[nt_2]} \sum_{i_3=1}^{[nt_3]} \dots \sum_{i_k=1}^{[nt_k]} X_i$$
$$= \sqrt{\frac{[nt_1][nt_2]...[nt_k]}{n^k}} \frac{1}{\sqrt{[nt_1][nt_2]...[nt_k]}} \sum_{i_1=1}^{[nt_1]} \sum_{i_2=1}^{[nt_2]} \sum_{i_3=1}^{[nt_3]} \dots \sum_{i_k=1}^{[nt_k]} X_i \xrightarrow{d} N(0,|t|) \stackrel{d}{=} W_t$$

where W_t is the k-dimesnional Brownian Sheet process at t $\in [0,1]^k$.
Hence, proved. □

This shows that the Brownian sheet is fundamental to the k-dimensional functional Central Limit Theorem.

Result 8.1.2. Let $\{X_i\}_{i\in\mathbb{N}}$ be a sequence of iid random variables with standard uniform distribution on $[0,1]^k$. Then the empirical cdf $F_n(t) = \frac{1}{n}\sum_{i=1}^{k} \mathbb{1}_{\{X_i \le t\}}$,

$$\sqrt{n}\{F_n(t) - F(t)\} \xrightarrow{d} \mathring{W}_t;$$

where $\mathring{W}_t$ is the k-dimensional pinned Brownian Sheet process at $t \in [0,1]^k$.

Proof. Use Theorem 2.1.4 on the iid sequence $\{Z_i = \mathbb{1}_{\{X_i \in [0,t]\}}\}_i$, and evaluate that $EZ_i = F(t) = |t|$, $\text{Var}Z_i = |t| - |t|^2$ to get

$$\sqrt{n}\{F_n(t) - F(t)\} \xrightarrow{d} N(0, |t| - |t|^2) \stackrel{d}{=} \mathring{W}_t.$$

□

The weak convergence of empirical processes requires limit processes to be continuous. Hence, analysing the continuity of processes like the pinned Brownian sheet is important.

A similar result can also be obtained with the empirical measure ν_n defined on $\mathfrak{B}_{[0,1]^k}$ as $\nu_n(A) = \frac{1}{n}\sum_{i=1}^{k} \mathbb{1}_{\{X_i \in A\}}$ instead of empirical cdf, and the Lebesgue measure λ instead of the actual cdf. We then have the following relationship:

$$\sqrt{n}\{\nu_n(A) - \lambda(A)\} \xrightarrow{d} N(0, \lambda(A) - \lambda^2(A)) \stackrel{d}{=} \mathring{W}(A)$$

where $\mathring{W}(A)$ is the 'set-indexed' pinned Brownian Sheet process at $A \in \mathfrak{B}_{[0,1]^k}$. The proof for this is similar to the proof above.

Some test statistics can be made better when used on a wider and richer class of sets. Hence, studying processes extended to a more general class of sets is of interest too.

8.2 Continuity of Brownian Sheet Processes

The Brownian Sheet on $[0,1]^k$

With enough motivation to analyse Brownian sheets, let us now see the use of Theorem 3.2.1 on Brownian Sheet Processes. A partition of $[0,1]^k$ is defined such that for each $\epsilon > 0$, the function inside the integral of Theorem 3.2.1 is bounded by a function which has a finite integral over $[0, \infty)$.

Proposition 8.2.1. The Brownian Sheet, W_t, and hence, the pinned Brownian Sheet, $\mathring{W}_t$ are continuous on $[0,1]^k$.

Proof. Step 1: Define $S_{(t,\delta)} = \{s \in [0,1]^k : t_i \le s_i \le t_i + \delta; i = 1, 2, ..., k\}$.
Step 2: Note that $\underset{s \in S_{(t,\delta)}}{sup}\ E(W_s - W_t)^2 = \underset{s \in S_{(t,\delta)}}{sup}\ (|s| - |t|) = \Pi_{i=1}^{i=k}(t_i + \delta) - \Pi_{i=1}^{i=k}(t_i)$.
Step 3: Prove using induction that $\Pi_{i=1}^{i=k}(t_i + \delta)$ - $\Pi_{i=1}^{i=k}(t_i) \le k\delta$: This is trivial for k=1, assume it holds for k ≥ 1. Then, since $t \in [0,1]^k$;

$$\prod_{i=1}^{k+1}(t_i + \delta) - \prod_{i=1}^{k}(t_i) = t_{k+1}(\prod_{i=1}^{k}(t_i + \delta) - \prod_{i=1}^{k}(t_i)) + \delta(\Pi_{i=1}^{i=k}(t_i + \delta)) \le (k+1)\delta$$

Step 4: Fix $\epsilon > 0$. Then, choosing $\delta \leq \frac{\epsilon^2}{k}$, it can be shown using result of Step 3 that $S(t, \frac{\epsilon^2}{k}) \subset B_d(t, \epsilon)$.
Step 5: Define lattice A= $\{(\frac{i_1\epsilon^2}{k}, \frac{i_2\epsilon^2}{k}, ..., \frac{i_k\epsilon^2}{k}) | i_1, i_2, ..., i_k \in \mathbb{N}_{\lfloor \frac{k}{\epsilon^2} \rfloor}\}$. It is easy to check that $[0,1]^k \subset \bigcup_{t \in A} S(t, \frac{\epsilon^2}{k}) \subset \bigcup_{t \in A} B_d(t, \epsilon)$.
This implies $N(\epsilon) \leq (\lfloor \frac{k}{\epsilon^2} \rfloor + 1)^k$.
Step 6: We can then bound the integral, $(log(N(\epsilon)))^{\frac{1}{2}}$ over [0,1] by taking a Riemann Upper sum with partition $[\frac{1}{(n+1)^2}, \frac{1}{n^2}]$ so that the bound is a convergent series. The integral is trvially finite over [1,∞).
Hence, from Theorem 3.2.1 (Main Result), W_t is continuous on $[0,1]^k$. Since W_t is continuous, $\mathring{W}_t = W_t - |t| W_1$ is also continuous on $[0,1]^k$. □

We proved that the Brownian sheet process is continuous on $[0,1]^k$. Is this also true for Gaussian White Noise processes indexed by more general sets? The answer happens to be no, unfortunately. This is clear from the next proposition.

The Brownian Sheet on Lower Layers in $[0,1]^k$

Definition 8.2.1. Lower Layers
Define a partial order on $\mathbb{R}_+^k$ such that s $< t \iff s_i < t_i \forall i \in \mathbb{N}_k$ (Similarly extended for '≤'). Considering $\mathbb{R}_+^k$ fitted with this partial order, a set A $\subset \mathbb{R}_+^k$ is said to be a **lower layer**, if for any two points s and t in $\mathbb{R}_+^k; s \leq t$ and $t \in A \implies s \in$ A.

Proposition 8.2.2. The Brownian Sheet defined on the lower layers in $[0,1]^2$ is discontinuous and unbounded with probability 1.

Proof. Step 1: Construction of Lower Layers:
Construct a sequence of disjoint triangles and squares in the following manner.

-We create a sequence of squares and a sequence of triangles first. To do this, start by dividing $[0,1]^k$ into 4 quadrants. In the same numbering as is standard for the 2-D plane, take the first quadrant as the first member of the sequence of squares. Now, begin the sequence of triangles with the triangle bordered by the x-axis, y-axis and the diagonal of $[0,1]^k$ opposite to the origin. Then, follow the patterns shown in 8.1 to build the sequences. For each $n \in \mathbb{N}$, for each $j \in \{1, 2, ..., 2^n\}$; the formal specifications of the triangles and squares taken in the manner shown in 8.1 are described below the figure itself.

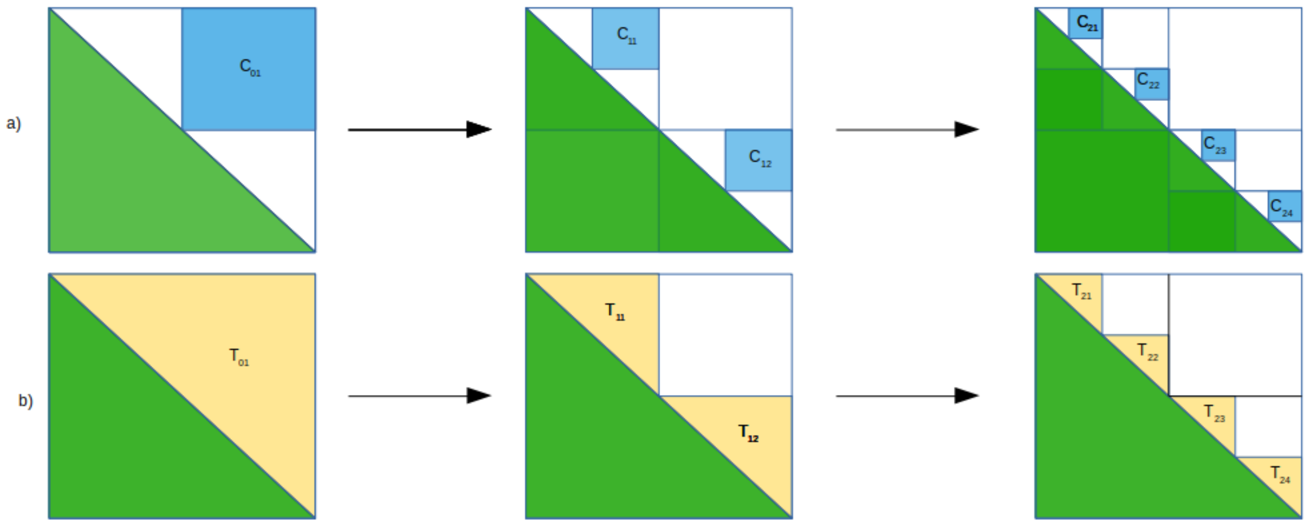

Figure 8.1:
a) C_{01} - square bounded by $\frac{1}{2} \le x \le 1; \frac{1}{2} \le t \le 1$.
C_{nj} - square bounded by $\frac{-(2j-1)}{2^{n+1}} \le s < \frac{j}{2^n}; 1 - \frac{(2j-1)}{2^{n+1}} \le t < 1 - \frac{j-1}{2^n}$.
b) T_{01} - triangle bounded by $x + y \ge 1; x \le 1; y \le 1.$;
T_{nj} - triangle bounded by $s + t \ge 1; \frac{j-1}{2^n} \le s < \frac{j}{2^n}; 1 - \frac{j}{2^n} < t \le 1 - \frac{j-1}{2^n}$

-Using those specifications we have a sequence of squares, $\{C_{nj}\}$ and a sequence of triangles, $\{T_{nj}\}$. Note the following:

- For each n, there are 2^n triangles and 2^n squares.
- $C_{nj} \subset T_{nj}$ for each n,j
- The construction is such that $\bigcap_{n=1}^{\infty} \bigcap_{j=1}^{2^n} C_{nj} = \phi$.
- The construction is such that for each n, $\bigcap_{j=1}^{2^n} T_{nj} = \phi$.

- Define $\{W(C_{nj}) : n \in \mathbb{N}, j \in \mathbb{N}_{2^n}\}$ are independent random variables and $EW^2(C_{nj}) = \lambda(C_{nj}) = 4^{-(n+1)}$. Define D as the diagonal $x + y = 1$ and let $L_{nj} = D \cap T_{nj}$ for each n,j. Note that for each $n \in \mathbb{N}$, if $p \in D \backslash \{(1,0)\}$; then $\exists$ unique $j(n,p) \in \mathbb{N}_{2^n}$ such that $p \in L_{nj}$.

-Let $M < \infty$. Define $E_{n,p} = \{W(C_{nj(n,p)} > \frac{M}{2^{-(n+1)}}\}$ for each $n \in \mathbb{N}, pinD$.
Now, $2^{n+1}W(C_{nj(n,p)}) \overset{d}{=} N(0,1)$.
Then, $P(E_{n,p}) = P(2^{n+1}W(C_{nj(n,p)}) > M) = P(X > M)$ for some $X \sim N(0,1)$.
$\therefore$ for each p, $\sum_{n=1}^{\infty} P(E_{n,p}) = \infty \underset{1.2.4}{\overset{Result}{\Longrightarrow}} P(E_{n,p} \text{ inf. often}) = 1$.

-Note that for fixed p, $E_{n,p}$ is the pre-image of a trivially Borel set in $\mathbb{R}$ through a measurable function (random variable), $W_{C_{nj(n,p)}}$ and is therefore measurable with respect to ω. Also, $E_{n,p}(\omega)$ = D when $\omega \in W^{-1}_{C_{nj}}(\frac{M}{2^{-(n+1)}}, \infty)$ and ϕ otherwise. Therefore $E_{n,p}$ is jointly measurable with p and ω. This implies that by Fubini-Tonelli, for each p, $E_{n,p}$ occurs with probability 1 for some n and the same is true for fixed ω too. Therefore, for all $\omega, p, E_{n,p}(\omega)$ occurs with probability 1 for some n. Therefore, $\exists k(p) \in \mathbb{N}$ such that $\mathrm{P}(E_{k(p),p}(\omega)) = 1$

-For each ω where $E_{n,p}(\omega)$ occurs with probability 1, define

$$V_\omega = \bigcup_{p \in \mathring{D}} T_{k(p)j(k(p),p)}; A_\omega = V_\omega \cup \{(s,t) : s + t \leq 1\}; B_\omega = A_\omega \setminus \bigcup_{p \in \mathring{D}} C_{k(p)j(k(p),p)}$$

where $\mathring{D}$ is the subset of D where $E_{n,p}(\omega)$ occurs with probability 1. Then, check that A_ω, B_ω are lower layers in $[0,1]^2$.

Step 2: Define Q = Note that for any Q, $W(B_\omega \cup Q) = W(B_\omega) + W(Q)$a.s;
$\implies W(A_\omega) - W(B_\omega) = W(Q)$ a.s.
Take Q to be the set of points on $\mathring{D}$ that have rational coordinates. Then, there are countable points in Q, so
$W(A_\omega) - W(B_\omega) = W(Q) \geq \sum_{q \in Q} \frac{M}{2^{k(q)+1}} = \sum_{n \in \mathbb{N}} \frac{M}{2^{n+1}} = \frac{M}{2}$ a.s.
Then, $|W(A)|, |W(B)| \geq \frac{M}{4}$ a.s.

Step 3: The above is true for any $M < \infty$. Take $M \uparrow \infty$ on $|W(A)|, |W(B)| \geq \frac{M}{4}$ a.s; to conclude that W is unbounded, and hence discontinuous, on the lower layers of $[0,1]^2$ with probability 1.

□

The Brownian Sheet on other index-sets

- A similar proof as done for lower layers above also shows that W is unbounded over the convex subsets of $[0,1]^3$.
- Entropy arguments show that W is also unbounded over convex subsets of the $[0,1]^k \forall k \geq 4$. How can entropy arguments be used to see if a process is discontinuous? Refer to the following example.

Example 8.2.1. 1) Consider the construction of $\mathbb{A}_\gamma = \{A_{nj}\}$ below.
Let $\gamma > 0$. Let $A_{01} = [0,1]^2$, and $A_n's$ be the closed rectangles whose left side is the right side of A_{n-1}, so that it has height 1 and width $2^{n(1-\gamma)}$. Further divide each A_n into 2^n equal horizontal slices, $A_{n1}, ..., A_{n2^n}$. Consider this class of sets, $\mathbb{A}_\gamma = \{A_{nj}\}$ and the Gaussian White Noise Process indexed by this class of sets.

2) $\gamma > 1$ Case: Check that the diam$(\mathbb{A}_\gamma) = 2 \underset{A,B \in \mathbb{A}_\gamma}{sup} \lambda(\bigcup_{n,k} A_{nk}) = 2(S_\gamma)^{0.5}$ for $\gamma > 1$.

There is a result that shows $\forall \gamma > 1, a_1 exp(b_1 \epsilon^{\frac{-2}{\gamma-1}}) \leq N(\epsilon) \leq a_2 exp(b_2 \epsilon^{\frac{-2}{\gamma-1}})$ which we will take for granted. Use this result to bound the entropy integral from Theorem 3.2.1 on both sides. Check using the upper bound and lower bound thus obtained to conclude that the entropy integral converges if and only if $\frac{1}{\gamma-1} < 1 \iff \gamma > 2$. Thus, using entropy arguments we not only obtain information about continuity but also about discontinuity since we can definitively say that W is discontinuous on $\mathbb{A}_\gamma$ for $1 < \gamma \leq 2$. Note: we assume $a_1 > 1$ in the result taken for granted here since $a_1 < 1$ gives a trivial bound. This assumption is required to prove that the entropy integral diverges for all $1 < \gamma \leq 2$ using the lower bound in the result.

3) The $\gamma \leq 1$ case gives $\text{diam}(\mathbb{A}_\gamma) = \infty$. Hence, it is not totally bounded, and as discussed in 7.1.9, there is no need to analyse this.

- Surprisingly, W is continuous over convex subsets of $[0, 1]^2$. This shows there is no relationship between the topological properties of the parameter space and sample path continuity since sets topologically similar show different results.

Conclusion

In the examples above, it is noted that similarly defined Gaussian processes are continuous and discontinuous when indexed by different parameter spaces. They also vary in terms of boundedness. This proves that as far as continuity and boundedness are concerned, the relationship between the process and its parameter space is indeed important. However, the above examples were still handled with a common analysis involving the canonical metric and hence, the covariance function. This already depicts the power of the General Theory and the modern attitudes.

Let us see some more examples of different Gaussian process families being analysed by the same tools.

Chapter 9

Gaussian Processes on $\mathbb{R}^k$

9.1 Real-indexed Gaussian Processes, k=1

Let X be a centred Gaussian process defined on a finite interval $[0,T]$, and define $^2(u) = \sup_{|s-t|\leq u} E|X_s - X_t|^2$. This is the same as $\rho^2(u)$ which was defined when the Canonical metric was introduced first.

Proposition 9.1.1. Let X be a centred Gaussian Process on a finite interval [0, T]. If for any $\delta > 0$, $\int_0^\delta (-log(u))^{\frac{1}{2}}$ dp(u) $< \infty$, then X is continuous on [0, T].

This theorem is a well-established result in the theory of real-indexed Gaussian processes. It has been derived and proven without the notion of entropy. However, avoiding the notion of entropy actually makes a proof of this theorem difficult. This is because relying only on the geometry of $\mathbb{R}$ shifts focus from the relationship between the process and the 'size' of the parameter space.

Let us now prove the above theorem using Theorem 3.2.1.

Proof of Proposition 9.1.1. 1. Define p(u) = $[Sup_{|s-t|\leq u} E(X_s - X_t)^2]^{\frac{1}{2}}$. Note that it is a monotone non-decreasing function of u. This means that an inverse p^{-1} exists and it is easy to check that $p^{-1}(u) = u$.

2. Note that diam(T) = p(T). Claim: $N_\epsilon = 1 \ \forall \epsilon > 2p(\frac{T}{2})$.
Take any $s \in [0,T] \implies |s - \frac{T}{2}| \leq \frac{T}{2} \implies d(s, \frac{T}{2}) \leq p(\frac{T}{2}) \implies B_d(\frac{T}{2}, \epsilon)$.
$\therefore [0,T] \subset B_d(\frac{T}{2}, \epsilon) \implies N(\epsilon) = 1$.

3. Partition [0,T] into $\lfloor \frac{T}{2p^{-1}(\epsilon)} \rfloor + 1$ intervals of length $2p^{-1}(\epsilon)$ in the Euclidean metric. Then, for any such interval I,

$x \in I \equiv [a,b] \implies d(x, a+p^{-1}(\epsilon)) = \sqrt{E(X_s - X_{a+p^{-1}(\epsilon)})^2} \le p^{-1}(p(\epsilon)) = \epsilon$. Therefore, $N(\epsilon) \le \lfloor \frac{T}{2p^{-1}(\epsilon)} \rfloor + 1$.

3. Bound the entropy integral in Theorem 3.2.1 using the above information and obtain an integral of the form in the sufficient condition given, after a appropriate change of variables.

$$\begin{aligned}
\int_0^\infty (\sqrt{log(N(\epsilon))}d\epsilon &= \int_0^{\frac{p(T)}{2}} \sqrt{log(N(\epsilon))})d\epsilon \\
&\le \int_0^{\frac{p(T)}{2}} \left(\sqrt{log(\lfloor \frac{T}{2p^{-1}(\epsilon)} \rfloor + 1)}\right)d\epsilon \\
&= \int_0^{\frac{T}{2}} \left(\sqrt{log(1+\frac{T}{2u})}\right)dp(u) \\
&\le \int_0^{\frac{T}{2}} \left(\sqrt{log(k+1) + log(\frac{T}{2}) - logu}\right)dp(u)
\end{aligned}$$

The last expression is bounded if $\int_0^\delta (-log(u))^{\frac{1}{2}}$ dp(u) $< \infty$ and since this holds, the Entropy integral is bounded and X is continuous on [0,T] by Theorem 3.2.1. □

Comments: Processes on $\mathbb{R}$

- Avoiding the notion of entropy leads to a more complicated analysis than the above.
- On the real line, it can be shown that if for some $0 < c < \infty$ and $\alpha, \eta > 0$;

  $$E|X_s - X_t|^2 \le \frac{c}{|log|s-t||^{1+\alpha}} \forall s,t \text{ such that } |s-t| < \eta$$

  ; then X is continuous on [0,T]. This condition actually implies the finiteness of the integral in Proposition 9.1.1 and hence follows. Note that this shows continuity as a consequence of the smoothness of the covariance function at the origin, which is a nice perspective to visualise continuity.
- Taking a look at the last few steps of the proof of Proposition 9.1.1, it is clear that the proposition is a specialisation of the Theorem 3.2.1 and loses the ability to determine continuity of some continuous processes.

  To make it more clear, consider a process X on a suitable compact set T ⊂ $\mathbb{R}$ that satisfies the integral condition of Proposition 9.1.1 and thereby is continuous. Let f be a homeomorphism from T to f(T).

  Then, the process defined by f(X_t) may not satisfy the integral argument of Proposition 9.1.1. However, as can be expected from the action of a homeomorphism, the process f(X_t) is indeed continuous. In fact the Theorem 3.2.1 would be able to prove this.

9.2 Processes on $\mathbb{R}^k, k \in \mathbb{N}$

Definition 9.2.1. Random Fields A stochastic process whose parameter space is either a k-dimensional Euclidean space or a k-dimensional lattice is called a **Random Field**.

This section is simply a short discussion on Gaussian Random Fields.

Now, many problems related to properties of sample paths of a Gaussian random field that depend on geometrical structure of the parameter space are very different and more difficult than the k=1 case. However, this is not the case for sample path continuity.

Problems related to sample path continuity are generally independent of dimension, which is fortunate.

The Proposition 9.1.1 done for the k=1 case holds true for any k with appropriate modifications as stated below.

Proposition 9.2.1. Let X be a centred Gaussian Process on a compact subset $K \subset \mathbb{R}^k$. Let p(u) be defined as

$$p^2(u) = \sup_{||s-t||_{\mathbb{R}^k} \leq \sqrt{k}u} E|X_s - X_t|^2$$

on K.
Then, $\int_K (-log\ u) dp(u) < \infty \implies$ X is continuous on K.

Similar to what was remarked in Section 9.1 for $\mathbb{R}$-indexed processes, it can be shown that if for some $0 < c < \infty$ and $\alpha, \eta > 0$;

$$E|X_s - X_t|^2 \leq \frac{c}{\left|log||s-t||_{\mathbb{R}^k}\right|^{1+\alpha}} \forall s,t \in K \text{ with } ||s-t||_{\mathbb{R}^k} < \eta$$

; then $\int_K (-log\ u) dp(u) < \infty \implies$ X is continuous on K.

Again, the proof of the proposition here is similar to the one for k=1 case, using Theorem 3.2.1, with appropriate modifications of the Euclidean norm everywhere.

Chapter 10

Function-indexed, Set-Indexed and Other Processes

10.1 Generalised Random Fields

Now, consider a centred Gaussian Random Field, X' on $\mathbb{R}^k$ with covariance function R'(s,t).

Let $\mathcal{F}$ be a family of functions on $\mathbb{R}^k$, and for $\phi \in \mathcal{F}$, define

$$X(\phi) = \int_{\mathbb{R}^k} \phi(t) X'(t) dt.$$

Note that each $X(\phi)$ is a valid Gaussian random variable.

Then, we obtain a centred Gaussian process indexed by functions in $\mathcal{F}$, whose covariance functional is given by:

$$R(\phi, \psi) = EX(\phi)X(\psi) = \int_{\mathbb{R}^k} \int_{\mathbb{R}^k} \phi(s) R'(s,t) \psi(t) ds dt$$

for each $\phi, \psi \in \mathcal{F}$.

Definition 10.1.1. The real-valued Gaussian Process X defined on a family of functions $\mathcal{F}$ with covariance functional $R(\phi, \psi)$ is called a **Generalised Random Field**.

Though we first assumed that R is a well-defined covariance function of a process X' that exists, this definition of allows us to define function-indexed processes with a covariance functional like the above, even when a point-indexed process X' with covariance R'(s,t) does not exist.

A centred Gaussian function indexed process can be defined on a family of functions,

$$\mathcal{F}_R = \{\phi | \int_{\mathbb{R}^k} \int_{\mathbb{R}^k} \phi(s) R'(s,t) \psi(t) ds dt < \infty\}$$

, for any positive semi-definite function R'.

What we are interested in is when are these processes continuous? On what family of functions?

Continuity of Generalised Random Fields

The below result describes a family of functions on which a well-defined Generalised Random Field happens to have a.s sample path continuity.

Result 10.1.1. Define $\mathcal{F}^q(T, C_0, .., C_q)$ where T is a bounded subset of $\mathbb{R}^k, q > 0$ and $p = \lfloor q \rfloor$ as the class of functions on T whose partial derivatives of orders 0,1,...,p are bounded by finite positive constants $C_0, ..C_p$ and the partial derivatives of order p satisfy Holder conditions of order q-p with constant C_q.

A centred Gaussian process with covariance function(al),

$$R(\phi, \psi) = EX(\phi)X(\psi) = \int_{\mathbb{R}^k} \int_{\mathbb{R}^k} \phi(s) R(s,t) \psi(t) ds dt$$

where R is a well-defined positive semi-definite function on $\mathbb{R}^k \times \mathbb{R}^k$; and

$$R(s,t) \le \frac{c}{||s-t||^{\alpha}_{\mathbb{R}^k}} \forall ||s-t||_{\mathbb{R}^k} \le \delta,$$

for some $c < \infty$ and $\delta > 0$ is continuous on $\mathfrak{F}^q(T, C_0, .., C_q)$ if $k > \alpha$ and $q > \frac{1+\alpha-k}{2}$.

The proof is along the same lines: trying to bound the entropy integral of Theorem 3.2.1 by an integrable function. It is however tedious while not being necessary to appreciate the power of the Main Entropy Result, which is our ultimate goal here. Hence, the Result is stated here without proof. Note that using Theorem 3.2.1 is still simpler than attempting to prove this directly using the geometry of the complicated parameter space of functions here.

Hence, Theorem 3.2.1 is able to characterise continuity of processes defined on spaces as complicated as the collection of functions in the above result to give a relatively simpler test for continuity of some processes. How this has further implications in other places can be appreciated from the remark below.

Remark 10.1.1.

- The function space taken in the above Result is useful in the study of infinite dimensional diffusion, stochastic PDEs, etc. Hence, a result determining continuity of processes defined as in the Result has several implications in these fields.
- The class of generalised fields considered here is also the 'Free Field' used in Euclidean Quantum Field Theory. Quantum Field Theory is a part of theoretical physics that is of great interest currently in theoretical physics.

10.2 Dudley-Class-indexed Processes

The Brownian Family of Processes that were seen earlier are set-indexed processes. There, the sets were Borel sets. What happens if we consider a parameter space of more general sets and general measures? And why should we bother?

The motivation behind analysing Gaussian processes on general sets is to develop better statistical tests. As discussed earlier in this report, the Kolmogorov-Smirnov test for example requires analysis of Brownian Sheets. Extending to a richer class of sets apparently makes such tests better. Hence, general set-indexed processes are of interest.

The focus in this section and the next is more on the interesting properties and results rather than specific details as we only seek to appreciate the power of the Main Entropy Result.

Let's start with the Dudley Class.

Definition 10.2.1. Construction of the Dudley Class of Sets:

1. Consider the standard smooth atlas $\{(V_j^+, F_j)\} \cup \{(V_j^-, F_j)\}$ on S^k, the k-dimensional sphere.
$V_j^+ = \{(x_1, ...x_{k+1}) | x_j > 0\}$
$V_j^- = \{(x_1, ...x_{k+1}) | x_j < 0\}$
$F_j : S^k (\subset \mathbb{R}^{k+1}) \to B^k$ is a diffeomorphism defined as

$$F_j((x_1, ..., x_{k+1})) = (\frac{x_1}{x_j}, ..., \frac{x_{j-1}}{x_j}, \frac{x_{j+1}}{x_j}, ..., \frac{x_k}{x_j})$$

, where B^k refers to the open unit ball in $\mathbb{R}^k$.
2. Just as $\mathcal{F}^{(q)}(T, C_0, C_1, ..., C_p, C_q)$ was defined in the Generalised Random Fields Discussion above, we define here a family of functions for each coordinate chart, say $V_j^{+/-}$ (denoting generally, any of the $V_j^+ {}'s$ and $V_j^- {}'s$ both) as:

Define $\mathcal{F}^{(q)}(V_j^{+/-}, M)$ as the set of all $\mathbb{R}$-valued functions ϕ on $V_j^{+/-}$ such that

$$\phi|_{V_j^{+/-}} \circ F_j^{-1} \in \mathcal{F}^{(q)}(B^k, M, ..., M)$$

. 3. Then, define $\mathcal{F}^{(q)}(S^k, M)$ as the set of all real-valued functions ϕ on S^k such that

$$\phi|_{V_j^{+/-}} \circ F_j^{-1} \in \mathcal{F}^{(q)}(V_j^{+/-}, M)$$

for some coordinate chart of S^k. Consider $D(k+1, q, M) = \Pi_{i=1}^{k+1} \mathcal{F}^{(q)}(S^k, M)$.
4. From this construction, each $\phi \in D(k+1, q, M)$ represents a k-1 dimensional surface in R^{k+1}. An application of algebraic geometry (which is not required for the goals of this report) on these surfaces yields a class of sets called the Dudley family of sets in $\mathbb{R}^{k+1}$.

Result 10.2.1. The Brownian Sheet is continuous on a bounded collection of Dudley sets in $\mathbb{R}^{k+1}$ with q times differentiable boundaries if $q > k \geq 1$. If $k > 1 \geq q > 0$ or if $k > q \geq 1$, then the Brownian Sheet is unbounded with probability one.

The proof requires algebraic geometry and is thus beyond the scope of this report. However, it is interesting to know that the proof for this result also follows from the doing the same thing that was done for the processes considered previously - bound the Entropy integral and use the Main Result. Doing so gives a nice relationship between k and q that imply continuity/unboundedness with probability 1.

10.3 The Vapnik-Chervonenkis (VC) Class

What happens when we move to a Gaussian White Noise process defined with more general σ-finite measure spaces rather than the usual Lebesgue measure?

Definition 10.3.1. Construction of the VC Class of Sets

Let E $\subset \mathbb{R}^k$ and ν be a probability measure on E. Given a class $\mathfrak{C}$ of subsets of E and a finite set F$\subset$E, let $\triangle^{\mathfrak{C}}$(F) be the number of different sets $C \cap F$ for $C \in \mathfrak{C}$.

For n=1,2,..., let $m^{\mathfrak{C}}(n) := \max\{\triangle^{\mathfrak{C}}(F)$:F has n elements$\}$.

$$\text{Set } V(\mathfrak{C}) = \begin{cases} inf\{n : m^C(n) < 2^n\} & m^{\mathfrak{C}}(n) < 2^n \text{ for some n} \\ \infty & m^{\mathfrak{C}}(n) = 2^n \text{ for all n} \end{cases}$$

The class $\mathfrak{C}$ is called a VC class if $m^{\mathfrak{C}}(n) < 2^n$ for some n, ie, if V($\mathfrak{C}$) < ∞. The value of V($\mathfrak{C}$) is called the VC-index of $\mathfrak{C}$.

This VC class of sets is interesting because it allows the consideration a very large class of sets across different general measure spaces. The following result makes such sets special and worth mentioning here.

Result 10.3.1. Let W be a Gaussian White Noise process based on probability measure ν on some measure space (E, ε, ν). If $\mathfrak{A}$ is a VC class of sets in ε with $V(\mathfrak{A}) = v$. Then there exists a constant K = K(v) independent of ν such that $0 < \epsilon \leq \frac{1}{2}$, the entropy function for W on $\mathfrak{A}$ satisfies

$$N_{\mathfrak{A}}(\epsilon) \leq K e^{-2v} |log\ \epsilon|^{v}.$$

The implication of this, given the Main Entropy Result 3.2.1 is stated below:

Corollary 10.3.1. *Let W be a Gaussian White Noise process based on probability measure ν on some measure space (E, ε, ν). Then, W is a.s continuous over any VC class of sets, $\mathfrak{A}$ in ε.*

Note how this corollary establishes sample path continuity of Gaussian White Noise processes defined on sets in any arbitrary space using the Entropy Result.

Chapter 11

A Little More to See

11.1 Gaussian Fourier Series

Definition 11.1.1. Gaussian Fourier Series
The sum represented by

$$\sum_{0}^{\infty} a_n Y_n e^{int}, t \in [0, 2\pi]$$

; where $\{a_n\} \subset \mathbb{R}$ such that

$$\sum_{0}^{\infty} a_n^2 = 1$$

and $\{Y_n\}$ is an iid sequence of standard Gaussian random variables; is called a **Gaussian Fourier Series**.

- The uniform convergence or divergence of Gaussian Fourier series hold a number of consequences in non-random harmonic analysis. Let us look at a result that depicts this.

Result 11.1.1. Let $\{Y_n\}$ and $\{Y_n'\}$ be two independent, infinite sequences of independent, standard normal random variables. Let $\{a_n\}$ be a non-increasing real sequence. Then, the process X defined as

$$X_t := \sum_{0}^{\infty} a_n (Y_n cos(nt) + Y_n' sin(nt)), t \in [0, 2\pi], \tag{11.1.1}$$

converges uniformly on $[0, 2\pi]$ if and only if the sum :

$$\sum_{j=2}^{\infty} \left(\frac{(\sum_{n=j}^{\infty} a_n^2)^{\frac{1}{2}}}{j(logj)^{\frac{1}{2}}} \right) \tag{11.1.2}$$

converges.

Since we are interested in seeing the use of the Entropy result, let us look at the proof of the sufficiency part of the theorem, ie, (11.1.2) $\Longrightarrow$ uniform convergence of (11.1.1).

Note that if X as in (11.1.1) is a process, then the continuity of X is equivalent to the convergence of X.

Proof. Claim 1: $\sum a_n$ is convergent.
Assume $\sum\limits_{n=1}^{\infty} a_n = \infty$. Then for any n $\in \mathbb{N}$, $a_1 + ... + a_n + \sum\limits_{i=n+1}^{\infty} = \infty$

$$\Longrightarrow \left(\frac{(\sum_{n=2}^{\infty} a_n^2)^{\frac{1}{2}}}{2(log2)^{\frac{1}{2}}}\right) + \left(\frac{(\sum_{n=3}^{\infty} a_n^2)^{\frac{1}{2}}}{3(log3)^{\frac{1}{2}}}\right) + ... = \infty \Longrightarrow \sum_{j=2}^{\infty}\left(\frac{(\sum_{n=j}^{\infty} a_n^2)^{\frac{1}{2}}}{j(logj)^{\frac{1}{2}}}\right) = \infty(\rightarrow\leftarrow)$$

Claim 2: The series $\sum_0^{\infty} a_n(Y_n cos(nt) + Y_n' sin(nt))$ is pointwise convergent at each $t \in [0, 2\pi]$.
Let $K_n = a_n(Y_n cos(nt) + Y_n' sin(nt))$. Clearly $EK_n = 0, EK_n^2 = a_n^2$.
Then using Chebyshev's Inequality (take $1 = (a_n)(a_n)^{-1}$), $\sum\limits_{n=0}^{\infty} P\{|K_n| \geq 1\} \leq \sum\limits_{n=0}^{\infty} a_n^2 < \infty$.
Define $Z_n = K_n \mathbb{1}_{\{|K_n| \leq 1\}}$.
Check that $\sum\limits_{n=0}^{\infty} EK_n < \infty$ and $\sum\limits_{n=0}^{\infty} VarZ_n < \infty$.
This means that Result 2.1.7 is applicable and $\sum\limits_{n=0}^{\infty} K_n < \infty$ a.s.
$\therefore X_t$ in (11.1.1) is well defined almost surely for each t $\in [0, 2\pi]$, ie, X_t is defined on $\Omega \backslash N_t$ where P(N_t) = 0.

Let T be a countable dense subset of $[0, 2\pi]$. Then, X_t is defined for each t in T, on $\Omega \backslash \underset{t \in T}{\cap} N_t$ where P($\underset{t \in T}{\cap} N_t$) $\leq \sum\limits_{t \in T} P(N_t) = 0$. Therefore, we can define a centred Gaussian process X on T, with X(t) = X_t which is well-defined a.s.

Claim 3: X is continuous at each t in T.
Since T is a compact set in $\mathbb{R}$, we can use Proposition 9.1.1. Check that the covariance function of this process R(s,t) = $\sum\limits_{n=1}^{\infty} a_n^2 cos(n(t-s))$.
Then,

$$p^2(u) = 2 \sup_{0 \leq t \leq u} |R(0) - R(t)| = 4 \sup_{0 \leq t \leq u} \sum_{n=1}^{\infty} a_n^2 (sin^2(\frac{nt}{2}))$$

.

Claim 3.1: $\sum\limits_{n=0}^{\infty} \frac{p(2^{-n})}{\sqrt{n}} < \infty$

1) First prove the bound $p^2(2^{-n}) \leq \frac{a_1^2}{2^{2n}} + 4\sum_{j=0}^{n} \frac{2^{2j}}{2^{2n}} A(2^j, 2^{j+1}) + B(2^n)$ where $A(m, n) = \sum_{j=m+1}^{n} a_j^2$, and B(n)=A(n,$\infty$). This bound follows by splitting the series appropriately and the facts that $sin^2 x \leq min\{1, x^2\} \leq 1 \ \forall \ x$. Note that $(|x| + |y|)^{0.5} \leq |x|^0.5 + |y|^0.5$.
Then,

$$\sum_{n=0}^{\infty} \frac{p(2^{-n})}{\sqrt{n}} \leq \sum_{n=0}^{\infty} \frac{a_1}{2^n \sqrt{n}} + \sum_{n=0}^{\infty} \frac{\sqrt{4\sum_{j=0}^{n} \frac{2^{2j}}{2^{2n}} A(2^j, 2^{j+1})}}{\sqrt{n}} + \sum_{n=0}^{\infty} \frac{\sqrt{B(2^n)}}{\sqrt{n}}$$

2) The first term on the right side of the inequality is trivially finite. To prove the finiteness of the second term and third term, we need to use the convergence of (11.1.2). For the second term, we also need another inequality additionally. Let us assume that the second term's finiteness is proven and move ahead. For the third term's finiteness, use Cauchy Condensation Test.

$$\sum_{j=2}^{\infty}\left(\frac{(\sum_{n=j}^{\infty} a_n^2)^{\frac{1}{2}}}{j(logj)^{\frac{1}{2}}}\right) < \infty \iff \sum_{j=2}^{\infty} 2^j\left(\frac{(\sum_{n=2^j}^{\infty} a_n^2)^{\frac{1}{2}}}{2^j(log(2^j))^{\frac{1}{2}}}\right) = \frac{1}{(log2)^{\frac{1}{2}}}\sum_{j=2}^{\infty}\left(\frac{(B(2^j))^{\frac{1}{2}}}{j^{\frac{1}{2}}}\right) < \infty$$

Then, we are done with Claim 3.1.

Claim 3.2: $\sum_{n=0}^{\infty}\frac{p(2^{-n})}{\sqrt{n}} < \infty \implies \int_0^{\delta}(-logu)^{\frac{1}{2}}dp(u) < \infty$
Let

$$I = \int_0^{\delta} p(e^{-x^2})dx = \lim_{N\to\infty}\int_0^N p(e^{-x^2})dx = \lim_{N\to\infty}\int_0^N \frac{p(e^{-t})}{2\sqrt{t}}dt \le \lim_{N\to\infty}\int_0^N \frac{p(e^{-t})}{\sqrt{t}}.$$

As t increases, $\frac{1}{\sqrt{t}}e^{-t}$ decreases $\implies$ $\frac{1}{\sqrt{t}}p(e^{-t})$ decreases.
Then,

$$I \le \lim_{N\to\infty}\sum_{0}^{N-1}\frac{p(e^{-n})}{\sqrt{n}} \le \lim_{N\to\infty}\sum_{0}^{N-1}\frac{p(2^{-n})}{\sqrt{n}} = \sum_{n=0}^{\infty}\frac{p(2^{-n})}{\sqrt{n}}$$

. The Claim 3.2 then follows.

Now, from Claim 3.2, Claim 3.1, and Proposition 9.1.1, Claim 3 follows.

From Theorem 6.2.3, since X is a.s continuous at each t in T, X is uniformly continuous over T a.s. This implies X uniformly converges over T. Since T is a dense set in $[0, 2\pi]$, the process X as in (11.1.1) converges uniformly over $[0, 2\pi]$ a.s. Hence proved.

□

- Notice how the entropy arguments that we used in proving Proposition 9.1.1 helps in determining the convergence of such a complicated looking random series in the above proof, which also has implications in other fields.

Before ending this section, an interesting topic, the 'Talagrand Exapnsion' is briefly introduced below. However, it does not directly depict a use of the entropy-based results and hence does not find a place in this Part as a chapter/section. After the below remark, the report describes some other kinds of processes that has not been given much thought so far in this report, and shows how the Main Entropy result comes handy there too.

Remark 11.1.1. The Talagrand Expansion

Just as a Fourier Series can be visualised as a basis representation of a function in the appropriate space, a Gaussian Fourier Series may also be viewed as an orthonormal decomposition of a Gaussian process over a finite interval.

The question that arises then is, does any general Gaussian process have such a decomposition? If no, under what conditions can a process be expected to have such a decomposition. The answer to these questions are: a Gaussian process X has such a decomposition if it is continuous over T.

In fact, the below result (due to Talagrand) makes use of these expansions to arrive at a conclusion that characterises continuous Gaussian processes, and also provides a way to construct continuous Gaussian processes. This result enables one to no longer need to find an orthonormal basis representation, which makes it easier in cases where finding such representations explicitly might be time consuming and/or extremely difficult.

Result 11.1.2. Let X be a centred Gaussian process on a compact metric space T. Then X is continuous $\iff$ X has a continuous covariance function and there exists a centred Gaussian sequence $\{Y_n\}$ with variances $\sigma^2(Y_n)$ such that $\forall t, X_t := \sum_{n=0}^{\infty} \alpha_n(t) Y_n$ with the following conditions satisfied:
- $lim_{n \longrightarrow \infty} (log(n))^{\frac{1}{2}} \sigma(Y_n) = 0$
- For each t $\in$ T, $\alpha_n(t) \geq 0$ and $\sum_{n=0}^{\infty} \alpha_n(t) \leq 1$. (14.2.1)

Notice that there are not many assumptions on $\{Y_n\}$ and $\{\alpha_n\}$. Even basic assumptions seen in other places like independence of $\{Y_n\}$, or continuity of the functions $\{\alpha_n\}$ are not required here. Then, any centred Gaussian sequence and functions satisfying the two conditions given in (11.1.2) can be used to construct a continuous process.

The only if condition makes the result more stronger because it implies that ALL continuous processes can be built in the manner above.

11.2 Vector Valued Processes

In this entire report, we have only considered real-valued Gaussian processes. What happens when it comes to a vector-valued processes.

Let X_t=$< X_t^1, X_t^2, ..., X_t^n >$ be an $\mathbb{R}^n$-valued Gaussian process on a totally bounded separable metric space (T, τ). Then, each X_t^i is a real-valued Gaussian process and there is a covariance 'matrix' function $R(s,t)$ whose elements $R_{ij}(s,t) = EX_s^i X_t^j$.

Calculus of several variables establishes without doubt that a vector valued function is continuous if and only if it's component functions are continuous. Using this, we can reduce the analysis of sample path continuity of processes of the above kind to analysing continuity of the real-valued component processes.

Then, the Main Entropy Result 3.2.1 again helps here because we can use it to determine continuity in each component process similar to the arguments covered in the last few chapters and hence determine continuity or something else about the vector-valued process.

11.3 Non-Gaussian Processes

Going a step ahead, can we use any of the tools introduced in this report to analyse non-Gaussian processes?

Suppose Y is a centred Gaussian process on T, and X is a stochastic process defined as $X_t = f(Y_t)$ where f is a continuous function defined appropriately. In general, X may then be a non-Gaussian process. However, since we have the tools to analyse continuity of Y, and f is continuous; we can essentially analyse the continuity of X with the same tools.

Basically the continuity of X comes down to the continuity of Y, which we are more familiar with.

What about processes that have no relationship to any Gaussian process whatsoever?

Interestingly, a result has been proved which gives a sufficient condition for any Banach space-valued stochastic process defined on a metric space. This result also requires a bound on the 'metric-entropy function' apart from other conditions. Though this metric entropy function is not defined using any canonical metric similar to what was defined here, the arguments used in proving the finiteness of the integral in Theorem 3.2.1 for Gaussian processes do come handy there. This also adds importance to the development of the General Theory for Gaussian processes.

Summary and Conclusion

The most fundamental tool of the modern approach towards a general theory to study all Gaussian processes at once, is the canonical metric, d. It is a custom-tailored tool that enables navigation around the geometry of the parameter space and focuses only on the 'size'. To measure this 'size' that it sees, we require two concepts, entropy and majorising measures.

When it comes to sample path continuity, the entropy arguments seem simplest. It is more easily applicable to several processes with very different properties at once. It also makes the analysis of some specific cases simpler than methods relying on the geometry of the parameter space. This is especially recognisable in the case of generalised random fields and set-indexed processes. Despite providing only a sufficient condition, the Entropy result establishes stronger results and even gives some information about boundedness in certain specific cases. However, majorising measures are more powerful in their analysis of both sample path continuity and boundedness. This is not a non-trivial observation, considering that a majorising measure-based result is what helps prove the Main Entropy Result. One of its results validates that we can disregard a whole bunch of processes when looking for continuity or boundedness (those that don't have totally bounded parameter spaces) which appears most interesting.

To think about the distribution of suprema, the Borell's inequality is powerful because it rigorously proves the asymptotic equivalence of the supremum of a process to a normal distribution with the variance equal to the highest variance across the process, which is something very intuitive and readily appreciable. When it comes to comparison inequalities, Kahane's inequality to is most important comparison inequality because it effectively implies Sudakov-Fernique, Slepian's, other similar inequalities, and hence whatever they imply.

Given everything, the general theory is indeed promising. Construction of majorising measures in a general sense is not known yet. Better understanding majorising measures could prove more stronger results in specific cases and provide more intuition behind its working. Analysing continuity of real-valued Gaussian processes also allows analysis of many non-Gaussian processes. A similar attempt to connect boundedness may also be possible. Hence, the study of general Gaussian processes is has far-reaching implications in different areas of science, mathematics and statistics.

References

[1] Robert J. Adler. An Introduction to Continuity, Extrema, and Related Topics for General Gaussian processes, volume 12. Institute of Mathematical Statistics, 1990.

[2] Robert J. Adler. The Geomtery of Random Fields. Society for Industrial and Applied Mathematics, 2010.

[3] L. Decreusefond and D. Nualari. Hitting times for gaussian processes. The Annals of Probability, 36:pp. 319–330, 2008.

[4] Lee H. Koltun V. Ghiasi, G. Exploring the structure of a real-time, arbitrary neural artistic stylization network. In Proceedings of the IEEE International Conference on Computer Vision, pages pp. 5587–5596, 2017.

[5] Klarner L. Moss H. et al. Griffiths, R. Gauche: A library for gaussian processes in chemistry. 2023.

[6] Ryan-Rhys Griffiths. Applications of gaussian processes at extreme lengthscales: From molecules to black holes. https://arxiv.org/abs/2303.14291, 2022.

[7] Tan X. Heenan M. Hou, E. and D. Wen. A global dataset of plant available and unavailable phosphorus in natural soils derived by hedley method. Scientific Data, 5, 2018.

[8] Shen J. Zhang Y. Li, Y. Adversarial ranking for language generation. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 1:204–213, 2017.

[9] Mikhail Lifshits. Lectures on Gaussian Processes. SpringerBriefs in Mathematics (BRIEFSMATH), 2012.

[10] J.G. Powles. Brownian motion. Physics Education, 13:310–312, 1978.

[11] Goyal A. Bender G. Zamir A. R. Ramesh, A. Zero-shot text-to-image generation. arXiv preprint arXiv:2102.12092.

[12] Jonathan E. Taylor Robert J. Adler. Random Fields and Geometry. Springer Science, 2007.

[13] Rolf Steye and Werner Nagel. Probability and Conditional Expectation. John Wiley and Sons Ltd, 2017.

[14] Petar Todorovic. An Introduction to Stochastic Processes and Their Applications. Springer-Verlag New York, 1992.