NEW VERSION OF OPTIMAL STOPPING PROBLEM

by

Wai-Lun Lam

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Applied Mathematics

Charlotte

2024

Approved by:

____

Dr. Stanislav Molchanov

____

Dr. Isaac Sonin

____

Dr. Zhiyi Zhang

____

Dr. Michael Grabchak

____

Dr. Paul Gaggl

ABSTRACT

WAI-LUN LAM. New Version of Optimal Stopping Problem. (Under the direction of DR. STANISLAV MOLCHANOV)

This dissertation contains several new results concerning Moser-type optimal stopping problems. In the simplest case we consider sequence of independent uniformly distributed points $X_1, X_2, \cdots, X_n$ on the compact Riemannian manifold $\mathcal{M}$ and give algorithm for the calculation of $S_n = \max_{\tau \le n} E[\mathcal{G}(X_\tau)]$ where $\mathcal{G}$ is a smooth function on $\mathcal{M}$ and τ is a random optimal stopping time. Description of the optimal τ depends on the structure of $\mathcal{G}$ near points of maximum. For different assumptions on this structure we calculate asymptotics of S_n.

DEDICATION

I dedicate this dissertation to my beloved mother, Chui-Kuen Leung, whose unwavering love, guidance, and sacrifices have shaped me into the person I am today. Her boundless wisdom and unfaltering support have been my guiding light through life's challenges. Without her endless encouragement and patience, I would not have achieved this milestone. This achievement unquestionably belongs not only to me, but to both of us. Though she resides among the stars now, her spirit and love continue to inspire me every day. With heartfelt gratitude and love, I dedicate this accomplishment to her, wishing her eternal happiness and peace.

ACKNOWLEDGEMENTS

I am truly grateful to have Dr. Stanislav Molchanov as my advisor on this academic journey. Dr. Molchanov is not only an exceptional mathematician but also an excellent mentor. His ability to simplify complex problems and his remarkable flexibility in thinking always amazes and inspires me. Above all, I appreciate his patience and support. Without his guidance and encouragement, completing this dissertation would undoubtedly have been a far more challenging endeavor. I will miss the days when I could attend his lectures and have mathematical conversations with him in his office.

I extend my sincere appreciation to Dr. Michael Grabchak for his invaluable teachings in probability theory. His openness and willingness to engage in discussions about mathematics, as well as matters concerning my career development, have been invaluable to me. His patience, rigor, and dedication have helped me equip with the necessary tools to face future challenges.

Special gratitude is extended to Dr. Isaac Sonin, an expert in Markov chain and decision theory, whose insights have significantly enriched my comprehension of the mathematical principles underlying decision-making processes.

I would also like to express my gratitude to Dr. Zhiyi Zhang and Dr. Paul Gaggl for agreeing to be members of my dissertation committee, and to the faculty of the mathematics department for their unwavering support and guidance.

Finally, I am indebted to my wife, my sister, and my brother-in-law for their loving support and encouragement every day of my life, accompanying me through the most difficult times I've faced. Their presence and encouragement have been my wellspring of strength, propelling me to overcome every obstacle encountered along the way.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

The Moser problem is a classic problem of optimal stopping theory. This branch of probability and decision theory focuses on determining the optimal strategy for making decisions in a sequential manner. The Moser problem, named after Leo Moser who proposed it in 1956, presents a captivating scenario where an individual must decide when to stop a sequential process in order to maximize the expected reward.

In Moser problem, the decision-maker is confronted with a sequence of options, each associated with a certain reward or penalty. The challenge lies in determining the optimal stopping rule - the point in the sequence at which the decision-maker should halt the process to attain the maximum expected reward. This problem is characterized by its simplicity in formulation but complexity in finding an optimal solution, making it an intriguing problem in the realm of decision theory.

Moser problem helps exploring the underlying principles of optimal stopping, seeking general strategies and insights that can be applied to a broader class of problems. It serves as a valuable case study, contributing to our understanding of decision-making under uncertainty and offering practical applications in diverse fields, including finance, operations research, and artificial intelligence. Analyzing the Moser problem provides a glimpse into the intricate balance between exploration and exploitation, shedding light on the delicate trade-offs inherent in sequential decision-making processes.

1.1 Motivation for Moser problem

Let's formulate the Moser problem. Suppose a gambler possesses n opportunities to randomly draw a number from the interval $[0, 1]$. After each draw, the gambler examines the number drawn. If the number is rejected, the gambler has the option to draw again from the remaining $n-1$ chances. This process repeats until the gambler decides to stop drawing, at which point he/she receives the value of the last drawn number. The question is how should the gambler achieve the maximum mean value in this game?

To translate this problem to a mathematical setting. Let $X_1, \cdots, X_n$ be i.i.d. random variables uniformly distributed on $[0, 1]$ and τ be the stopping times for this sequence such that $\forall k \geq 1,\ \{\tau = k\} \in \mathcal{F}_k = \sigma(X_1, \cdots, X_k)$. We are interested in the maximum expected reward

$$S_n = \max_{\tau \leq n} E[X_\tau].$$

By using the Bellman's principle one can find the recursive relation

$$\begin{cases} S_{n+1} = \frac{1+S_n^2}{2},\ n \geq 1 \\ S_1 = \frac{1}{2}. \end{cases}$$

Then by applying the standard formulas for the asymptotics of the iterations $x_{n+1} = g(x_n)$ with appropriate conditions on $g(x)$, one can prove that

$$S_n = 1 - \frac{1}{n} + o(\frac{1}{n}).$$

See details in [1], [2].

More general problem for i.i.d. random variables (not necessarily uniformly distributed but supported on the finite interval, say $[0, L]$) with continuous positive density $f(x)$ on $[0,1]$ has the similar form. One can find

$$S_n = \max_{\tau \le n} E[X_\tau], \quad \{\tau = k\} \in \sigma(X_1, \cdots, X_k)$$

Again the Bellman's principle gives the recursive formula

$$\begin{cases} S_{n+1} = H(S_n), \ S_1 = E[X_1] \\ H(x) = \int_x^L z f(z) dz + x \int_0^x f(z) dz. \end{cases}$$

It can be proved that as $n \to \infty$, the sequence S_n monotonically increases towards L, i.e. $S_n \uparrow L, \ n \to \infty$. Moreover, under some regularity condition on $f(x)$ near $x = L$, one can find asymptotics of $S_n, \ n \to \infty$. Simplest regularity condition

$$f(x) \sim c(L-x)^\alpha \mathcal{L}\left(\frac{1}{L-x}\right), \ x \uparrow L$$

where $\alpha > -1$ and $\mathcal{L}(z)$ is slowly varying function if $z \to +\infty$.

Furthermore, the first publication addressing the Moser problem with unbounded random variables is attributed to Karlin [3]. Let $X_1, \cdots, X_n$ are i.i.d. exp(1) random variables, i.e. $P\{X_1 > x\} = e^{-x}$. Then

$$\begin{cases} S_{n+1} = S_n + e^{-S_n} \\ S_n = \ln n + \frac{1}{n} + o(\frac{1}{n}), \ n \to \infty. \end{cases}$$

See details in appendix A. From this point on, when addressing the Moser problem in the context of non-uniform probability distributions, we will refer to it as the Moser-type problem. The term "Moser problem" will be reserved for scenarios involving uniformly distributed random variables on $[0,1]$.

In the subsequent chapters, we'll study the scenarios involving random variables $X_1, \cdots, X_n$ characterized by distributions with an unknown parameter and involve an atom. Subsequently, we explore scenarios concerning the maximum expected value within an open set of a compact Riemannian manifold with a special function.

CHAPTER 2: MOSER-TYPE PROBLEMS

2.1 Non-stationary Moser-type problem

New results in this area of Moser-type problem concern the situation of the reward function $\mathcal{G} : \mathcal{M} \to \mathbb{R}$ occurs on an open set of a compact Riemannian manifold $\mathcal{M}$. A nonnegative integer-valued random variable τ is called a stopping time if the event $\{\tau = k\} \in \mathcal{F}_k = \sigma(X_1, \cdots, X_k)$. Our interest lies in solving the optimal stopping problem defined as:

$$\max_{\tau \leq n} E[\mathcal{G}(X_\tau)]$$

Here the maximum is taken over all stopping times τ that are less than or equal to n. For simplicity we focus on optimizing over stopping times of the form:

$$\tau = \min\{1 \leq i \leq n : \mathcal{G}(X_i) \geq h_i\}$$

where $1 > h_1 > h_2 > \cdots > h_n > 0$ are the predetermined thresholds. In this case the goal is to find the optimal choice of h_i's. This optimization problem is sometimes denoted as:

$$S_n = \max_{\{h_i\}} E[\mathcal{G}(X_\tau)].$$

Consider the simplest example when $\mathcal{M} = [0,1]$. Suppose that $X_1, \cdots, X_n$ are i.i.d. random variables uniformly distributed on interval $[0,1]$ and let $\mathcal{G} : [0,1] \to [0,1]$ be

$$\mathcal{G}(x) = \begin{cases} \frac{x}{1-\epsilon} & , \ 0 \le x < 1-\epsilon \\ 1 & , \ 1-\epsilon \le x \le 1 \end{cases} \tag{2.1}$$

where $0 < \epsilon << 1$. Since such a function resembles a plateau, let's just call this the plateau reward function. It worth to note that the plateau function is more general than a linear function. Notice that if we set $\epsilon = 0$, the problem will be reduced back to the classical Moser problem. Let's fix some threshold for each step $1 > h_1 \ge h_2 \ge \cdots \ge h_n > 0$ and let

$$Y_1 = \mathcal{G}(X_1), \cdots, Y_n = \mathcal{G}(X_n)$$

be i.i.d. random variables distributed on $[0,1]$. The probability distribution of Y is

$$P\{Y \le y\} = \begin{cases} y(1-\epsilon) & , \ 0 \le y < 1 \\ 1 & , \ y = 1 \end{cases}$$

and the corresponding probability density is

$$f(y) = \begin{cases} 1-\epsilon & , \ 0 \le y < 1 \\ \epsilon\delta(y-1). \end{cases}$$

where $\delta(\cdot)$ is the delta function.

Lemma 2.1. Let $X_1, \cdots, X_n$ be i.i.d. random variables uniformly distributed on the interval $[0,1]$. Let $\mathcal{G} : [0,1] \to [0,1]$ such that

$$\mathcal{G}(x) = \begin{cases} \frac{x}{1-\epsilon} & , 0 \leq x < 1-\epsilon \\ 1 & , 1-\epsilon \leq x \leq 1 \end{cases}$$

where $0 < \epsilon << 1$. Let $Y_1 = \mathcal{G}(X_1), \cdots, Y_n = \mathcal{G}(X_n)$ be i.i.d. random variables distributed on $[0,1]$. Let $\tilde{\tau} = \min\{t : Y_t = 1\}$, then $\tilde{\tau}$ is geometric distributed, i.e. $P\{\tilde{\tau} = k\} = (1-\epsilon)^{k-1}\epsilon, k \geq 1$.

Proof. Let $0 < \epsilon << 1$ and fix $k \geq 1$, then

$$\begin{aligned} P\{\tilde{\tau} = k\} &= P\{Y_1 < 1, Y_2 < 1, \cdots, Y_k = 1\} \\ &= P\{Y_1 < 1\}P\{Y_2 < 1\} \cdots P\{Y_k = 1\} \\ &= (1-\epsilon)^{k-1}\epsilon. \end{aligned}$$

□

Lemma 2.1 suggests that the mean time of hitting the plateau 1 is of order $\frac{1}{\epsilon}$. This gives us a sense of time when the random variable will hit the plateau.

Now let's turn our attention to the calculation of the maximum expectation, i.e, $S_n = \max\limits_{\tau \leq n} E[Y_n]$, with the plateau function in (2.1). By the law of total expectation of Y, we have

$$\begin{aligned} S_n &= \max_{\tau \leq n} \Big(E[Y_n | Y_n \geq h_n] P\{Y_n \geq h_n\} + E[Y_n | Y_n < h_n] P\{Y_n < h_n\} \Big) \\ &= \max_{\tau \leq n} \left(\int_{h_n}^{1} y \left(1 - \epsilon + \epsilon\delta(y-1)\right) dy + S_{n-1} h_n (1-\epsilon) \right) \\ &= \max_{\tau \leq n} \left(\left(\frac{1-h_n^2}{2} \right) (1-\epsilon) + \epsilon \cdot 1 + S_{n-1} h_n (1-\epsilon) \right). \end{aligned} \tag{2.2}$$

Then take the derivative of above equation with respect to h_n and set zero implies

$$h_n = S_{n-1}.$$

Replace all the h_n in equation (2.2) to be S_{n-1}, we have

$$S_n = (1-\epsilon)\left(\frac{1+S_{n-1}^2}{2}\right) + \epsilon$$

Now rewrite the recursive relation as a function

$$g(x) = (1-\epsilon)\left(\frac{1+x^2}{2}\right) + \epsilon.$$

By the fix point theorem, $g(x) = x$ gives the solution $x = 1$. Since $g'(1) < 1$, g is contractive and $g(S_n) \to 1$ as $n \to \infty$. Let

$$h_n = 1 - g(S_n)$$

implies

$$1 - h_n = (1-\epsilon)\left(\frac{1+(1-h_{n-1})^2}{2}\right) + \epsilon$$

implies

$$h_n = (1-\epsilon)h_{n-1} - (1-\epsilon)\frac{h_{n-1}^2}{2}. \tag{2.3}$$

In the classical Moser problem, one may apply the Pólya and Szëgo theorem (see details in [4] problem 174 on p.38) to equation (2.3) to get the asymptotic solution of h_n. Since in this case the coefficient of h_{n-1} is $1-\epsilon$, we cannot directly apply this theorem on equation (2.3). Let's analyze equation (2.3) with the following two cases:

Case 1. To estimate h_n from above, let $\epsilon n >> 1$. From (2.3), since

$$h_n \leq (1-\epsilon)h_{n-1}$$

implies

$$h_n \leq (1-\epsilon)^n \leq e^{-\epsilon n}.$$

Then

$$\begin{aligned} S_n &= 1 - h_n \\ &\geq 1 - (1-\epsilon)^n \\ &\geq 1 - e^{-\epsilon n}. \end{aligned}$$

Case 2. To estimate h_n from below, let $\epsilon n << 1$. Since

$$h_n \geq h_{n-1} - \frac{h_{n-1}^2}{2}$$

By the classical Moser problem, we have

$$h_n < \frac{2}{n}.$$

Then

$$S_n \geq 1 - \frac{2}{n}.$$

The above calculation indicates that there is a phase transition region between the two cases.

Lemma 2.2. Let $X_1, \cdots, X_n$ be i.i.d. uniformly distributed random variables on the interval [0,1]. Without loss of generality, let the maximum value of the smooth function $\mathcal{G}$ be 1. Furthermore, let $\mathcal{G} : [0,1] \to [0,1]$ such that $\mathcal{G}(x) \sim 1 - c(1-x)^{\beta}$ where $c > 0, \beta > 0$ and

$$Y_1 = \mathcal{G}(X_1), \cdots, Y_n = \mathcal{G}(X_n)$$

be i.i.d. random variables. Then the maximum expectation is

$$S_n \sim 1 + \frac{A}{n^{\beta}}$$

where A is a constant depends on c and β.

Proof. The probability distribution of Y is

$$\begin{aligned}
P\{Y \geq y\} &= P\{\mathcal{G}(X) \geq y\} \\
&= P\{1 - c(1-X)^{\beta} \geq y\} \\
&= P\left\{X \geq 1 - \left(\frac{1-y}{c}\right)^{1/\beta}\right\} \\
&= 1 - \left(1 - \left(\frac{1-y}{c}\right)^{1/\beta}\right) \\
&= \left(\frac{1-y}{c}\right)^{1/\beta}
\end{aligned}$$

The probability density is

$$\begin{aligned}
f(y) &= \frac{d}{dy} P\{Y \leq y\} \\
&= \frac{d}{dy}\left(1 - \left(\frac{1-y}{c}\right)^{1/\beta}\right) \\
&= \frac{1}{c\beta}\left(\frac{1-y}{c}\right)^{\frac{1}{\beta}-1}
\end{aligned}$$

One can calculate the maximum expectation with change of variables.

$$\begin{aligned}
S_n &= \max_{\tau\le n}\left(\int_{h_n}^{1} yf(y)dy + S_{n-1}P\{Y\le h_n\}\right)\\
&= \max_{\tau\le n}\left(\int_{h_n}^{1} \frac{y}{c\beta}\left(\frac{1-y}{c}\right)^{\frac{1}{\beta}-1} dy + S_{n-1}\left(1-\left(\frac{1-h_n}{c}\right)^{1/\beta}\right)\right)\\
&= \max_{\tau\le n}\left(\frac{1}{c^{\frac{1}{\beta}}}\left((1-h_n)^{\frac{1}{\beta}} - \frac{(1-h_n)^{\frac{1}{\beta}+1}}{1+\beta}\right) + S_{n-1}\left(1-\left(\frac{1-h_n}{c}\right)^{1/\beta}\right)\right) \qquad (2.4)
\end{aligned}$$

then take derivative of above equation and set zero implies

$$h_n = S_{n-1}.$$

Now substitute this back to equation (2.4), we have

$$S_n = \frac{1}{c^{\frac{1}{\beta}}}\left((1-S_{n-1})^{\frac{1}{\beta}} - \frac{(1-S_{n-1})^{\frac{1}{\beta}+1}}{1+\beta}\right) + S_{n-1}\left(1-\left(\frac{1-S_{n-1}}{c}\right)^{1/\beta}\right).$$

Now rewrite the recursive relation as a function

$$g(x) = \frac{1}{c^{\frac{1}{\beta}}}\left((1-x)^{\frac{1}{\beta}} - \frac{(1-x)^{\frac{1}{\beta}+1}}{1+\beta}\right) + x\left(1-\left(\frac{1-x}{c}\right)^{1/\beta}\right)$$

by fixed point theorem, $g(x) = x$ gives the solution $x = 1$.

Since $g'(1) = 1$, g is contractive and $g(S_n) \to 1$ as $n \to \infty$. Let

$$h_n = 1 - g(S_n)$$

implies

$$\begin{aligned}
1 - h_n = \frac{1}{c^{\frac{1}{\beta}}}&\left((1-(1-h_{n-1}))^{\frac{1}{\beta}} - \frac{(1-(1-h_{n-1}))^{\frac{1}{\beta}+1}}{1+\beta}\right)\\
&+ (1-h_{n-1})\left(1-\left(\frac{1-(1-h_{n-1})}{c}\right)^{1/\beta}\right)
\end{aligned}$$

implies

$$h_n = h_{n-1} - \frac{\beta c^{-\frac{1}{\beta}}}{1+\beta} h_{n-1}^{1+\frac{1}{\beta}}$$

Now let $k = 1 + \frac{1}{\beta}$ and $a = \frac{\beta c^{-\frac{1}{\beta}}}{1+\beta}$. Then by the Pólya and Szëgo theorem,

$$n^{\beta} h_n \to \left(\frac{\beta c^{-\frac{1}{\beta}}}{1+\beta} \cdot \frac{1}{\beta} \right)^{-\frac{1}{1/\beta}}.$$

That is

$$h_n \to c \left(\frac{1+\beta}{n} \right)^{\beta}$$

as $n \to \infty$. Then the maximum expectation becomes

$$S_n \sim 1 + c \left(\frac{1+\beta}{n} \right)^{\beta}$$

when n is large. □

2.2 Probability distribution with incomplete information

When the distribution of the random variables has an unknown parameter, one can apply statistical method such as maximum likelihood estimation to estimate it. The maximum likelihood method provides us a way to develop the sense of stopping in the game.

Let $X_1, \cdots, X_n$ be i.i.d. uniform random variables on interval $[0, a]$ where a is an unknown positive constant. Since the player does not have any information at all when the game starts, then he/she should always observe X_1. To establish the sense of stopping, the maximum likelihood estimation can be applied on unknown a. Let

$1 \leq m \leq n$, then the log-likelihood function can be written as

$$L(a|X_1, \cdots, X_m) = \prod_{i=1}^{m} f(x_i|a) = \frac{1}{a^m}.$$

Then the log-likelihood becomes

$$\log L(a|X_1, \cdots, X_m) = -m \log a.$$

Then the derivative of the log-likelihood function is

$$\frac{d}{da} \log L(a|X_1, \cdots, X_m) = -\frac{m}{a}.$$

Since the derivative is a monotone decreasing function, the estimated parameter is

$$\hat{a}_m = \max(X_1, \cdots, X_m).$$

Let $\hat{M}_m = \frac{m}{m-1}\hat{a}_m$. Since $E\left[\hat{M}_m\right] = E\left[\frac{m}{m-1}\hat{a}_m\right] = a = M_m$, an unbiased estimator of the maximum is

$$\hat{M}_m = \frac{m}{m-1}\hat{a}_m.$$

The concept of employing the maximum likelihood method prompts us to iteratively refine the unbiased estimation of parameter a at each step of the process. This iterative approach allows for continual improvement in the accuracy of our estimation as more data is collected, resulting in a more robust and reliable estimation of the parameter a over time. However, it worth to note that a fundamental trade-off arises between augmenting the dataset to refine the estimate and allocating resources for optimal stopping. This dilemma gives rise to the multi-armed bandit problem.

Let's calculate the estimated distribution of X. For each step m,

$$P\{X \leq x\} = \frac{x}{\hat{a}_m}$$

and the corresponding density is

$$f(x) = \frac{1}{\hat{a}_m}\mathbf{1}(0 \leq x \leq \hat{a}_m).$$

Then the maximum expectation becomes

$$\begin{aligned} S_n &= \max_{\tau \leq n}\Big(E[X_n|X_n \geq h_n]P\{X_n \geq h_n\} + E[X_n|X_n < h_n]P\{X_n < h_n\}\Big) \\ &= \max_{\tau \leq n}\left(\int_{h_n}^{\hat{a}_n} \frac{x}{\hat{a}_n}dx + \frac{h_n}{\hat{a}_n}S_{n-1}\right) \\ &= \max_{\tau \leq n}\left(\frac{\hat{a}_n^2 - h_n^2}{2\hat{a}_n} + \frac{h_n}{\hat{a}_n}S_{n-1}\right). \end{aligned} \tag{2.4}$$

Take the derivative of the above equation with respect to h_n and set zero implies

$$-\frac{h_n}{\hat{a}_n} + \frac{S_{n-1}}{\hat{a}_n} = 0$$

implies

$$h_n = S_{n-1}.$$

Substitute this result back to (2.4), then

$$S_n = \frac{\hat{a}_n^2 + S_{n-1}^2}{2\hat{a}_n}.$$

Now rewrite the recursive relation as a function

$$g(x) = \frac{\hat{a}_n^2 + x^2}{2\hat{a}_n}$$

by fixed point theorem, $g(x) = x$ gives

$$x = \frac{\hat{a}_n^2 + x^2}{2\hat{a}_n}$$

and the solution is

$$x = \hat{a}_n.$$

Since $g'(\hat{a}_n) \leq 1$, g is contractive and $g(S_n) \to \hat{a}_n$ as $n \to \infty$. Let

$$h_n = \hat{a}_n - g(S_n)$$

implies

$$\hat{a}_n - h_n = \frac{\hat{a}_n^2 + (\hat{a}_n - h_{n-1})^2}{2\hat{a}_n}$$

implies

$$h_n = h_{n-1} - \frac{h_{n-1}^2}{2\hat{a}_n}.$$

Now let $k = 2$ and $a = \frac{1}{2\hat{a}_n}$. Then by Pólya and Szëgo theorem,

$$nh_n \to 2\hat{a}_n \;\; as \;\; n \to \infty$$

implies he following asymptotic relationship when n is large

$$S_n \sim \hat{a}_n \left(1 - \frac{2}{n}\right)$$

that is $S_n \to \hat{a}_\infty$ as $n \to \infty$.

2.3 Probability distribution with an atom

The inclusion of atoms in probability distributions within decision-making theory traces back to the mid-20th century, with seminal contributions from mathematicians such as Leonard J. Savage [5], [6]. His work laid the groundwork for understanding decision-making under uncertainty, highlighting the importance of considering rare events or extreme outcomes in probabilistic models to better reflect real-world scenarios.

Here we discuss about one simple scenario with distribution with an atom. Consider $X_1, \cdots, X_n$ are i.i.d. random variables with density contains an atom of unknown mass π_0 such that

$$f(x) = \begin{cases} \pi_0 \delta(x - \frac{a}{2}) \\ 1 - \pi_0 & , \; x \in [0, \frac{a}{2}) \cup (\frac{a}{2}, a] \end{cases}$$

where a is an unknown constant. In this case, since the location of the atom is unknown, large sample size is the key to reveal the location of the atom.

Suppose this game is repeated one million times with mass of atom $\pi_0 = \frac{1}{2}$, then $(1 - \frac{1}{2})^{10^6}$ is the probability that the outcome does not hit the atom. This probability is extremely small. In contrast, the probability of getting the atom is extremely high. The outcomes will occur within $\frac{a}{2} \pm \sqrt{10^6(\frac{1}{2})(1 - \frac{1}{2})}$ which is a narrow region. It means that if the player see an exact outcome occurs twice or repeatedly, he/she can be sure that is the atom and the time for the atom appears repeatedly is called the collision time.

If the mass of an atom is very small, then the effect is negligible and the situation reduces back to the previous examples.

2.4 Stationary Moser-type problem

In Moser problem, the optimal strategy depends on the time interval, that is the number of the random variables in the sequence $X_1, \cdots, X_n$. One can consider a similar model with stationary strategy. That means one can consider fixing a single threshold h for the game instead of having a sequence of thresholds.

Let's illustrate the stationary Moser problem in more details. Consider for each step $t = 1, \cdots, n$, a judge of the game will flip a coin. Let $0 << \delta < 1$. With probability δ, the judge would give a "green light" for the player to continue the game. With probability $1-\delta$ the judge would end the game and the player receives zero reward. Let's fix a level h, the player would cash in if $X_i \geq h$, otherwise he/she would continue the game. Then, by the law of total expectation, the expected reward in each step is

$$S(\delta, h) = (1-\delta)\cdot 0 + \delta \cdot \left[\int_h^\infty x f(x)dx + F(h)\cdot S(\delta, h)\right]$$

where $F(h) = P\{X \leq h\}$, then it implies

$$S(\delta, h) = \frac{\delta \int_h^\infty x f(x)dx}{1-\delta F(h)}$$

To maximize the expectation in each step over the level h, it is necessary to take the derivative of $S(\delta, h)$ with respect to h and set it equals zero. Then the optimal level h is

$$h_{\text{opt}} = \frac{\delta}{(1-\delta)}\int_{h_{\text{opt}}}^\infty (1-F(x))dx \tag{2.5}$$

and the maximum expectation is

$$S(\delta) = \frac{\delta \int_{h_{\text{opt}}}^\infty x f(x)dx}{1-\delta F(h_{\text{opt}})}. \tag{2.6}$$

This computation provides a straightforward formula for determining both the optimal threshold and the maximum expectation of the game, showcasing its inherent elegance.

Example 2.1. Let $X_1, \cdots, X_n$ be i.i.d. exponential random variables with mean 1. Then by equation (2.5), the optimal level is

$$h_{\mathrm{opt}} = \frac{\delta e^{-h_{opt}}}{1-\delta}$$

That is

$$h_{\mathrm{opt}} = \ln \frac{1}{1-\delta} - \ln \ln \frac{1}{1-\delta} + o(1)$$

and since h_{opt} is large, by equation (2.6)

$$\begin{aligned}
S(\delta) &= \frac{\delta h_{\mathrm{opt}} e^{-h_{opt}}}{1-\delta(1-e^{-h_{\mathrm{opt}}})} \\
&= \frac{h_{\mathrm{opt}}^2(1-\delta)}{1-\delta+(1-\delta)h_{\mathrm{opt}}} \\
&= \frac{h_{\mathrm{opt}}^2}{1+h_{\mathrm{opt}}} \\
&= h_{\mathrm{opt}} \left(\frac{1}{1+\frac{1}{h_{\mathrm{opt}}}} \right) \\
&= h_{\mathrm{opt}} \left(1 - \frac{1}{h_{\mathrm{opt}}} + \frac{1}{h_{\mathrm{opt}}^2} + \cdots \right) \\
&= h_{\mathrm{opt}} + o(1).
\end{aligned}$$

That is

$$S(\delta) \sim h_{\mathrm{opt}}.$$

Example 2.2. Let $X_1, \cdots, X_n$ be i.i.d. uniform random variables on [0,1]. Then by equation (2.5), the optimal level is

$$h_{\text{opt}} = \frac{\delta}{1-\delta}\left(\frac{(1-h_{opt})^2}{2}\right)$$

That is

$$\begin{aligned} h_{\text{opt}} &= \frac{1}{\delta} + \sqrt{\frac{1}{\delta^2} - 1} \\ &\sim \frac{2}{\delta} \end{aligned}$$

and the maximum expectation is

$$\begin{aligned} S(\delta) &= \frac{\delta(1-h_{\text{opt}}^2)}{2(1-\delta h_{\text{opt}})} \\ &\sim h_{\text{opt}} - \frac{1}{h_{opt}}. \end{aligned}$$

CHAPTER 3: TECHNICAL TOOLS OF RIEMANNIAN GEOMETRY

On a compact Riemannian manifold, there is no global coordinate system. That means there is no one single coordinate system can cover the entire manifold. For example, consider a sphere, $\mathbb{S}^2$, in a three-dimensional Euclidean space. The cartesian coordinate system cannot cover the equator while the polar coordinate system cannot cover the north and south poles. In other words, singularities appear on every coordinate system on close surfaces. To over come this problem, multiple local coordinate systems may be used to cover the manifold. These covers with coordinate systems on top of them are called maps or charts. When more than one map are employed to cover the manifold, there will be some overlapping regions and the maps are required to be agreed on the same region that they cover.

3.1 Introduction to Riemannian geometry

Let $\mathcal{M}$ be a 2-dimensional compact Riemannian manifold and $\boldsymbol{\varphi}$ be a map (chart) from an open set $U \subset \mathbb{R}^2$ to $\mathcal{M}$. Consider a curve $\mathbf{r}(t) = (x_1(t), x_2(t)), t \in [a, b]$ on U and let $\boldsymbol{\gamma}(t) = \boldsymbol{\varphi}(x_1(t), x_2(t)), t \in [a, b]$ be a curve on $\mathcal{M}$. Then the magnitude of the velocity of a particle moving along the curve $\boldsymbol{\gamma}$ is

$$|\boldsymbol{V}(t)| = |\gamma'(t)| = \sqrt{(\boldsymbol{\varphi}' \cdot \boldsymbol{\varphi}')}(t) = \sqrt{(\boldsymbol{\varphi}_{x_1} x_1 + \boldsymbol{\varphi}_{x_2} x_2)) \cdot (\boldsymbol{\varphi}_{x_1} x_1 + \boldsymbol{\varphi}_{x_2} x_2))}(t)$$

$$= \sqrt{(\boldsymbol{\varphi}_{x_1} \cdot \boldsymbol{\varphi}_{x_1})(x_1')^2 + 2(\boldsymbol{\varphi}_{x_1} \cdot \boldsymbol{\varphi}_{x_2}) x_1' x_2' + (\boldsymbol{\varphi}_{x_2} \cdot \boldsymbol{\varphi}_{x_2})(x_2')^2}(t)$$

Note that $|\boldsymbol{V}(t)| > 0$, the reason for this is to ensure the curve is smooth at all points. Now let $E(x_x, x_2) = \boldsymbol{\varphi}_{x_1} \cdot \boldsymbol{\varphi}_{x_1}, F(x_x, x_2) = \boldsymbol{\varphi}_{x_1} \cdot \boldsymbol{\varphi}_{x_2}, G(x_x, x_2) = \boldsymbol{\varphi}_{x_2} \cdot \boldsymbol{\varphi}_{x_2}$, we define the first quadratic form of $\mathcal{M}$ as

$$ds^2 = Edx_1^2 + 2Fdx_1dx_2 + Fdx_2^2$$

and the arc length of the curve $\boldsymbol{\gamma}$ is

$$L(\boldsymbol{\gamma(t_0)}) = \int_a^{t_0} |\boldsymbol{V}(t)|dt = \int_a^{t_0} \frac{ds}{dt}dt = \int_a^{t_0} \sqrt{E(x_1')^2 + 2F(x_1'x_2') + G(x_2')^2}(t)dt$$

To establish a measure for area on the surface of the compact Riemannian manifold $\mathcal{M}$, we first fix a point, $\boldsymbol{\varphi}(x_1, x_2)$, then take the derivative of $\boldsymbol{\varphi}$ with respect to x_1 and x_2 to obtain the bases for the tangent plane at the point $\boldsymbol{\varphi}(x_1, x_2)$, that is , $\boldsymbol{\varphi}_{x_1}dx_1$ and $\boldsymbol{\varphi}_{x_2}dx_2$. Then by the parallelogram law, the infinitesimal area on M can be written as

$$dA(x_1, x_2) = |(\boldsymbol{\varphi}_{x_1}dx_1) \times (\boldsymbol{\varphi}_{x_2}dx_2)| = |\boldsymbol{\varphi}_{x_1} \times \boldsymbol{\varphi}_{x_2}|dx_1dx_2 = \sqrt{|\boldsymbol{\varphi}_{x_1} \times \boldsymbol{\varphi}_{x_2}|^2}dx_1dx_2$$

$$= \sqrt{(\boldsymbol{\varphi}_{x_1} \times \boldsymbol{\varphi}_{x_2}) \cdot (\boldsymbol{\varphi}_{x_1} \times \boldsymbol{\varphi}_{x_2})}dx_1dx_2 = \sqrt{\det \begin{bmatrix} \boldsymbol{\varphi_{x_1} \cdot \varphi_{x_1}} & \boldsymbol{\varphi_{x_1} \cdot \varphi_{x_2}} \\ \boldsymbol{\varphi_{x_1} \cdot \varphi_{x_2}} & \boldsymbol{\varphi_{x_2} \cdot \varphi_{x_2}} \end{bmatrix}}(x_1, x_2)dx_1dx_2$$

So the surface area is

$$A = \int_U \sqrt{EG - F^2}dx_1dx_2.$$

From the above calculation, let us generalize the compact Riemannian manifold formally. Let $(\mathcal{M}, g)$ be a compact Riemannian manifold where g is a positive-definite inner product, i.e.

$$g = \begin{bmatrix} E(x_1, x_2) & F(x_1, x_2) \\ F(x_1, x_2) & G(x_1, x_2) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\varphi_{x_1}} \cdot \boldsymbol{\varphi_{x_1}} & \boldsymbol{\varphi_{x_1}} \cdot \boldsymbol{\varphi_{x_2}} \\ \boldsymbol{\varphi_{x_1}} \cdot \boldsymbol{\varphi_{x_2}} & \boldsymbol{\varphi_{x_2}} \cdot \boldsymbol{\varphi_{x_2}} \end{bmatrix}$$

Let $\boldsymbol{\varphi}$ be a map from $U \subset \mathbb{R}^2$ to $\mathcal{M}$. Suppose the curve $\boldsymbol{\gamma}(t) \subset \mathcal{M}, t \in [a, b]$ and $|\boldsymbol{V}(t)| > 0$. Then the local coordinate $\boldsymbol{x} = (x_1, x_2)$, the Riemannian manifold is equipped with

(i) **The first quadratic form**:

$$ds^2 = g_{ij}(x^i, x^j)dx^i dx^j$$

(ii) **Arc Length**:

$$L(t) = \int_a^t \sqrt{g_{ij}(x^i, x^j)dx^i dx^j}du$$

by minimizing the arc length function, we obtain the geodesic on the Riemannian manifold.

(iii) **Area (Measure)**:

$$A = \int\int_U \sqrt{EG - F^2}dx_1 dx_2 \quad \text{and} \quad \mu(d\boldsymbol{x}) = \sqrt{\det g}\; d\boldsymbol{x}$$

(iv) **Laplace-Beltrami Operator**:

$$\Delta f = \lim_{\delta \to 0} \frac{\int_{U_\delta(\boldsymbol{x})} f(u)\mu(du) - f(\boldsymbol{x})}{\delta^2} = \frac{1}{\sqrt{\det g}} \frac{\partial}{\partial x^i}\left(\sqrt{\det g}(g^{ij})\frac{\partial}{\partial x^j}\right)$$

Now let us look at some examples.

3.2 Riemannian metric and Laplacian on a sphere

Consider $\mathcal{M} \subset \mathbb{R}^3$ which is a sphere $\mathbb{S}^2 = \{(x, y, z) : x^2 + y^2 + z^2 = 1\}$. The parametric equations are

$$\begin{cases} x(\theta, \phi) = \sin\theta\cos\phi \\ y(\theta, \phi) = \sin\theta\sin\phi \\ z(\theta, \phi) = \cos\theta \end{cases}$$

where $\theta \in [0, \pi]$, $\phi \in [0, 2\pi)$.

Let $\boldsymbol{\gamma}(t) = \boldsymbol{\varphi}(\theta(t), \phi(t)) = (\sin\theta(t)\cos\phi(t), \sin\theta(t)\sin\phi(t), \cos\theta(t))$. Then

$$\boldsymbol{\varphi}_\theta = (\cos\theta\cos\phi, \cos\theta\sin\phi, -\sin\theta)$$

$$\boldsymbol{\varphi}_\phi = (-\sin\theta\sin\phi, \sin\theta\cos\phi, 0)$$

and

$$E(\theta, \phi) = \boldsymbol{\varphi}_\theta \cdot \boldsymbol{\varphi}_\theta = 1$$

$$F(\theta, \phi) = \boldsymbol{\varphi}_\theta \cdot \boldsymbol{\varphi}_\phi = 0$$

$$G(\theta, \phi) = \boldsymbol{\varphi}_\phi \cdot \boldsymbol{\varphi}_\phi = \sin^2\theta$$

then

$$g = \begin{bmatrix} 1 & 0 \\ 0 & \sin^2\theta \end{bmatrix} \quad \text{and} \quad g^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{\sin^2\theta} \end{bmatrix}$$

which implies that the first quadratic form is

$$ds^2 = d\theta^2 + \sin^2\theta d\phi^2$$

and the arc length is

$$L(t) = \int_0^t \sqrt{\left(\frac{d\theta}{du}\right)^2 + \sin^2\theta \left(\frac{d\phi}{du}\right)^2} du$$

especially on the meredian, i.e. $d\phi^2 = 0$, the arc length becomes

$$L(t) = \int_0^{\theta_0} d\theta = \theta_0$$

and the area measure is

$$\mu(d(\theta, \phi)) = \sin\theta d\theta d\phi$$

$$A = \int_0^{\phi_0} \int_0^{\theta_0} \sin\theta d\theta d\phi = \phi_0(1 - \cos\theta_0)$$

and the laplacian on a sphere is

$$\Delta = \frac{1}{\sqrt{\det g}} \frac{\partial}{\partial x^1}\left(\sqrt{\det g}(g^{11})\frac{\partial}{\partial x^1}\right) + \frac{1}{\sqrt{\det g}} \frac{\partial}{\partial x^1}\left(\sqrt{\det g}(g^{12})\frac{\partial}{\partial x^2}\right)$$

$$+\frac{1}{\sqrt{\det g}} \frac{\partial}{\partial x^2}\left(\sqrt{\det g}(g^{21})\frac{\partial}{\partial x^1}\right) + \frac{1}{\sqrt{\det g}} \frac{\partial}{\partial x^2}\left(\sqrt{\det g}(g^{22})\frac{\partial}{\partial x^2}\right)$$

$$= \frac{1}{\sin\theta}\left[\frac{\partial}{\partial\theta}\left(sin\theta\frac{\partial}{\partial\theta}\right) + \frac{\partial}{\partial\phi}\left(\frac{1}{\sin\theta}\frac{\partial}{\partial\phi}\right)\right]$$

$$= \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}.$$

3.3 Riemannian metric and Laplacian on a torus

Consider $\mathcal{M} \subset \mathbb{R}^3$ which is a torus $T^2 = S^1 \times S^1$. The parametric equations are

$$\begin{cases} x(\theta,\phi) = (R + r\cos\theta)\cos\phi \\ y(\theta,\phi) = (R + r\cos\theta)\sin\phi \\ z(\theta,\phi) = r\sin\theta \end{cases}$$

where $\theta, \phi \in [0, 2\pi)$ and R is the distance from the middle of the torus to the middle of the tube and r is the radius of circle of the tube.

Let $\boldsymbol{\gamma}(t) = \boldsymbol{\varphi}(\theta(t), \phi(t)) = ((R{+}r\cos\theta(t))\cos\phi(t), (R{+}r\cos\theta(t))\sin\phi(t), rsin\theta(t))$. Then

$$\boldsymbol{\varphi}_\theta = (-r\cos\phi\sin\theta, -r\sin\phi\sin\theta, r\cos\theta)$$

$$\boldsymbol{\varphi}_\phi = (-(R + r\cos\theta)\sin\phi, (R + r\cos\theta)\cos\phi, 0)$$

and

$$E(\theta,\phi) = \boldsymbol{\varphi}_\theta \cdot \boldsymbol{\varphi}_\theta = r^2$$

$$F(\theta,\phi)=\boldsymbol{\varphi}_\theta\cdot\boldsymbol{\varphi}_\phi=0$$

$$G(\theta,\phi)=\boldsymbol{\varphi}_\phi\cdot\boldsymbol{\varphi}_\phi=(R+r\cos\theta)^2$$

then

$$g=\begin{bmatrix} r^2 & 0 \\ 0 & (R+r\cos\theta)^2 \end{bmatrix} \quad \text{and} \quad g^{-1}=\begin{bmatrix} \frac{1}{r^2} & 0 \\ 0 & \frac{1}{(R+r\cos\theta)^2} \end{bmatrix}$$

which implies that the first quadratic form is

$$ds^2=r^2d\theta^2+(R+r\cos\theta)^2d\phi^2$$

and the arc length is

$$L(t)=\int_0^t\sqrt{r^2\left(\frac{d\theta}{du}\right)^2+(R+r\cos\theta)^2\left(\frac{d\phi}{du}\right)^2}du$$

especially on the meredian, i.e. $d\phi^2=0$, the arc length becomes

$$L(t)=\int_0^{\theta_0}rd\theta=r\theta_0$$

and the area measure is

$$\mu(d(\theta,\phi))=\sqrt{r^2(R+r\cos\theta)^2}d\theta d\phi$$

$$A=\int_0^{\phi_0}\int_0^{\theta_0}(rR+r^2\cos\theta)d\theta d\phi=rR\theta_0\phi_0+r^2\phi_0\sin\theta_0$$

and the laplacian on a torus is

$$\Delta = \frac{1}{\sqrt{\det g}}\frac{\partial}{\partial x^1}\left(\sqrt{\det g}(g^{11})\frac{\partial}{\partial x^1}\right) + \frac{1}{\sqrt{\det g}}\frac{\partial}{\partial x^1}\left(\sqrt{\det g}(g^{12})\frac{\partial}{\partial x^2}\right)$$

$$+\frac{1}{\sqrt{\det g}}\frac{\partial}{\partial x^2}\left(\sqrt{\det g}(g^{21})\frac{\partial}{\partial x^1}\right) + \frac{1}{\sqrt{\det g}}\frac{\partial}{\partial x^2}\left(\sqrt{\det g}(g^{22})\frac{\partial}{\partial x^2}\right)$$

$$= \frac{1}{r(R+\cos\theta)}\left[\frac{\partial}{\partial\theta}\left(\frac{1}{r^3(R+\cos\theta)}\frac{\partial}{\partial\theta}\right) + \frac{\partial}{\partial\phi}\left(\frac{1}{r(R+\cos\theta)^3}\frac{\partial}{\partial\phi}\right)\right]$$

$$= \frac{1}{r(R+\cos\theta)}\frac{\partial}{\partial\theta}\left(\frac{1}{r^3(R+\cos\theta)}\frac{\partial}{\partial\theta}\right) + \frac{1}{r(R+\cos\theta)^4}\frac{\partial^2}{\partial\phi^2}.$$

CHAPTER 4: PROBABILITY DISTRIBUTION AND CRITICAL POINTS ON COMPACT RIEMANNIAN MANIFOLDS

4.1 Single maximum point on the surface of a sphere

Now we'll formulate a different version of the Moser-type problem. Let $\mathcal{M}$ be a compact Riemannian manifold with the metric $ds^2 = g_{ij}(x)dx^i dx^j$ defined on the system of maps $X : \mathbb{R}^2 \to \mathcal{M}$ covering $\mathcal{M}$ and $d\sigma = \sqrt{\det g_{ij}(x)}dx$ be the differential of the Lebesgue measure on $\mathcal{M}$. One can select the metric tensor $g_{ij}(x)$ in such a way that $\int\limits_{\mathcal{M}} d\sigma = \int\limits_{\mathcal{M}} \sqrt{\det g(x)}dx = 1$.

Let $X_1, \cdots, X_n$ be the points on the compact Riemannian manifold $\mathcal{M}$ with uniform distribution measure $d\sigma$ and $\mathcal{G}(X) : \mathcal{M} \to \mathbb{R}$ be the function of C^2 class on $\mathcal{M}$ such that

$$Y_1 = \mathcal{G}(X_1), \cdots, Y_n = \mathcal{G}(X_n)$$

are the scalar i.i.d. random variables. Our goal is to find $S_n = \max\limits_{\tau \le n} E[\mathcal{G}(X_\tau)] = \max\limits_{\tau \le n} E[Y_\tau]$. To do this, one needs to find the distribution function of Y_i with $P\{Y_i \ge y\} = m(\{X_i \in \mathcal{M} : \mathcal{G}(X_i) \ge y\}), i = 1, \cdots, n$. For large n the asymptotics of S_n depends on the structure of top extrema of $\mathcal{G}(\cdot)$. Literatures relate to this can be found in [7], [8].

If $Y_1, \cdots, Y_n$ are i.i.d. random variables uniformly distributed on [0,1], the problem reduces back to the classical Moser problem. Otherwise, we need to consider the structure near the critical point of the reward function $\mathcal{G}$ on the compact Riemannian manifold.

Without loss of generality, suppose there exists only one global non-degenerated maximum point X_* on the entire manifold $\mathcal{M} = \mathbb{S}^2$ and $\mathcal{G}(X) \in C^2(\mathcal{M})$ such that

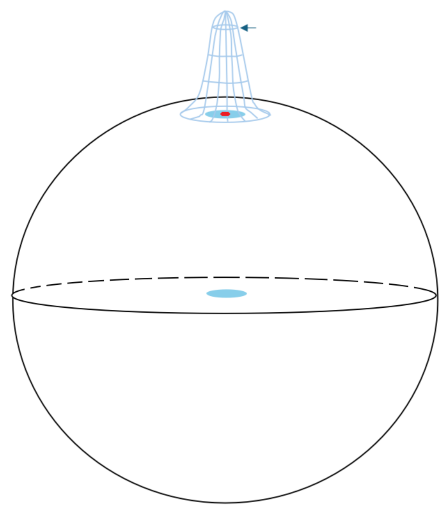

Figure 4.1: This figure illustrates a single maximum point (red point) of $\mathcal{G}$, i.e. X_*, on the surface of a sphere. The arrow indicates the threshold level and the blue regions is a set projection of the threshold level on the manifold and xy-plane.

$\mathcal{G}(X_*) = 1$, that is $\mathcal{G}(X) < 1$, $\forall\ X \neq X_*$. Then there exists an appropriate coordinate system near X_* such that

$$\mathcal{G}(X) = 1 + \frac{1}{2}\sum_{i=1}^{d} \lambda_i (X^i - X_*^i)^2 + o((X - X_*)^2),$$

$$d = \dim \mathcal{M}, \quad \lambda_i = \frac{\partial^2 \mathcal{G}}{\partial X^{i^2}}(X_*) < 0.$$

See details in [9].

Since the result of the maximum expectation, S_n, is an asymptotic result, when $n \to \infty$, meaning when we wait long enough eventually there is high chance that one of the random variables will hit the maximum or get very close to the maximum. So let's assume for a second that $\epsilon > 0$ and the threshold level to be $h_n = 1 - \epsilon$, $\forall n$. When h_n is extremely close to the maximum value 1, the area of the projected set of the threshold level of $\mathcal{G}$ onto the manifold can be approximated by the one onto the xy-plane. We can regard the tail probability, $P\{Y > y\}$, as the measure of the projection of the function $\mathcal{G}$ onto the xy-plane instead of the manifold. See figure 4.1. Then one can calculate the tail probability distribution of Y as follow.

$$\begin{aligned}
P\{Y > y\} &= m\{X : \mathcal{G}(X) > y\} \\
&\approx m\left\{X : \sum_{i=1}^{d} |\lambda_i|(X^i - X_*^i)^2 < 2(1-y)\right\} \\
&= m\left\{X : \sum_{i=1}^{d} \frac{(X^i - X_*^i)^2}{\frac{2(1-y)}{|\lambda_i|}} < 1\right\} \\
&= \frac{2\pi^{\frac{d}{2}}}{d\Gamma(\frac{d}{2})} \prod_{i=1}^{d} \frac{\sqrt{2}(1-y)^{\frac{1}{2}}}{\sqrt{|\lambda_i|}} \\
&= \frac{2^{(\frac{d}{2}+1)}\pi^{\frac{d}{2}}}{d\Gamma(\frac{d}{2})\sqrt{|\lambda_1|\cdots|\lambda_d|}}(1-y)^{\frac{d}{2}}.
\end{aligned}$$

We used here the formula for the volume of d-dimensional ellipsoid.

Then the corresponding probability density is

$$f(y) = \frac{d}{dy}P\{Y \leq y\} = \frac{(2\pi)^{\frac{d}{2}}}{\Gamma(\frac{d}{2})\sqrt{|\lambda_1|\cdots|\lambda_d|}}(1-y)^{\frac{d}{2}-1}.$$

Now let $X_1, \cdots, X_n$ be a sequence of independent random points on $\mathcal{M}$. Let the height of the maximum point to be $\mathcal{G}(X_*) = 1$ and fix a threshold. If $Y_i \geq h_i$, one would stop, otherwise he would continue the game if $Y_i < h_i$. Then the maximum expectation in each step is

$$
\begin{aligned}
S_n &= \max_{\tau \le n} \left(E[Y_n | Y_n \ge h_n] P\{Y_n \ge h_n\} + E[Y_n | Y_n < h_n] P\{Y_n < h_n\} \right) \\
&= \max_{\tau \le n} \left(\int_{h_n}^{1} y f(y) dy + S_{n-1} \left(1 - \frac{2^{(\frac{d}{2}+1)} \pi^{\frac{d}{2}}}{d\Gamma(\frac{d}{2}) \sqrt{|\lambda_1| \cdots |\lambda_d|}} (1-h_n)^{\frac{d}{2}} \right) \right) \\
&= \max_{\tau \le n} \left(\frac{(2\pi)^{\frac{d}{2}}}{\Gamma(\frac{d}{2}) \sqrt{|\lambda_1| \cdots |\lambda_d|}} \int_{h_n}^{1} y (1-y)^{\frac{d}{2}-1} dy \right. \\
&\qquad\qquad \left. + S_{n-1} \left(1 - \frac{2^{(\frac{d}{2}+1)} \pi^{\frac{d}{2}}}{d\Gamma(\frac{d}{2}) \sqrt{|\lambda_1| \cdots |\lambda_d|}} (1-h_n)^{\frac{d}{2}} \right) \right) \\
&= \max_{\tau \le n} \left(\frac{(2\pi)^{\frac{d}{2}}}{\Gamma(\frac{d}{2}) \sqrt{|\lambda_1| \cdots |\lambda_d|}} \int_{h_n}^{1} [1-(1-y)] (1-y)^{\frac{d}{2}-1} dy \right. \\
&\qquad\qquad \left. + S_{n-1} \left(1 - \frac{2^{(\frac{d}{2}+1)} \pi^{\frac{d}{2}}}{d\Gamma(\frac{d}{2}) \sqrt{|\lambda_1| \cdots |\lambda_d|}} (1-h_n)^{\frac{d}{2}} \right) \right) \\
&= \max_{\tau \le n} \left(\frac{(2\pi)^{\frac{d}{2}}}{\Gamma(\frac{d}{2}) \sqrt{|\lambda_1| \cdots |\lambda_d|}} \left(\int_{h_n}^{1} (1-y)^{\frac{d}{2}-1} dy - \int_{h_n}^{1} (1-h_n)^{\frac{d}{2}} dy \right) \right. \\
&\qquad\qquad \left. + S_{n-1} \left(1 - \frac{2^{(\frac{d}{2}+1)} \pi^{\frac{d}{2}}}{d\Gamma(\frac{d}{2}) \sqrt{|\lambda_1| \cdots |\lambda_d|}} (1-h_n)^{\frac{d}{2}} \right) \right) \\
&= \max_{\tau \le n} \left(\frac{(2\pi)^{\frac{d}{2}}}{\Gamma(\frac{d}{2}) \sqrt{|\lambda_1| \cdots |\lambda_d|}} \left[-\frac{(1-y)^{d/2}}{\frac{d}{2}} + \frac{(1-y)^{\frac{d}{2}+1}}{\frac{d}{2}+1} \right]_{h_n}^{1} \right. \\
&\qquad\qquad \left. + S_{n-1} \left(1 - \frac{2^{(\frac{d}{2}+1)} \pi^{\frac{d}{2}}}{d\Gamma(\frac{d}{2}) \sqrt{|\lambda_1| \cdots |\lambda_d|}} (1-h_n)^{\frac{d}{2}} \right) \right) \\
&= \max_{\tau \le n} \left(\frac{(2\pi)^{\frac{d}{2}}}{\Gamma(\frac{d}{2}) \sqrt{|\lambda_1| \cdots |\lambda_d|}} \left[\frac{(1-h_n)^{d/2}}{\frac{d}{2}} - \frac{(1-h_n)^{\frac{d}{2}+1}}{\frac{d}{2}+1} \right] \right. \\
&\qquad\qquad \left. + S_{n-1} \left(1 - \frac{2^{(\frac{d}{2}+1)} \pi^{\frac{d}{2}}}{d\Gamma(\frac{d}{2}) \sqrt{|\lambda_1| \cdots |\lambda_d|}} (1-h_n)^{\frac{d}{2}} \right) \right).
\end{aligned}
$$

Take the derivative with respect to h_n and set zero implies

$$
\frac{(2\pi)^{\frac{d}{2}}}{\Gamma(\frac{d}{2}) \sqrt{|\lambda_1| \cdots |\lambda_d|}} (1-h_n)^{\frac{d}{2}-1} \left[-1 + (1-h_n) - S_{n-1} \right] = 0
$$

implies

$$
h_n = S_{n-1}.
$$

This gives

$$S_n = \frac{(2\pi)^{\frac{d}{2}}}{\Gamma(\frac{d}{2})\sqrt{|\lambda_1|\cdots|\lambda_d|}}\left[\frac{(1-S_{n-1})^{d/2}}{\frac{d}{2}} - \frac{(1-S_{n-1})^{\frac{d}{2}+1}}{\frac{d}{2}+1}\right] + S_{n-1}\left(1 - \frac{2^{(\frac{d}{2}+1)}\pi^{\frac{d}{2}}}{d\Gamma(\frac{d}{2})\sqrt{|\lambda_1|\cdots|\lambda_d|}}(1-S_{n-1})^{\frac{d}{2}}\right)$$

Now rewrite the recursive relation as a function

$$\begin{aligned} g(x) &= \frac{(2\pi)^{\frac{d}{2}}}{\Gamma(\frac{d}{2})\sqrt{|\lambda_1|\cdots|\lambda_d|}}\left[\frac{(1-x)^{d/2}}{\frac{d}{2}} - \frac{(1-x)^{\frac{d}{2}+1}}{\frac{d}{2}+1}\right] + x\left(1 - \frac{2^{(\frac{d}{2}+1)}\pi^{\frac{d}{2}}}{d\Gamma(\frac{d}{2})\sqrt{|\lambda_1|\cdots|\lambda_d|}}(1-x)^{\frac{d}{2}}\right) \\ &= \frac{(2\pi)^{\frac{d}{2}}}{\Gamma(\frac{d}{2})\sqrt{|\lambda_1|\cdots|\lambda_d|}}(1-x)^{d/2}\left[\frac{2}{d} - \frac{(1-x)}{\frac{d}{2}+1} - \frac{2}{d}x\right] + x \end{aligned}$$

by fixed point theorem, set $g(x) = x$ gives

$$(1-x)^{d/2}\left[\frac{2}{d} - \frac{(1-x)}{\frac{d}{2}+1} - \frac{2}{d}x\right] = 0$$

by solving this equation, the solution is $x = 1$. Since $g'(1) \leq 1$, g is contractive and $g(S_n) \to 1$ as $n \to \infty$. Let

$$h_n = 1 - g(S_n)$$

implies

$$1 - h_n = \frac{(2\pi)^{\frac{d}{2}}}{\Gamma(\frac{d}{2})\sqrt{|\lambda_1|\cdots|\lambda_d|}}(1-(1-h_{n-1}))^{d/2}\left[\frac{2}{d} - \frac{(1-(1-h_{n-1}))}{\frac{d}{2}+1} - \frac{2}{d}(1-h_{n-1})\right] + (1-h_{n-1})$$

implies

$$h_n = -\frac{(2\pi)^{\frac{d}{2}}}{\Gamma(\frac{d}{2})\sqrt{|\lambda_1|\cdots|\lambda_d|}} h_{n-1}^{d/2}\left[\frac{2}{d} - \frac{h_{n-1}}{\frac{d}{2}+1} - \frac{2}{d}(1-h_{n-1})\right] + h_{n-1}$$
$$= -\frac{(2\pi)^{\frac{d}{2}}}{\Gamma(\frac{d}{2})\sqrt{|\lambda_1|\cdots|\lambda_d|}}\left[\frac{2}{d(\frac{d}{2}+1)}\right] h_{n-1}^{\frac{d}{2}+1} + h_{n-1}$$

Now let $k = \frac{d}{2}+1$ and $a = \frac{2^{\frac{d}{2}+1}\pi^{\frac{d}{2}}}{d(\frac{d}{2}+1)\Gamma(\frac{d}{2})\sqrt{|\lambda_1|\cdots|\lambda_d|}}$. Then by the Pólya and Szëgo theorem,

$$n^{\frac{2}{d}} h_n \to \left[\frac{2^{\frac{d}{2}+1}\pi^{\frac{d}{2}}}{d(\frac{d}{2}+1)\Gamma(\frac{d}{2})\sqrt{|\lambda_1|\cdots|\lambda_d|}}\frac{d}{2}\right]^{-\frac{2}{d}}$$

That is,

$$h_n \to \left[\frac{(2\pi)^{\frac{d}{2}} n}{(\frac{d}{2}+1)\Gamma(\frac{d}{2})\sqrt{|\lambda_1|\cdots|\lambda_d|}}\right]^{-\frac{2}{d}}$$

Therefore,

$$S_n \to 1 - \left[\frac{(2\pi)^{\frac{d}{2}} n}{(\frac{d}{2}+1)\Gamma(\frac{d}{2})\sqrt{|\lambda_1|\cdots|\lambda_d|}}\right]^{-\frac{2}{d}}$$

as $n \to \infty$.

In general, the asymptotic depends on the classification of the critical points on the Riemannian manifolds or on $\mathbb{R}^d$ due to locality problems. This example only shows the case of smooth non-degenerated critical point.

One can also expand the function $\mathcal{G}(X)$ to a more general form. Let $\mathcal{G}(X) = 1 - \sum_{i=1}^{d} |\lambda_i| X_i^{2p_i}$ be a smooth function with a non-degenerate critical point where $p_i \in \mathbb{N}$. Let's apply change of variable $X_i = \frac{(1-y)^{\frac{1}{2p_i}}}{\lambda_i^{\frac{1}{2p_i}}} t_i$ and $dX_i = \frac{(1-y)^{\frac{1}{2p_i}}}{\lambda_i^{\frac{1}{2p_i}}} dt_i$, then the probability distribution of Y is

$$
\begin{aligned}
P\{Y > y\} &= m\left\{X : \mathcal{G}(X) > y\right\} \\
&= m\left\{X : 1 - \sum_{i=1}^{d} |\lambda_i| X_i^{2p_i} > y\right\} \\
&= m\left\{X : \sum_{i=1}^{d} |\lambda_i| X_i^{2p_i} < 1 - y\right\} \\
&= m\left\{X : \sum_{i=1}^{d} \left(\frac{|\lambda_i|^{\frac{1}{2p_i}} X_i}{(1-y)^{\frac{1}{2p_i}}}\right)^{2p_i} < 1\right\} \\
&= 2^d m\left\{X : \sum_{i=1}^{d} \left(\frac{|\lambda_i|^{\frac{1}{2p_i}} |X_i|}{(1-y)^{\frac{1}{2p_i}}}\right)^{2p_i} < 1\right\} \\
&= 2^d \int \cdots \int_{\mathbb{R}^d} \mathbf{1}\left\{X : \sum_{i=1}^{d} \left(\frac{|\lambda_i|^{\frac{1}{2p_i}} |X_i|}{(1-y)^{\frac{1}{2p_i}}}\right)^{2p_i} < 1\right\} dx_1 \cdots dx_d \\
&= 2^d \int \cdots \int_{\mathbb{R}^d} \mathbf{1}\left\{t : \sum_{i=1}^{d} |t_i|^{2p_i} < 1\right\} \frac{(1-y)^{\frac{1}{2p_1}}}{|\lambda_1|^{\frac{1}{2p_1}}} \cdots \frac{(1-y)^{\frac{1}{2p_d}}}{|\lambda_d|^{\frac{1}{2p_d}}} dt_1 \cdots dt_d \\
&= \frac{2^d}{\Gamma(d+1)} \prod_{i=1}^{d} \left(\frac{1-y}{|\lambda_i|}\right)^{\frac{1}{2p_i}}.
\end{aligned}
$$

Due to the symmetry of the shape, the constant 2^d appeared in the middle of calculation helps simplify the calculation by focusing on the first octant. Also it worths to note that the above integral is the so-called Dirichlet integral. See [10].

Then the probability density is

$$
\begin{aligned}
f(y) &= \frac{d}{dy} P\{Y < y\} \\
&= \frac{d}{dy}\left(1 - \frac{2^d}{\Gamma(d+1)} \prod_{i=1}^{d} \left(\frac{1-y}{|\lambda_i|}\right)^{\frac{1}{2p_i}}\right) \\
&= -\frac{2^d}{\Gamma(d+1)} \frac{d}{dy} \exp\left(\sum_{i=1}^{d} \frac{1}{2p_i} \log\left(\frac{1-y}{|\lambda_i|}\right)\right)
\end{aligned}
$$

$$= -\frac{2^d}{\Gamma(d+1)}\left(\sum_{i=1}^{d}\frac{1}{2p_i}\frac{d}{dy}\log\left(\frac{1-y}{|\lambda_i|}\right)\right)\exp\left(\sum_{i=1}^{d}\frac{1}{2p_i}\log\left(\frac{1-y}{|\lambda_i|}\right)\right)$$
$$= -\frac{2^d}{\Gamma(d+1)}\frac{1}{y-1}\left(\sum_{i=1}^{d}\frac{1}{2p_i}\right)\exp\left(\sum_{i=1}^{d}\frac{1}{2p_i}\log\left(\frac{1-y}{|\lambda_i|}\right)\right)$$
$$= \frac{2^d}{(1-y)\Gamma(d+1)}\left(\sum_{i=1}^{d}\frac{1}{2p_i}\right)\prod_{i=1}^{d}\left(\frac{1-y}{|\lambda_i|}\right)^{\frac{1}{2p_i}}.$$

Now let $X_1, \cdots, X_n$ be a sequence of independent random points on $\mathcal{M}$. Let the height of the maximum point to be $\mathcal{G}(X_*) = 1$. If $Y_i \geq h_i$, one would stop, otherwise he would continue the game if $Y_i < h_i$. Then the maximum expectation in each step is

$$\begin{aligned}
S_n &= \max_{\tau\leq n}\Big(E[Y_n|Y_n \geq h_n]P\{Y_n \geq h_n\} + E[Y_n|Y_n < h_n]P\{Y_n < h_n\}\Big)\\
&= \max_{\tau\leq n}\left(\int_{h_n}^{1} yf(y)dy + S_{n-1}\left(1-\frac{2^d}{\Gamma(d+1)}\prod_{i=1}^{d}\left(\frac{1-h_n}{|\lambda_i|}\right)^{\frac{1}{2p_i}}\right)\right)\\
&= \max_{\tau\leq n}\left(\int_{h_n}^{1}\frac{2^d(\sum\limits_{i=1}^{d}\frac{1}{2p_i})y}{(1-y)\Gamma(d+1)}\prod_{i=1}^{d}\left(\frac{1-y}{|\lambda_i|}\right)^{\frac{1}{2p_i}}dy\right.\\
&\qquad\qquad\left. + S_{n-1}\left(1-\frac{2^d}{\Gamma(d+1)}\prod_{i=1}^{d}\left(\frac{1-h_n}{|\lambda_i|}\right)^{\frac{1}{2p_i}}\right)\right)\\
&= \max_{\tau\leq n}\left(\left(\sum_{i=1}^{d}\frac{1}{2p_i}\right)\frac{2^d}{\Gamma(d+1)}\prod_{i=1}^{d}\left(\frac{1}{|\lambda_i|}\right)^{\frac{1}{2p_i}}\right.\\
&\qquad\times\left\{\left[\frac{(1-y)^{1+\sum_{i=1}^{d}\frac{1}{2p_i}}}{1+\sum_{i=1}^{d}\frac{1}{2p_i}}\right]_{h_n}^{1} - \left[\frac{(1-y)^{\sum_{i=1}^{d}\frac{1}{2p_i}}}{\sum_{i=1}^{d}\frac{1}{2p_i}}\right]_{h_n}^{1}\right\}\\
&\qquad\qquad\left. + S_{n-1}\left(1-\frac{2^d}{\Gamma(d+1)}\prod_{i=1}^{d}\left(\frac{1-h_n}{|\lambda_i|}\right)^{\frac{1}{2p_i}}\right)\right)\\
&= \max_{\tau\leq n}\left(\left(\sum_{i=1}^{d}\frac{1}{2p_i}\right)\frac{2^d}{\Gamma(d+1)}\prod_{i=1}^{d}\left(\frac{1}{|\lambda_i|}\right)^{\frac{1}{2p_i}}\left\{\frac{(1-h_n)^{\sum_{i=1}^{d}\frac{1}{2p_i}}}{\sum_{i=1}^{d}\frac{1}{2p_i}} - \frac{(1-h_n)^{1+\sum_{i=1}^{d}\frac{1}{2p_i}}}{1+\sum_{i=1}^{d}\frac{1}{2p_i}}\right\}\right.\\
&\qquad\qquad\left. + S_{n-1}\left(1-\frac{2^d}{\Gamma(d+1)}\prod_{i=1}^{d}\left(\frac{1-h_n}{|\lambda_i|}\right)^{\frac{1}{2p_i}}\right)\right).
\end{aligned} \tag{4.1}$$

Now take the derivative with respect to h_n and set zero, that is

$$\frac{2^d}{\Gamma(d+1)}\left(\sum_{i=1}^{d}\frac{1}{2p_i}\right)\prod_{i=1}^{d}\left(\frac{1}{|\lambda_i|}\right)^{\frac{1}{2p_i}}\left\{-(1-h_n)^{\sum_{i=1}^{d}\frac{1}{2p_i}-1}+(1-h_n)^{\sum_{i=1}^{d}\frac{1}{2p_i}}\right\}$$
$$+S_{n-1}\frac{2^d}{\Gamma(d+1)}\left(\sum_{i=1}^{d}\frac{1}{2p_i(1-h_n)}\right)\exp\left(\sum_{i=1}^{d}\frac{1}{2p_i}\log\left(\frac{1-h_n}{|\lambda_i|}\right)\right)=0$$

$$-h_n(1-h_n)^{\sum_{i=1}^{d}\frac{1}{2p_i}-1}\left(\sum_{i=1}^{d}\frac{1}{2p_i}\right)\prod_{i=1}^{d}\left(\frac{1}{|\lambda_i|}\right)^{\frac{1}{2p_i}}$$
$$+\frac{S_{n-1}}{1-h_n}\left(\sum_{i=1}^{d}\frac{1}{2p_i}\right)\prod_{i=1}^{d}\left(\frac{1-h_n}{|\lambda_i|}\right)^{\frac{1}{2p_i}}=0$$

Solve the above equation, we have

$$h_n = S_{n-1}$$

Substitute this back to (4.1),

$$S_n=\left(\sum_{i=1}^{d}\frac{1}{2p_i}\right)\frac{2^d}{\Gamma(d+1)}\prod_{i=1}^{d}\left(\frac{1}{|\lambda_i|}\right)^{\frac{1}{2p_i}}\left\{\frac{(1-S_{n-1})^{\sum_{i=1}^{d}\frac{1}{2p_i}}}{\sum_{i=1}^{d}\frac{1}{2p_i}}-\frac{(1-S_{n-1})^{1+\sum_{i=1}^{d}\frac{1}{2p_i}}}{1+\sum_{i=1}^{d}\frac{1}{2p_i}}\right\}$$
$$+S_{n-1}\left(1-\frac{2^d}{\Gamma(d+1)}\prod_{i=1}^{d}\left(\frac{1-S_{n-1}}{|\lambda_i|}\right)^{\frac{1}{2p_i}}\right)$$

Now rewrite the recursive relation as a function

$$g(x)=\left(\sum_{i=1}^{d}\frac{1}{2p_i}\right)\frac{2^d}{\Gamma(d+1)}\prod_{i=1}^{d}\left(\frac{1}{|\lambda_i|}\right)^{\frac{1}{2p_i}}\left\{\frac{(1-x)^{\sum_{i=1}^{d}\frac{1}{2p_i}}}{\sum_{i=1}^{d}\frac{1}{2p_i}}-\frac{(1-x)^{1+\sum_{i=1}^{d}\frac{1}{2p_i}}}{1+\sum_{i=1}^{d}\frac{1}{2p_i}}\right\}$$
$$+x\left(1-\frac{2^d}{\Gamma(d+1)}\prod_{i=1}^{d}\left(\frac{1-x}{|\lambda_i|}\right)^{\frac{1}{2p_i}}\right)$$

by the fixed point theorem, $g(x) = x$ gives

$$x = \left(\sum_{i=1}^{d} \frac{1}{2p_i}\right) \frac{2^d}{\Gamma(d+1)} \prod_{i=1}^{d} \left(\frac{1}{|\lambda_i|}\right)^{\frac{1}{2p_i}} \left\{ \frac{(1-x)^{\sum_{i=1}^{d} \frac{1}{2p_i}}}{\sum_{i=1}^{d} \frac{1}{2p_i}} - \frac{(1-x)^{1+\sum_{i=1}^{d} \frac{1}{2p_i}}}{1+\sum_{i=1}^{d} \frac{1}{2p_i}} \right\} + x\left(1 - \frac{2^d}{\Gamma(d+1)} \prod_{i=1}^{d} \left(\frac{1-x}{|\lambda_i|}\right)^{\frac{1}{2p_i}}\right)$$

Solving this equation, we obtain

$$x = 1$$

Since $g'(1) \leq 1$, g is contractive and $g(S_n) \to 1$ as $n \to \infty$. Let

$$h_n = 1 - g(S_n)$$

implies

$$1 - h_n = \left(\sum_{i=1}^{d} \frac{1}{2p_i}\right) \frac{2^d}{\Gamma(d+1)} \prod_{i=1}^{d} \left(\frac{1}{|\lambda_i|}\right)^{\frac{1}{2p_i}} \times \left\{ \frac{(1-(1-h_{n-1}))^{\sum_{i=1}^{d} \frac{1}{2p_i}}}{\sum_{i=1}^{d} \frac{1}{2p_i}} - \frac{(1-(1-h_{n-1}))^{1+\sum_{i=1}^{d} \frac{1}{2p_i}}}{1+\sum_{i=1}^{d} \frac{1}{2p_i}} \right\} + (1-h_{n-1})\left(1 - \frac{2^d}{\Gamma(d+1)} \prod_{i=1}^{d} \left(\frac{1-(1-h_{n-1})}{|\lambda_i|}\right)^{\frac{1}{2p_i}}\right)$$

$$1 - h_n = \left(\sum_{i=1}^{d} \frac{1}{2p_i}\right) \frac{2^d}{\Gamma(d+1)} \prod_{i=1}^{d} \left(\frac{1}{|\lambda_i|}\right)^{\frac{1}{2p_i}} \left\{ \frac{h_{n-1}^{\sum_{i=1}^{d} \frac{1}{2p_i}}}{\sum_{i=1}^{d} \frac{1}{2p_i}} - \frac{h_{n-1}^{1+\sum_{i=1}^{d} \frac{1}{2p_i}}}{1+\sum_{i=1}^{d} \frac{1}{2p_i}} \right\} + (1-h_{n-1})\left(1 - \frac{2^d}{\Gamma(d+1)} \prod_{i=1}^{d} \left(\frac{h_{n-1}}{|\lambda_i|}\right)^{\frac{1}{2p_i}}\right)$$

$$h_{n-1} - h_n + \frac{2^d}{\Gamma(d+1)} \prod_{i=1}^{d} \left(\frac{h_{n-1}}{|\lambda_i|}\right)^{\frac{1}{2p_i}} - \frac{2^d}{\Gamma(d+1)} h_{n-1} \prod_{i=1}^{d} \left(\frac{h_{n-1}}{|\lambda_i|}\right)^{\frac{1}{2p_i}} =$$

$$\frac{2^d}{\Gamma(d+1)} \left(\sum_{i=1}^{d} \frac{1}{2p_i}\right) \prod_{i=1}^{d} \left(\frac{1}{|\lambda_i|}\right)^{\frac{1}{2p_i}} \left\{ \frac{h_{n-1}^{\sum_{i=1}^{d} \frac{1}{2p_i}}}{\sum_{i=1}^{d} \frac{1}{2p_i}} - \frac{h_{n-1}^{1+\sum_{i=1}^{d} \frac{1}{2p_i}}}{1+\sum_{i=1}^{d} \frac{1}{2p_i}} \right\}$$

$$h_n = h_{n-1} + \frac{2^d}{\Gamma(d+1)} \prod_{i=1}^{d} \left(\frac{h_{n-1}}{|\lambda_i|}\right)^{\frac{1}{2p_i}} - \frac{2^d}{\Gamma(d+1)} h_{n-1} \prod_{i=1}^{d} \left(\frac{h_{n-1}}{|\lambda_i|}\right)^{\frac{1}{2p_i}}$$
$$- \frac{2^d}{\Gamma(d+1)} \left(\sum_{i=1}^{d} \frac{1}{2p_i}\right) \prod_{i=1}^{d} \left(\frac{1}{|\lambda_i|}\right)^{\frac{1}{2p_i}} \left(h_{n-1}^{\sum_{i=1}^{d} \frac{1}{2p_i}}\right) \left\{ \frac{1}{\sum_{i=1}^{d} \frac{1}{2p_i}} - \frac{h_{n-1}}{1+\sum_{i=1}^{d} \frac{1}{2p_i}} \right\}$$

$$h_n = h_{n-1} + \frac{2^d}{\Gamma(d+1)} \prod_{i=1}^{d} \left(\frac{h_{n-1}}{|\lambda_i|}\right)^{\frac{1}{2p_i}} - \frac{2^d}{\Gamma(d+1)} h_{n-1} \prod_{i=1}^{d} \left(\frac{h_{n-1}}{|\lambda_i|}\right)^{\frac{1}{2p_i}}$$
$$- \frac{2^d}{\Gamma(d+1)} \prod_{i=1}^{d} (\frac{1}{|\lambda_i|})^{\frac{1}{2p_i}} \left(h_{n-1}^{\sum_{i=1}^{d} \frac{1}{2p_i}}\right) + \frac{2^d}{\Gamma(d+1)} \frac{\sum\limits_{i=1}^{d} \frac{1}{2p_i}}{1+\sum_{i=1}^{d} \frac{1}{2p_i}} \prod_{i=1}^{d} \left(\frac{1}{|\lambda_i|}\right)^{\frac{1}{2p_i}} \left(h_{n-1}^{1+\sum_{i=1}^{d} \frac{1}{2p_i}}\right)$$

$$h_n = h_{n-1} + \frac{2^d}{\Gamma(d+1)} \left(\frac{\sum\limits_{i=1}^{d} \frac{1}{2p_i}}{1+\sum_{i=1}^{d} \frac{1}{2p_i}} - 1 \right) \prod_{i=1}^{d} \left(\frac{1}{|\lambda_i|}\right)^{\frac{1}{2p_i}} \left(h_{n-1}^{1+\sum_{i=1}^{d} \frac{1}{2p_i}}\right)$$

$$h_n = h_{n-1} - \left(\frac{2^d}{\Gamma(d+1)(1+\sum_{i=1}^{d} \frac{1}{2p_i})} \prod_{i=1}^{d} \left(\frac{1}{|\lambda_i|}\right)^{\frac{1}{2p_i}} \right) h_{n-1}^{1+\sum_{i=1}^{d} \frac{1}{2p_i}}.$$

Now let $k = 1 + \sum\limits_{i=1}^{d} \frac{1}{2p_i}$ and $a = \frac{2^d}{\Gamma(d+1)(1+\sum_{i=1}^{d} \frac{1}{2p_i})} \prod_{i=1}^{d} \left(\frac{1}{|\lambda_i|}\right)^{\frac{1}{2p_i}}$. Then by the Pólya and Szëgo theorem,

$$n^{\frac{1}{\sum_{i=1}^{d}\frac{1}{2p_i}}} h_n \to \left[\frac{2^d}{\Gamma(d+1)}\frac{\sum_{i=1}^{d}\frac{1}{2p_i}}{1+\sum_{i=1}^{d}\frac{1}{2p_i}}\prod_{i=1}^{d}\left(\frac{1}{|\lambda_i|}\right)^{\frac{1}{2p_i}}\right]^{-\frac{1}{\sum_{i=1}^{d}\frac{1}{2p_i}}}$$

That is,

$$h_n \to \left[\frac{2^d n}{\Gamma(d+1)}\frac{\sum_{i=1}^{d}\frac{1}{2p_i}}{1+\sum_{i=1}^{d}\frac{1}{2p_i}}\prod_{i=1}^{d}\left(\frac{1}{|\lambda_i|}\right)^{\frac{1}{2p_i}}\right]^{-\frac{1}{\sum_{i=1}^{d}\frac{1}{2p_i}}}$$

Therefore,

$$S_n \to 1 - \left[\frac{2^d n}{\Gamma(d+1)}\frac{\sum_{i=1}^{d}\frac{1}{2p_i}}{1+\sum_{i=1}^{d}\frac{1}{2p_i}}\prod_{i=1}^{d}\left(\frac{1}{|\lambda_i|}\right)^{\frac{1}{2p_i}}\right]^{-\frac{1}{\sum_{i=1}^{d}\frac{1}{2p_i}}}$$

as $n \to \infty$. Note that if $p_i = 1, \forall i$, then the above formula reduces to the previous example.

4.2 Maximum along a parallel on the surface of a sphere

Suppose the maximum of a smooth function $\mathcal{G}$ is not just sitting at one point of a sphere but along a parallel, i.e. $\theta = \theta_*$, of a sphere. One can choose the Euler angles coordinate system, (θ, φ), such that the smooth function $\mathcal{G} : \mathcal{M} \to \mathbb{R}$ near the maximum along the parallel can be defined as

$$\mathcal{G}(\theta, \varphi) \sim 1 + \frac{K(\varphi)}{2}(\theta - \theta_*)^2$$

where $K(\varphi) < 0$ is the second derivative of $\mathcal{G}$ at the maximum points on the parallel, $\theta \in [0, \pi]$ is the angle along the meridian and $\varphi \in [0, 2\pi)$ is the angle along the parallel. $K(\varphi)$ is also a quantity describing the curvature of the maximum in the direction orthogonal to the parallel. Without loss of generality, let $\mathcal{G}(\theta_*, \varphi) = 1$ be the maximum value along the parallel $\theta = \theta_*$ for all $\varphi \in [0, 2\pi)$. See figure 4.2. Furthermore, let

$$Y_1 = \mathcal{G}(\theta_1, \varphi_1), \cdots, Y_n = \mathcal{G}(\theta_n, \varphi_n)$$

be i.i.d. random variables on $[0, 1]$.

Let's calculate the probability distribution of Y. Let the path $\gamma = \{(\theta, \varphi) : \theta = \theta_*\}$ be the parallel. Then

$$\begin{aligned}
P\{Y > y\} &= m\left\{(\theta, \varphi) : \mathcal{G}(\theta, \varphi) > y\right\} \\
&= m\left\{(\theta, \varphi) : 1 - \frac{|K(\varphi)|}{2}(\theta - \theta_*)^2 > y\right\} \\
&= m\left\{(\theta, \varphi) : \frac{|K(\varphi)|}{2(1-y)}(\theta - \theta_*)^2 < 1\right\} \\
&= 2\sqrt{2(1-y)} \int_0^{2\pi} |K(\varphi)|^{-1} d\varphi
\end{aligned}$$

and the probability density of Y is

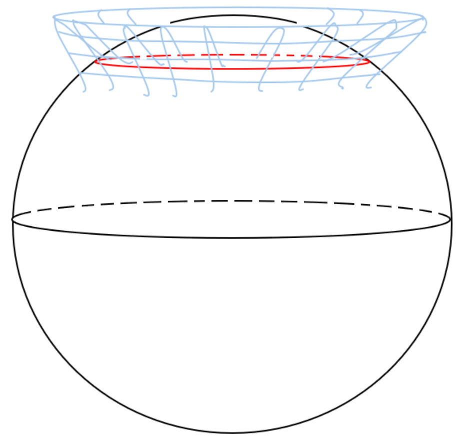

Figure 4.2: This figure illustrates the maxima along the parallel (red curve), $\theta = \theta_*$, form a volcano shape on the surface of a sphere. The projection of the threshold level of this volcano shape will form a band wrap around the red curve on the surface of the sphere.

$$
\begin{aligned}
f(y) &= \frac{d}{dy}\left(1 - 2\sqrt{2(1-y)}\int_0^{2\pi} |K(\varphi)|^{-1} d\varphi\right) \\
&= \sqrt{\frac{2}{1-y}}\int_0^{2\pi} |K(\varphi)|^{-1} d\varphi
\end{aligned}
$$

The maximum expectation can be calculated as

$$
\begin{aligned}
S_n &= \max_{\tau \le n}\Bigg(E[Y_n|Y_n \ge h_n]P\{Y_n \ge h_n\} + E[Y_n|Y_n < h_n]P\{Y_n < h_n\}\Bigg) \\
&= \max_{\tau \le n}\left(\int_{h_n}^1 yf(y)dy + S_{n-1}\left(1 - 2\sqrt{2(1-h_n)}\int_0^{2\pi} |K(\varphi)|^{-1} d\varphi\right)\right) \\
&= \max_{\tau \le n}\Bigg(\int_{h_n}^1 y\sqrt{\frac{2}{1-y}}\int_0^{2\pi} |K(\varphi)|^{-1} d\varphi dy \\
&\qquad\qquad + S_{n-1}\left(1 - 2\sqrt{2(1-h_n)}\int_0^{2\pi} |K(\varphi)|^{-1} d\varphi\right)\Bigg) \\
&= \max_{\tau \le n}\Bigg(\frac{\sqrt{2}\sqrt{1-h_n}(2h_n+4)}{3}\int_0^{2\pi} |K(\varphi)|^{-1} d\varphi \\
&\qquad\qquad + S_{n-1}\left(1 - 2\sqrt{2(1-h_n)}\int_0^{2\pi} |K(\varphi)|^{-1} d\varphi\right)\Bigg)
\end{aligned} \tag{4.2}
$$

Now take the derivative of above equation with respect to h_n and set zero, that is

$$\frac{\sqrt{2}}{3}\int_0^{2\pi}|K(\varphi)|^{-1}d\varphi\left(2\sqrt{1-h_n}-\frac{1}{2}(1-h_n)^{-\frac{1}{2}}(2h_n+4)\right)$$
$$-2\sqrt{2}S_{n-1}(-\frac{1}{2}(1-h_n)^{-\frac{1}{2}})\int_0^{2\pi}|K(\varphi)|^{-1}d\varphi=0$$

implies

$$\sqrt{2}(1-h_n)^{-\frac{1}{2}}\int_0^{2\pi}|K(\varphi)|^{-1}d\varphi\left[\frac{2}{3}(1-h_n)-\frac{1}{3}(h_n+2)+S_{n-1}\right]=0$$

implies

$$h_n=S_{n-1}$$

Substitute back to (4.2), then

$$S_n=\frac{\sqrt{2}\sqrt{1-S_{n-1}}(2S_{n-1}+4)}{3}\int_0^{2\pi}|K(\varphi)|^{-1}d\varphi$$
$$+S_{n-1}\left(1-2\sqrt{2(1-S_{n-1})}\int_0^{2\pi}|K(\varphi)|^{-1}d\varphi\right)$$

Now rewrite the recursive relation as a function

$$g(x)=\frac{\sqrt{2}\sqrt{1-x}(2x+4)}{3}\int_0^{2\pi}|K(\varphi)|^{-1}d\varphi+x\left(1-2\sqrt{2(1-x)}\int_0^{2\pi}|K(\varphi)|^{-1}d\varphi\right)$$

Set $g(x)=x$ and solve the equation, we have

$$x=\frac{\sqrt{2}\sqrt{1-x}(2x+4)}{3}\int_0^{2\pi}|K(\varphi)|^{-1}d\varphi+x-2x\sqrt{2(1-x)}\int_0^{2\pi}|K(\varphi)|^{-1}d\varphi$$

$$2\sqrt{2}\int_0^{2\pi}|K(\varphi)|^{-1}d\varphi\left(\frac{x+2}{3}-x\right)=0$$

implies

$$x=1$$

Since $g'(1) \leq 1$, g is contractive and $g(S_n) \to 1$ as $n \to \infty$. Let

$$h_n = 1 - g(S_n)$$

implies

$$\begin{aligned} 1 - h_n = & \frac{\sqrt{2}\sqrt{1-(1-h_{n-1})}(2(1-h_{n-1})+4)}{3} \int_0^{2\pi} |K(\varphi)|^{-1} d\varphi \\ & + (1-h_{n-1}) \left(1 - 2\sqrt{2(1-(1-h_{n-1}))} \int_0^{2\pi} |K(\varphi)|^{-1} d\varphi\right) \end{aligned}$$

implies

$$h_n = h_{n-1} - \left(\frac{4\sqrt{2}}{3} \int_0^{2\pi} |K(\varphi)|^{-1} d\varphi\right) h_{n-1}^{\frac{3}{2}}$$

Now let $k = \frac{3}{2}$ and $a = \frac{4\sqrt{2}}{3} \int_0^{2\pi} |K(\varphi)|^{-1} d\varphi$. Then by the Pólya and Szëgo theorem,

$$n^2 h_n \to \left(\frac{2\sqrt{2}}{3} \int_0^{2\pi} |K(\varphi)|^{-1} d\varphi\right)^{-2}$$

implies

$$h_n \to \frac{1}{\left[\left(\frac{2\sqrt{2}}{3} \int_0^{2\pi} |K(\varphi)|^{-1} d\varphi\right) n\right]^2}$$

Therefore,

$$S_n \to 1 - \frac{1}{\left[\left(\frac{2\sqrt{2}}{3} \int_0^{2\pi} |K(\varphi)|^{-1} d\varphi\right) n\right]^2}$$

as $n \to \infty$.

4.3 Maximum along a path on higher dimensional surface of a sphere.

Suppose the maximum of a smooth function $\mathcal{G}$ is sitting on a path γ on the surface of a $(d+1)$-dimensional sphere, let it be $\mathcal{M}$ with dimension d. Then the Euler angles coordinate system becomes $(\phi_1, \cdots, \phi_d)$ on $\mathcal{M}$, such that the smooth function $\mathcal{G} : \mathcal{M} \to \mathbb{R}$ near the maximum along the path can be defined as

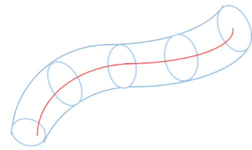

Figure 4.3: This illustration depicts the projection of the threshold level of the maximum of a function $\mathcal{G}$ along a curve γ (the red line) onto a higher-dimensional surface. The projected region is no longer a two dimensional band, but a higher dimensional snake shape.

$$\mathcal{G}(\phi^1, \cdots, \phi^d) \sim 1 + \sum_{i=2}^{d} \lambda_i(\phi^1)(\phi^i - \phi^i_*)^2$$

where $\lambda_i(\phi^1) < 0, i = 2, \cdots, d$ are the terms contain the second derivatives of $\mathcal{G}$ on the directions orthogonal to γ, ϕ^1 is the angle measures the deviation of the tangent vector at each point of γ from γ and $\phi^i \in [0, 2\pi), \forall i = 2, \cdots, d$ are the angles orthogonal to γ. Without loss of generality, let $\mathcal{G}(\phi^1, \phi^2_* \cdots, \phi^d_*) = 1$ to be the maximum value along γ for all ϕ^1. Let

$$Y_1 = \mathcal{G}(\phi^1_1, \cdots, \phi^d_1), \cdots, Y_n = \mathcal{G}(\phi^1_n, \cdots, \phi^d_n)$$

be i.i.d. random variables on the manifold $\mathcal{M}$. Let's calculate the probability distribution of Y.

$$\begin{aligned}
P\{Y > y\} &= m\left\{(\phi^1, \cdots, \phi^d) : \mathcal{G}(\phi^1, \cdots, \phi^d) > y\right\} \\
&= m\left\{(\phi^1, \cdots, \phi^d) : 1 - \sum_{i=2}^{d} |\lambda_i(\phi^1)|(\phi^i - \phi^i_*)^2 > y\right\} \\
&= m\left\{(\phi^1, \cdots, \phi^d) : \sum_{i=2}^{d} \frac{|\lambda_i(\phi^1)|}{1-y}(\phi^i - \phi^i_*)^2 < 1\right\} \\
&= m\left\{(\phi^1, \cdots, \phi^d) : \sum_{i=2}^{d} \frac{(\phi^i - \phi^i_*)^2}{\left(\sqrt{\frac{1-y}{|\lambda_i(\phi^1)|}}\right)^2} < 1\right\} \\
&= \frac{2\pi^{\frac{d-1}{2}}(1-y)^{\frac{d-1}{2}}}{(d-1)\Gamma(\frac{d-1}{2})} \int_0^{L(\gamma)} \prod_{i=2}^{d} |\lambda_i(\phi^1)|^{-1} d\phi^1
\end{aligned}$$

where $L(\gamma)$ is the length of γ. The probability density is

$$
\begin{aligned}
f(y) &= \frac{d}{dy}\left(1 - \frac{2\pi^{\frac{d-1}{2}}(1-y)^{\frac{d-1}{2}}}{(d-1)\Gamma(\frac{d-1}{2})}\int_0^{L(\gamma)}\prod_{i=2}^{d}|\lambda_i(\phi^1)|^{-1}d\phi^1\right)\\
&= \frac{\pi^{\frac{d-1}{2}}(1-y)^{\frac{d-3}{2}}}{\Gamma(\frac{d-1}{2})}\int_0^{L(\gamma)}\prod_{i=2}^{d}|\lambda_i(\phi^1)|^{-1}d\phi^1
\end{aligned}
$$

The maximum expectation can be calculated as

$$
\begin{aligned}
S_n &= \max_{\tau\le n}\Big(E[Y_n|Y_n\ge h_n]P\{Y_n\ge h_n\} + E[Y_n|Y_n<h_n]P\{Y_n<h_n\}\Big)\\
&= \max_{\tau\le n}\left(\int_{h_n}^{1} yf(y)dy + S_{n-1}\left(1 - \frac{2\pi^{\frac{d-1}{2}}(1-h_n)^{\frac{d-1}{2}}}{(d-1)\Gamma(\frac{d-1}{2})}\int_0^{L(\gamma)}\prod_{i=2}^{d}|\lambda_i(\phi^1)|^{-1}d\phi^1\right)\right)\\
&= \max_{\tau\le n}\Bigg(\int_{h_n}^{1}\frac{\pi^{\frac{d-1}{2}}y(1-y)^{\frac{d-3}{2}}}{\Gamma(\frac{d-1}{2})}\int_0^{L(\gamma)}\prod_{i=2}^{d}|\lambda_i(\phi^1)|^{-1}d\phi^1 dy\\
&\qquad + S_{n-1}\left(1 - \frac{2\pi^{\frac{d-1}{2}}(1-h_n)^{\frac{d-1}{2}}}{(d-1)\Gamma(\frac{d-1}{2})}\int_0^{L(\gamma)}\prod_{i=2}^{d}|\lambda_i(\phi^1)|^{-1}d\phi^1\right)\Bigg)\\
&= \max_{\tau\le n}\Bigg(\frac{\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d-1}{2})}\left(\frac{2}{d-1}(1-h_n)^{\frac{d-1}{2}} - \frac{2}{d+1}(1-h_n)^{\frac{d+1}{2}}\right)\int_0^{L(\gamma)}\prod_{i=2}^{d}|\lambda_i(\phi^1)|^{-1}d\phi^1\\
&\qquad + S_{n-1}\left(1 - \frac{2\pi^{\frac{d-1}{2}}(1-h_n)^{\frac{d-1}{2}}}{(d-1)\Gamma(\frac{d-1}{2})}\int_0^{L(\gamma)}\prod_{i=2}^{d}|\lambda_i(\phi^1)|^{-1}d\phi^1\right)\Bigg)
\end{aligned}
\tag{4.3}
$$

Now take the derivative of above equation with respect to h_n and set zero, that is

$$
\begin{aligned}
&\frac{\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d-1}{2})}\left((1-h_n)^{\frac{d-1}{2}} - (1-h_n)^{\frac{d-3}{2}}\right)\int_0^{L(\gamma)}\prod_{i=2}^{d}|\lambda_i(\phi^1)|^{-1}d\phi^1\\
&\qquad + S_{n-1}\left(\frac{\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d-1}{2})}(1-h_n)^{\frac{d-3}{2}}\right)\int_0^{L(\gamma)}\prod_{i=2}^{d}|\lambda_i(\phi^1)|^{-1}d\phi^1
\end{aligned}
$$

implies

$$
(1-h_n)^{\frac{d-1}{2}} - (1-h_n)^{\frac{d-3}{2}} + S_{n-1}(1-h_n)^{\frac{d-3}{2}} = 0
$$

implies

$$h_n = S_{n-1}$$

Substitute back to (4.3),

$$S_n = \frac{\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d-1}{2})}\left(\frac{2}{d-1}(1-S_{n-1})^{\frac{d-1}{2}} - \frac{2}{d+1}(1-S_{n-1})^{\frac{d+1}{2}}\right)\int_0^{L(\gamma)}\prod_{i=2}^{d}|\lambda_i(\phi^1)|^{-1}d\phi^1$$
$$+ S_{n-1}\left(1 - \frac{2\pi^{\frac{d-1}{2}}(1-S_{n-1})^{\frac{d-1}{2}}}{(d-1)\Gamma(\frac{d-1}{2})}\int_0^{L(\gamma)}\prod_{i=2}^{d}|\lambda_i(\phi^1)|^{-1}d\phi^1\right)$$

Now rewrite the recursive relation as a function

$$g(x) = \frac{\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d-1}{2})}\left(\frac{2}{d-1}(1-x)^{\frac{d-1}{2}} - \frac{2}{d+1}(1-x)^{\frac{d+1}{2}}\right)\int_0^{L(\gamma)}\prod_{i=2}^{d}|\lambda_i(\phi^1)|^{-1}d\phi^1$$
$$+ x\left(1 - \frac{2\pi^{\frac{d-1}{2}}(1-x)^{\frac{d-1}{2}}}{(d-1)\Gamma(\frac{d-1}{2})}\int_0^{L(\gamma)}\prod_{i=2}^{d}|\lambda_i(\phi^1)|^{-1}d\phi^1\right)$$

Set $g(x) = x$ and solve the equation, we have

$$x = \frac{\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d-1}{2})}\left(\frac{2}{d-1}(1-x)^{\frac{d-1}{2}} - \frac{2}{d+1}(1-x)^{\frac{d+1}{2}}\right)\int_0^{L(\gamma)}\prod_{i=2}^{d}|\lambda_i(\phi^1)|^{-1}d\phi^1$$
$$+ x\left(1 - \frac{2\pi^{\frac{d-1}{2}}(1-x)^{\frac{d-1}{2}}}{(d-1)\Gamma(\frac{d-1}{2})}\int_0^{L(\gamma)}\prod_{i=2}^{d}|\lambda_i(\phi^1)|^{-1}d\phi^1\right)$$

implies

$$\frac{\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d-1}{2})}\left((1-x)^{\frac{d-1}{2}}\left(\frac{4}{(d+1)(d-1)}\right)\right)\int_0^{L(\gamma)}\prod_{i=2}^{d}|\lambda_i(\phi^1)|^{-1}d\phi^1 = 0$$

implies

$$x = 1$$

Since $g'(1) \leq 1$, g is contractive and $g(S_n) \to 1$ as $n \to \infty$. Let

$$h_n = 1 - g(S_n)$$

implies

$$\begin{aligned} 1 - h_n = \frac{\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d-1}{2})} & \left(\frac{2}{d-1}(1-(1-h_{n-1}))^{\frac{d-1}{2}} - \frac{2}{d+1}(1-(1-h_{n-1}))^{\frac{d+1}{2}} \right) \\ & \times \int_0^{L(\gamma)} \prod_{i=2}^{d} |\lambda_i(\phi^1)|^{-1} d\phi^1 \\ & + (1-h_{n-1}) \left(1 - \frac{2\pi^{\frac{d-1}{2}}(1-(1-h_{n-1}))^{\frac{d-1}{2}}}{(d-1)\Gamma(\frac{d-1}{2})} \int_0^{L(\gamma)} \prod_{i=2}^{d} |\lambda_i(\phi^1)|^{-1} d\phi^1 \right) \end{aligned}$$

implies

$$h_n = h_{n-1} - \left(\frac{4d\pi^{\frac{d-1}{2}}}{(d+1)(d-1)\Gamma(\frac{d-1}{2})} \int_0^{L(\gamma)} \prod_{i=2}^{d} |\lambda_i(\phi^1)|^{-1} d\phi^1 \right) h_{n-1}^{\frac{d+1}{2}}$$

Now let $k = \frac{d+1}{2}$ and $a = \frac{4d\pi^{\frac{d-1}{2}}}{(d+1)(d-1)\Gamma(\frac{d-1}{2})} \int_0^{L(\gamma)} \prod_{i=2}^{d} |\lambda_i(\phi^1)|^{-1} d\phi^1$. Then by the Pólya and Szëgo theorem,

$$n^{\frac{2}{d-1}} h_n \to \left[\frac{d-1}{2} \frac{4d\pi^{\frac{d-1}{2}}}{(d+1)(d-1)\Gamma(\frac{d-1}{2})} \int_0^{L(\gamma)} \prod_{i=2}^{d} |\lambda_i(\phi^1)|^{-1} d\phi^1 \right]^{-\frac{2}{d-1}}$$

That is,

$$h_n \to \left[\left(\frac{2d\pi^{\frac{d-1}{2}}}{(d+1)\Gamma(\frac{d-1}{2})} \int_0^{L(\gamma)} \prod_{i=2}^{d} |\lambda_i(\phi^1)|^{-1} d\phi^1 \right) n \right]^{-\frac{2}{d-1}}$$

Therefore,

$$S_n \to 1 - \left[\left(\frac{2d\pi^{\frac{d-1}{2}}}{(d+1)\Gamma(\frac{d-1}{2})} \int_0^{L(\gamma)} \prod_{i=2}^{d} |\lambda_i(\phi^1)|^{-1} d\phi^1 \right) n \right]^{-\frac{2}{d-1}}$$

as $n \to \infty$.

4.4 Minkowski-type formula near extreme values of a function on compact Riemannian manifold

In this section, we will explore the connection between the Minikowski-type formula and the Laplace method, then explore it's similarity with the Moser-type problem on compact Riemannian manifolds.

Minkowski formula plays a crucial role in understanding the behavior of geometric quantities near extreme values on Riemannian manifolds. Minkowski formula provides a means of calculating the volume or surface area of a convex set and its ϵ-neighborhood. The formula expresses these geometric quantities in terms of integrals involving the curvature of the boundary of the set. By examining the behavior of volume or area measures as they approach maximum or minimum points, Minkowski formula provides insight into the local geometry and curvature of the manifold.

Let's introduce the Minkowski formula. Consider a convex set $\mathcal{D}_0$ with a smooth boundary surface $\partial\mathcal{D}$ of class C^2. In this scenario, the fundamental quadratic forms $Q_1(du, dv)$ and $Q_2(du, dv)$ are well-defined. Define the set $\mathcal{D}_\epsilon = \{x \in \mathbb{R}^3 : d(x, \mathcal{D}_0) \leq \epsilon\}$ as the ϵ-neighborhood of $\mathcal{D}_0$. Then, we have the following expression for the volume $Vol(\mathcal{D}_\epsilon)$:

$$Vol(\mathcal{D}_\epsilon) = Vol(\mathcal{D}_0) + \epsilon Ar(\partial\mathcal{D}) + \epsilon^2 H_1(\partial\mathcal{D}_0) + \epsilon^3 K_1(\partial\mathcal{D}_0)$$

where

$$H_1(\partial\mathcal{D}_0) = \int_{\partial\mathcal{D}_0} H(\sigma)d\sigma$$

$$K_1(\partial\mathcal{D}_0) = \int_{\partial\mathcal{D}_0} K(\sigma)d\sigma.$$

Here, $Vol(\cdot)$ is the volume of the region, $Ar(\cdot)$ is the area of the region, $H(\sigma) = \frac{K_1+K_2}{2}(\sigma)$, and $K(\sigma) = K_1K_2(\sigma)$. Moreover, $K_1(\sigma)$ and $K_2(\sigma)$ denote the principal curvatures of $\partial\mathcal{D}_0$ at the point $\sigma \in \partial\mathcal{D}_0$, while $H(\sigma)$ and $K(\sigma)$ signify the mean and Gaussian curvature at $\sigma \in \partial\mathcal{D}_0$ respectively.

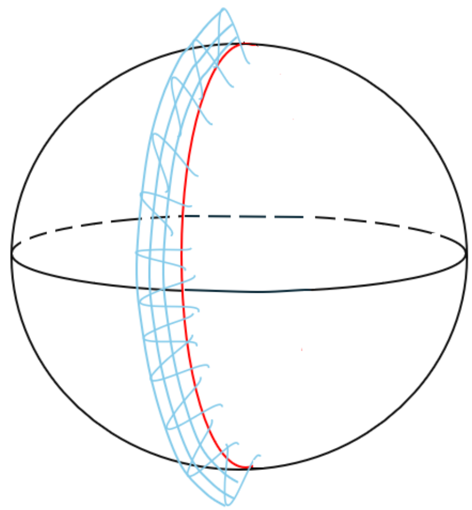

Figure 4.4: This figure shows the non-degenerated global maximum along a meridian (the red curve) on the surface of a sphere.

Let's now turn our attention to the Laplace method. Laplace method is a technique for obtaining the asymptotic behavior of integrals in which the large parameter $t \to \infty$, appears in the exponent of a function $\zeta(s) = e^{ts}$. Let $\gamma = \{(\theta, \varphi) : \varphi = \varphi_*, \theta \in [0, \pi]\}$ be a meridian on the surface of a unit sphere (see Figure 4.4) and let $\mathcal{G} : \mathcal{M} \to \mathbb{R}$ be a smooth function near its maximum points. Suppose the maximum value $\mathcal{G}(\theta, \varphi_*) = 1$ for all $\theta \in [0, \pi]$ on the meridian, such that

$$\mathcal{G}(\theta, \varphi) \sim 1 - K(\theta)(\varphi - \varphi_*)^2 \tag{4.4}$$

where $0 < c_0 < K(\theta) < c_1 < \infty$, and c_0 and c_1 are constants. Here, $\theta \in [0, \pi]$ and $\varphi \in [0, 2\pi]$. Notice that (4.4) is actually the Minkowski formula with only the second order term. Then there exists a neighborhood $\mathcal{U}_\delta, 0 < \delta << 1$ around $\varphi = \varphi_*$, then when $t \to \infty$

$$
\begin{aligned}
I(t) &= \int_{\mathcal{U}_\delta} e^{t\mathcal{G}(x)}\mu(dx) \\
&\sim e^t \int_0^\pi \left(\int_{-\delta}^{\delta} e^{-t(\varphi-\varphi_*(\theta))^2 K(\theta)} d\varphi \right) d\theta \\
&= \frac{e^t\sqrt{\pi}}{\sqrt{t}} \int_0^\pi K(\theta)^{-\frac{1}{2}} d\theta.
\end{aligned}
$$

where μ is the measure of Riemannian manifold.

It's intriguing to observe that the computational intricacies involved in solving optimal stopping problems bear a striking resemblance to the methodological intricacies encountered when calculating integrals using the Laplace method for asymptotics. This similarity shows the deep connection between decision theory and mathematical analysis, shedding light on the underlying symmetries and connections between seemingly disparate fields of study.

CHAPTER 5: MARKOV CHAIN ON COMPACT RIEMANNIAN MANIFOLDS

In this chapter, we will explore the basic principles of Markov chains operating on compact Riemannian manifolds. In this formulation of the Markov stopping time, applicable to discrete-time chains and extending to diffusion processes, the phase space can be arbitrary. We would like to study the stationary optimal stopping problem (i.e. not to fix number of the steps).

Let $\mathcal{M}$ be a compact Riemannian manifold, partitioned into two distinct regions $\mathcal{M}_1$ and $\mathcal{M}_2$. That is $\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2$, $\mathcal{M}_1 \cap \mathcal{M}_2 = \emptyset$. Let $W_0, W_1, W_2, \cdots$ be a Markov chain with two states $\{1, 2\}$ such that the transition matrix is

$$P = \begin{bmatrix} p_1 & q_1 \\ q_2 & p_2 \end{bmatrix}$$

where p_1 is the probability continues staying on state 1, q_1 is the probability to jump from state 1 to state 2, p_2 is the probability continue staying on state 2, q_2 is the probability to jump from state 2 to state 1.

Let $X_0, X_1, X_2, \cdots$ and $Y_0, Y_1, Y_2, \cdots$ be two sequences of i.i.d. uniformly distributed random variables on $\mathcal{M}_1$ and $\mathcal{M}_2$ respectively. Let $Z_0, Z_1, Z_2, \cdots$ be a Markov chain on $\mathcal{M}$ given by

$$Z_i = \begin{cases} X_i & , W_i = 1 \\ Y_i & , W_i = 2 \end{cases}$$

$i = 1, 2 \cdots$. The stationary distribution of W_0 (i.e. the solution of equation $\pi P = \pi$)

has the form

$$\pi_1 = \frac{q_2}{q_1 + q_2}, \qquad \pi_2 = \frac{q_1}{q_1 + q_2}$$

for $W_0 = 1$ and $W_0 = 2$ respectively. We will take this as the initial distribution of W_0.

Let us introduce a small killing probability $\epsilon > 0$, on each step the Markov chain will cease with probability ϵ, yielding a reward of 0. Conversely, with probability $1-\epsilon$, the Markov chain proceeds to the next step. At each point $Z_t \in \mathcal{M}_1$, a decision is made to transition. Specifically, with probability p_1, the chain remains uniformly distributed on $\mathcal{M}_1$, while with complementary probability q_1, it transitions uniformly to $\mathcal{M}_2$. A similar scenario unfolds for $Z_t \in \mathcal{M}_2$. Here, with probability p_2, the chain transitions uniformly within $\mathcal{M}_2$, and with probability q_2, it moves back to $\mathcal{M}_1$, again following a uniform distribution. This framework provides a probabilistic interpretation of our Markov chain operating on $\mathcal{M}$.

Now let the reward function $\mathcal{G} : \mathcal{M} \to \mathbb{R}$ be

$$\mathcal{G}(z) = \begin{cases} h_1 & , \ z \in \Delta_1 \subset \mathcal{M}_1 \\ h_2 & , \ z \in \Delta_2 \subset \mathcal{M}_2 \\ 0 & , \ \text{otherwise.} \end{cases}$$

Let $\delta_1 = \frac{\mu(\Delta_1)}{\mu(\mathcal{M})}$ and $\delta_2 = \frac{\mu(\Delta_2)}{\mu(\mathcal{M})}$ where $\mu(\cdot)$ is the Riemannian measure and $0 < \mu(\mathcal{M}) < \infty$. We will consider $h_1 > h_2$ and $\delta_1 < \delta_2$, where δ_1 and δ_2 represent small parameters that will be compared with the killing probability ϵ. Figure 5.1 illustrates this scenario on the simplest compact Riemannian manifold $[0, 1]$. Our objective is to determine the maximum expectation

$$S = \max_{\tau \geq 0} E[\mathcal{G}(Z_\tau)].$$

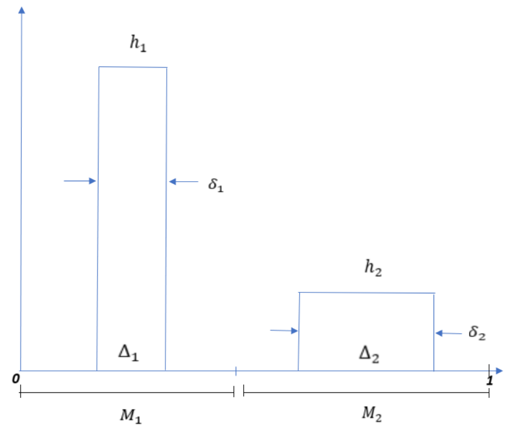

Figure 5.1: This figure illustrates an example of the reward function $\mathcal{G}$ on the partitioned manifolds $\mathcal{M}_1, \mathcal{M}_2$ of $\mathcal{M} = [0,1]$.

Now set

$$S = \pi_1 S_1 + \pi_2 S_2$$

where S_1 represents the optimal value if the Markov chain starts from a point $Z_t \in \mathcal{M}_1$, and S_2 represents the optimal value starting from an initial point $Z_t \in \mathcal{M}_2$. At some moment $t \geq 0$, if $Z_t \in \Delta_1$, i.e., $\mathcal{G}(Z_t) = h_1$, then the decision-maker must stop. If $Z_t \in (\mathcal{M}_1 \setminus \Delta_1) \cup (\mathcal{M}_2 \setminus \Delta_2)$, the decision-maker can continue. Finally, if $Z_t \in \Delta_2$, i.e., $\mathcal{G}(Z_t) = h_2$, the chain will stop with probability α and proceed to the next step with probability $1 - \alpha$. The parameter α is the only variable in the optimization problem. We compute $S = S(\alpha)$ and then find $\max\limits_{0 \leq \alpha \leq 1} S(\alpha) = S_*$ (the optimum).

For the functions $S_1(\alpha), S_2(\alpha)$ which are the optimal results for a fixed α and initial points from $\mathcal{M}_1$ and $\mathcal{M}_2$ respectively, we have the usual Bellman's equations:

$$\begin{cases} S_1 = (1-\epsilon)\delta_1 h_1 + (1-\epsilon)(1-\delta_1)(p_1 S_1 + q_1 S_2) \\ S_2 = (1-\epsilon)\delta_2 h_2 \alpha + [(1-\epsilon)\delta_2(1-\alpha) + (1-\epsilon)(1-\delta_2)] \cdot (p_2 S_2 + q_2 S_1) \end{cases}$$

By solving this linear system, which is inherently non-linear with respect to the parameter α, and employing asymptotic analysis of the solution and its optimization, we aim to determine the solution. Due to the computational complexity of the calculations, we will focus on formulating several qualitative results instead:

Since the chain stops at each step with probability $\epsilon > 0$, that is, the total time of a game has order $\frac{1}{\epsilon}$.

a) If $\frac{1}{\epsilon} >> \frac{1}{\delta_1}$ then the decision-maker has to wait for the first visit of Δ_1.

b) If $\frac{1}{\delta_1} << \frac{1}{\epsilon}$ and $\frac{1}{\delta_2} << \frac{1}{\epsilon}$, then the decision-maker must stop on Δ_2.

c) If $\frac{1}{\epsilon} = \frac{1}{\delta_1} >> \frac{1}{\delta_2}$, then the decision-maker selects an α such that $\frac{1}{\delta_1} = \frac{1}{\alpha\delta_2}$.

REFERENCES

[1] L. Moser, "On a problem of cayley," 1956. Scripta mathematica a quarterly journal.

[2] E. Dynkin and A. A. Yushkevich, "Markov processes: theorems and problems," 1969. Plenum press.

[3] S. Karlin and H. E. Taylor, "A first course in stochastic processes," 1975. Academic Press, p.245-247.

[4] G. Pólya and G. Szëgo, "Problems and theorems in analysis vol 1," 1998. Springer, p.38, Problem 174.

[5] L. J. Savage, "The foundations of statistics," 1954. John Wiley Sons.

[6] L. J. Savage, "The theory of statistical decision," 1954. University of Chicago Press.

[7] V. I. Arnol'd, S. M. Gusein-Zade, and A. N. Varchenko, "Singularities of differentiable maps: Volume 1: Classification of critical points, caustics and wave fronts.," 1986. Springer.

[8] S. A. Molchanov, "Diffusion processes and riemannian geometry," 1975. Russian mathematical surveys, 30(1): 1-63.

[9] M. Morse, "Relations between the critical points of a real function of n independent variables," 1925. Transactions of the American Mathematical Society, 27(3), 345-396.

[10] W. Rudin, "Principles of mathematical analysis," 1976. McGraw-Hill Education.

[11] V.I.Arnol'd, "On the stationary phase method and coxeter numbers," 1973. Russian Math.Surveys 28:5, 19-48.

[12] V. I. Arnol'd, "Remarks on the stationary phase method and coxeter numbers," 1973. Russian mathematical surveys, 28: 19-48.

[13] V. I. Arnol'd, "Integrals of rapidly oscillating functions and singularities of projections of lagrangian manifolds," 1972. Functional Analysis and Its Applications, 6: 222-224.

[14] V. I. Arnol'd, "Normal forms for functions near degenerate critical points, the weyl groups of a_k, d_k, e_k and lagrangian singularities," 1972. Functional Analysis and Its Applications, 6: 254-272.

[15] J. Gilbert and F. Mosteller, "Recognizing the maximum of a sequence," 1966. Journal of the American statistical association, 61: 35-73.

[16] I. Guttman, "On a problem of l.moser," 1960. Canadian Mathematics Bulletin, 3(1): 35-39.

APPENDIX A: MOSER-TYPE PROBLEMS

A.1 Example of Moser-type problem with unbounded random variables

Let $X_1, \cdots, X_n \sim \exp(1)$ be i.i.d. random variables, i.e. $f(x) = e^{-x}\mathbf{1}_{x>0}(x)$. Let $\varphi(x) = x$ such that $S_n = \max_{\tau \leq n} E[X_\tau] \uparrow \infty$.

Proof. By the law of total expectation,

$$
\begin{aligned}
S_n &= \int_{S_{n-1}}^{\infty} xe^{-x}dx + S_{n-1} \cdot \int_0^{S_{n-1}} e^{-x}dx \\
&= (S_{n-1}+1)e^{-S_{n-1}} + S_{n-1}(1-e^{-S_{n-1}}) \\
&= S_{n-1}e^{-S_{n-1}} + e^{-S_{n-1}} + S_{n-1} - S_{n-1}e^{-S_{n-1}} \\
&= S_{n-1} + e^{-S_{n-1}}.
\end{aligned}
$$

That is $S_n = S_{n-1} + e^{-S_{n-1}}$. Now let $S_n = \ln n + \delta_n$ where δ_n is a small error. Then

$$
\begin{aligned}
S_{n+1} &= S_n + e^{-S_n} \\
&= (\ln n + \delta_n) + e^{-\ln n - \delta_n} \\
&= (\ln n + \delta_n) + \frac{1}{n}e^{-\delta_n} \\
&= \ln n + \delta_n + \frac{1}{n}(1 - \frac{\delta_n}{n}) \\
&= \ln n + \frac{1}{n} + \delta_n(1 - \frac{1}{n}) \\
&\sim \ln n + \ln(1 + \frac{1}{n}) \\
&= \ln(n+1).
\end{aligned}
$$

That is $S_{n+1} \sim \ln(n+1)$. It implies

$$S_n \sim \ln(n)$$

$S_n \to \infty$ as $n \to \infty$. $\square$

A.2 Example of Moser-type problem with Beta Distribution

Let $X_1, \cdots, X_n \sim$ i.i.d. beta distribution with parameters $\alpha, \beta > 0$

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$$

where

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

One can calculate the maximum expectation by change of variables as follow.

$$\begin{aligned}
S_n &= \int_{S_{n-1}}^1 x \cdot \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}dx + S_{n-1} \cdot \int_0^{S_{n-1}} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}dx \\
&= \int_{S_{n-1}}^1 (x - S_{n-1} + S_{n-1}) \cdot \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}dx + S_{n-1} \cdot \int_0^{S_{n-1}} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}dx \\
&= \int_{S_{n-1}}^1 (x - S_{n-1}) \cdot \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}dx + S_{n-1} \\
&= \int_{S_{n-1}}^1 [(1 - S_{n-1}) - (1-x)] \cdot \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}dx + S_{n-1} \\
&\sim \frac{1}{B(\alpha,\beta)} \left[(1-S_{n-1}) \int_{S_{n-1}}^1 (1-x)^{\beta-1}dx - \int_{S_{n-1}}^1 (1-x)^{\beta}dx\right] + S_{n-1} \\
&= \frac{1}{B(\alpha,\beta)} \left\{(1-S_{n-1}) \left[-\frac{(1-x)^{\beta}}{\beta}\right]_{S_{n-1}}^1 + \left[\frac{(1-x)^{\beta+1}}{\beta+1}\right]_{S_{n-1}}^1\right\} + S_{n-1} \\
&= \frac{(1-S_{n-1})^{\beta+1}}{\beta(\beta+1)B(\alpha,\beta)} + S_{n-1}.
\end{aligned}$$

We have the fifth step since near $x = 1$, the term $(1-x)^{\beta-1}$ dominates the other terms, so we can approximate $x^{\alpha-1}(1-x)^{\beta-1}$ by $(1-x)^{\beta-1}$. Now let the recursive relation to be the function g(x) as

$$g(x) = \frac{(1-x)^{\beta+1}}{\beta(\beta+1)B(\alpha,\beta)} + x$$

then by the fix point theorem $g(x) = x$ gives

$$x = 1.$$

Since $g'(1) < 1$, g is contractive and $g(S_n) \to 1$ as $n \to \infty$. Let

$$h_n = 1 - g(S_n)$$

implies

$$1 - h_n = \frac{(1 - (1 - h_{n-1}))^{\beta+1}}{\beta(\beta+1)B(\alpha,\beta)} + (1 - h_{n-1})$$

implies

$$h_n = h_{n-1} - \frac{h_{n-1}^{\beta+1}}{\beta(\beta+1)B(\alpha,\beta)}.$$

Now let $k = \beta + 1$ and $a = \frac{1}{\beta(\beta+1)B(\alpha,\beta)}$. Then by the Pólya and Szëgo theorem,

$$n^{\frac{1}{\beta}} h_n \to \left[\frac{1}{(\beta+1)B(\alpha,\beta)}\right]^{-\frac{1}{\beta}}.$$

That is,

$$h_n \to \left[\frac{n}{(\beta+1)B(\alpha,\beta)}\right]^{-\frac{1}{\beta}}.$$

Therefore,

$$S_n \to 1 - \left[\frac{n}{(\beta+1)B(\alpha,\beta)}\right]^{-\frac{1}{\beta}}$$

as $n \to \infty$.