An Alternative Method for Characterization and Comparison of Plant Root Shapes

A thesis submitted to the
College of Graduate and Postdoctoral Studies
in partial pulfillment of the requirements
for the degree of Master of Science
in the Department of School of Environment and Sustainability
University of Saskatchewan
Saskatoon

By
Yujie Pei

Contents

Existed Morphological Descriptors for Root Systems

An Alternatively Mathematical Method for Shape Description

2.1 Monte Carlo Simulations

In this section, several generally utilized numerical methods [?][?] [?] [?] for solving the heat equation, and their limitations in practice are presented. Also, one of the non-deterministic algorithms, Monte Carlo methods (MCM) [?] [?], and its application in approximating the solution of the PDEs are proposed. As the weaknesses and challenges of applying the numerical techniques in solving $2-$dimensional heat equation defined in the real root images with millions of pixels and extremely complex root systems, the alternative fixed-time step Monte Carlo simulations, lattice random walks (LRWs) and Pearson's random walks (PRWs), are designed. The most outstanding advantage of the proposed random walk models is that the integration, named the heat content, can be approximated directly based upon the probabilistic interpretation of Brownian motion and the heat equation. Finally, the methods to analyze the output of the Monte Carlo simulations and solve the sampling-related problems in the simulations are brought up theoretically.

2.1.1 Background

In the subsection **??**, the analytical solutions of the heat equation defined in the annulus with the initial and boundary conditions have been derived. The analytical heat content, calculated by the integration over Ω, implies the geometrical properties of the annulus. However, the analytical method for solving the heat equation has many restrictions, and its applications to practical problems will exhibit difficulties. Firstly, the numerical evaluation of the analytical solutions is usually by no means trivial because they are in the form of infinite series. Secondly, either irregular geometries or discontinuities lead to the complexities, so the explicit algebraic solutions are close to non-existed. Thirdly, the purely analytical techniques can apply strictly only to the linear form of the boundary conditions and to constant diffusion properties [?]. Therefore, numerical methods and computer simulations are more helpful and applicable to find solutions to the heat equation than calculating pure analytical solutions.

Numerical Methods

The techniques for solving initial-boundary value problems (IBVPs) based on numerical approximations have existed for a long time and been developed considerably including the finite-difference method (FDM), finite element method (FEM), finite volume method (FVM), boundary element method (BEM), and so forth.

FDM is frequently utilized to converting the heat equation into a system of algebraically solvable equations [?]. The basic idea is to replace the derivatives in the equation by the difference quotients. For example, the FTCS (Forward Time Centered Space) scheme [?] discretize the Laplace operator in space and the

[Yuge 1] The writing of this part has been revised. Dave, please give me the feedback or comments on it.

time derivative and implement the boundary conditions on the staggered grid for representing the original continuous problem, but it is conditionally stable [**?**]. If the spatial resolution becomes doubled, the time-step should be reduced by a factor of four to maintain the numerical stability, which causes the extremely tiny time-step in the high-resolution calculations. There are three kinds of errors needed to be considered when using FDM. First of all, in the derivation of the finite-difference equations, the higher-order terms in the Taylor series are neglected, constituting the truncation error. If the time and space interval tends to 0, the truncation errors will approach 0, or the FDM is incompatible or inconsistent with the original heat equation [**?**]. Another class of error appearing in FDM, called round-off error, results from the loss of precision due to the computer rounding of decimal quantities. [**?**]. The last type of error is the discretization error, which can be reduced by decreasing the time size, grid size, or both [**?**]. Moreover, DFM becomes less accurate and hard to implement when the problem is defined in the irregular geometries since the heat equation must be transformed before applying the Taylor series.

Unlike the FDM, FEM [**?**] divides the complicated and irregular geometries and boundaries into the union of smaller and simpler subdomains or finite elements [**?**], e.g. lattice, triangle, curvilinear polygons, etc. Each subdomain is locally represented by the element equation, continuous piecewise shape functions, which are finally assembled into a larger system of algebraic equations for modelling the entire problem. The numerical solution can be obtained by minimizing the associated error function to meet the certain specification of the accuracy. The smaller size of the finite element mesh, the more accurate the approximate solution. FEM has great flexibilities or adaptivity [**?**]. For instance, FEM can provide higher fidelity or accuracy in a specific local region and keep elsewhere identical. Nevertheless, FEM requires an amount of human involvement in building the FE model, checking the result, detecting and updating the model design. Moreover, compared with FDM, FEM demands a longer execution time and a larger amount of input data.

FVM, closely related to FEM, converts the original heat equation into the integral forms [**?**]. However, the accuracy of FVM is related to the numerical integration over time and space dimensions. Unlike the domain-type methods (e.g. FDM, FEM, FVM, etc.), BEM transforms the heat equation into an integral equation over the boundary of the domain using the boundary integral equation method [**?**]. Especially, when the domain extends to infinity or the boundary is complex, BEM is more efficient in computation than other methods because of the smaller surface or volume ratio [**?**] since it only discretizes the boundary and fits the boundary values into the integral equation [**?**]. However, it is arduous to solve the matrics generated in BEM, since they are generally unsymmetric and fully populated [**?**].

In summary, all the described numerical methods have an intrinsically similar feature - mesh discretization in the time and space dimension. In this thesis, the heat equation is defined in a 2-dimensional domain, bounded by the border of the image and that of the root system, with millions of pixels, the extremely complex roots and various boundary conditions. It may be possible to calculate the heat content contained in this domain by the numerical integration of the solutions approximated by those numerical techniques. However, some practical problems have to be taken into consideration. For instance, the far more efforts

are required when applying FDM and FVM because of the complicated boundary of the roots and non-continuous issues. Although the whole 2-dimensional root image can be viewed as a discretized domain, it is still time-consuming and challenging to trace and identify the boundary of roots, label the nodes, and generate the coordinates and connectivities among the nodes in the preprocessing stage of FEM. The finer discretization, the more accurate approximated solutions of the original IBVP and the longer computational time spent by the numerical methods. More importantly, the heat content, which is the integration of the numerical solution over the space dimension, should also be approximated numerically resulting in extra effort and errors.

Probabilistic Interpretation

In the subsection **??**, the heat equation describes the temperature distribution of a homogeneous and isotropic domain [**?**], and its solution characterizes how the temperature changes over the position and time. From the probabilistic perspective, the heat equation and its solution can also be understood by the Brownian motion [**?**]. The Brownian motion also called the Wiener process, is a continuous-time and continuous-space stochastic process [**?**] with the continuous sample paths and stationary independent increments [**?**]. This process also has the Markov property: the future state depends only on the present state [**?**]. In the probability theory, if a large number of free particles undergoing the Brownian motion independently, the density of particles at a specific time becomes a deterministic process called diffusion, which satisfies the heat equation [**?**][**?**].

Survival Probability For simplicity, we only investigate the probabilistic interpretation of the heat equation defined in the annulus with the boundary and initial conditions as same as described in the subsection **??**. Consider a particle undergoing the Brownian motion from $(\hat{r_0}, \theta_0) \in \Omega$ at $\tau = 0$, and let $\rho(\hat{r}, \theta, \tau | \hat{r_0}, \theta_0, 0)$ be the conditional probability of finding the particle at $(\hat{r}, \theta) \in \Omega$ at time $\tau > 0$. Moreover, particle's initial position $(\hat{r_0}, \theta_0)$ is distributed uniformly over the whole domain Ω. $\rho(\hat{r}, \theta, \tau | \hat{r_0}, \theta_0, 0)$ satisfies the following equations

$$\rho_\tau = \rho_{\hat{r}\hat{r}} + \frac{1}{\hat{r}}\rho_{\hat{r}} + \frac{1}{\hat{r}^2}\rho_{\theta\theta} \qquad \text{for } (\hat{r}, \theta) \in \Omega \tag{2.1}$$

$$\rho = 0 \qquad \text{for } (\hat{r}, \theta) \in \partial\Omega_1 \tag{2.2}$$

$$\rho_{\hat{r}} = 0 \qquad \text{for } (\hat{r}, \theta) \in \partial\Omega_2 \tag{2.3}$$

$$\tag{2.4}$$

The "local" survival probability $S(\tau | \hat{r_0}, \theta_0, 0)$, represents the probability that a particle, localized at $(\hat{r_0}, \theta_0)$ at $\tau = 0$, keeps diffusing in Ω at time $\tau > 0$ and is unabsorbed by the boundary Ω_1.

$$S(\tau | \hat{r_0}, \theta_0, 0) = \iint_\Omega \hat{r}\rho(\hat{r}, \theta, \tau | \hat{r_0}, \theta_0, 0) d\hat{r} d\theta \tag{2.5}$$

Our interest is the "global" survival probability $S(\tau)$, the average of the local survival probability over Ω, expressed as

$$S(\tau) = \frac{1}{|\Omega|} \iint_{\Omega} \hat{r_0} S(\tau|\hat{r_0}, \theta_0, 0) d\hat{r_0} d\theta_0 \tag{2.6}$$

Eq. 2.5 and Eq. 2.6 reveal that the heat content $Q_\Omega(\tau)$ expressed in Eq. **??** is proportional to the survival probability $S(\tau)$ [**?**].

Mean First-Passage Time The first passage phenomena play a fundamental role in the stochastic processes triggered by a first-passage event [**?**]. In this thesis, one of the essential first-passage-related quantities is the first-passage time or the first-hitting time [**?**], which is the time taken by particle undergoing the Brownian motion from an initial position to any sites of Ω_1 for the first time. Particles' mean first-passage time $\langle\tau\rangle$, also called the average first-passage time, has a closed relationship with the survival probability [**?**]

$$\begin{aligned} \langle\tau\rangle &= \int_0^\infty \tau dS(\tau) \\ &= \sum_{n=1}^{\infty} \frac{4}{\mu^2 - 1} \frac{1}{\lambda_{0,n}^2 \left\{ \left[\frac{J_0(\sqrt{\lambda_{0,n}})}{J_0'(\mu\sqrt{\lambda_{0,n}})} \right]^2 - 1 \right\}} \end{aligned} \tag{2.7}$$

From Eq. 2.7 and Eq. **??**, it is clear that $\langle\tau\rangle$ implies an overall property of the annulus since it only depends on the radius ratio of annulus μ.

Brownian Motion and Random Walks Brownian motion, the irregular motion of individual particles, has been existed for a long time before the random-walk theory was developed. At the beginning of the twentieth century, the term, random walk, was initially proposed by Karl Pearson [**?**]. He utilized the isotropic planar random flights to model how mosquitoes migrate and invade randomly in the cleared jungle regions. At each time step, the mosquito moves to a random direction with a fixed step length. Rayleigh [**?**] answered Pearson's question in the same year, the distributions of mosquitos after many steps have been taken, is identical to superposition the sound vibrations with unit amplitude and arbitrary phase [**?**]. At almost the same time, Louis Bachelier [**?**] designed a model for the financial time series by the random walks. He also explored the connection between discrete random walks and the continuous heat equation. During the development of random-walk theory, many other scientific fields, including the random processes, random noise, spectral analysis, and stochastic equations, were developed by some physicists [**?**] [**?**] [**?**]. The continuous Brownian motion can be considered as the limit of the discrete random walks as the time and space increments go to zero [**?**][**?**].

Lattice Random Walks (LRWs) Let us consider a large number of particles performing the simple random walk on the d-dimensional integer grid $\mathbb{Z}^d$. It is a discrete-space and discrete-time symmetric hopping

process [?] on the lattice. At each time step, the particle moves to one of its $2d$ nearest neighbours with probability $\frac{1}{2d}$. If $d \leq 2$, the random walk is recurrent [?], which means the particle will return to its origin infinitely often with the probability 1. If $d \geq 3$, the random walk is transient [?], which indicates the particle will return to its origin only finitely often with the probability 1 [?].

2-dimensional lattice random walks (LRWs) are considered and bec.

Let $\triangle l$ be the distance between two sites and δ be the time step. Define $p(x, t|x_S)$ as the conditional probability of a particle to be in position x at time t starting from x_S at time $t = 0$. $p(x, t + \delta|x_S)$ is defined as probability of a particle to be in position x at time $t + \delta$ starting from x_S at time $t = 0$, [?]

$$p(x, t + \delta|x_S) = \frac{p(x - \triangle l e_x, t|x_S) + p(x + \triangle l e_x, t|x_S) + p(x - \triangle l e_y, t|x_S) + p(x + \triangle l e_y, t|x_S)}{4} \tag{2.8}$$

where e_x and e_y are unit vectors of $x-$ axis and $y-$ axis, respectively.

When $\delta \to 0$, $\triangle l \to 0$, and $\delta \sim (\triangle l)^2$,

$$p(x, t|x_S) + \delta p(x, t|x_S) = p(x, t|x_S) + \frac{(\triangle l)^2}{4}(\frac{\partial^2 p(x, t|x_S)}{\partial x^2} + \frac{\partial^2 p(x, t|x_S)}{\partial y^2}) \tag{2.9}$$

Finally, the 2-dimensional heat equation is

$$\frac{\partial p(x, t|x_S)}{\partial t} = \frac{(\triangle l)^2}{4\delta}(\frac{\partial^2 p(x, t|x_S)}{\partial x^2} + \frac{\partial^2 p(x, t|x_S)}{\partial y^2}) \tag{2.10}$$

where $D = \frac{(\triangle l)^2}{4\delta}$ is the diffusion coefficent. This derivation shows a tight relationship between lattice random walks and diffusion.

Pearson's Random Walks (PRWs) Based on Pearson's problem and Rayleigh's answer, Stadje [?] and Masoliver et al. [?] considered a two-dimensional continuous-time and continuous-space random walk, defined as Pearson's random walks (PRWs) in this thesis. In PRWs, particle moves with constant speed and with random directions distributed uniformly in $[0, 2\pi)$. Moreover, the lengths of the straight-line paths and the turn angles are stochastically independent. If the mean step length approaches zero and the walking time is big enough, the behaviours of particles in PRWs weakly converges to the Wiener Process [?], which satisfies the traditional heat equation.

Monte Carlo Simulations For Solving PDEs

Monte Carlo methods (MCMs), the commonly used computational techniques, aim to generate samples from a given probability distribution, estimate the functions' expectations under this distribution, and optimize the complicated objective functions by using random numbers [?] [?]. MCMs can be used to solve the IBVPs by generating the random numbers to simulate the successive positions of the trajectory of a stochastic process at fixed instants [?][?], since the original continuous problem can be represented by the probabilistic interpretation and the solution can be approximated by the expectation of some functional of the trajectories

of a stochastic process [**?**][**?**]. Therefore, unlike the numerical techniques proposed in the subsubsection 2.1.1, the nondeterministic Monte Carlo simulations are grid-free on the domain, boundary, and the boundary conditions of the problem [**?**].

Monte Carlo simulations have been applied frequently in solving the elliptic partial differential equation, for example, the Laplace's and Poisson's equations [**?**] [**?**] [**?**]. For example, let $u(P_0)$ be the value of the solution of the elliptic partial differential equation at a specific point P_0 in a bounded domain. $u(P_0)$ can be estimated by a point P_1, which is sampled uniformly on the largest circle C_0 centered at P_0 with radius r_0 lying entirely in the domain. If P_1 gets closed to the target bounday within an error, $u(P_1)$ is known, and it can be considered as a particle's estimate of $u(P_0)$ by multiplying the particle's statistical weight. If not, $u(P_1)$ is should be estimated in the same way as $u(P_0)$, that is, P_2 is sampled uniformly on the largest circle C_1 centered at P_1 with radius r_1 lying entirely in the domain. Check the position of P_2, and the procedure will be repeated until the simulation is terminates on the traget boundary, which is defined as one particle's estimate of $u(P_0)$. Finally, averaging a larger number of one-particle estimates, $u(P_0)$ will be more accurate.

However, Monte Carlo simulations are barely applied in solving parabolic and hyperbolic partial differential equations, such as the $1-$ dimensional and $2-$ dimensional time-dependent heat problem. As introduced in the papers [**?**] [**?**], after obtaining the probabilistic interpretation of the finit-difference approximation of the heat equation, the solution of the heat equation at a specific space-time point can be approximated by averaging a large number of random-walking particles, whose trajectories are simulated by the Monte Carlo methods untill they hit the any sites of the target boundary. However, the drawbacks of this method are obvious. Firstly, there is the error appeared in the finit-difference approximation. Secondly, there has the statistical sampling errors inherent in the Monte Carlo simulations. Last but not least, it will be time-comsuming to evaluate the solution defined in the whole domain, the real root images with millions pixels in this thesis, since this method can only be used to approximate the solution of the heat equation at one point at a time.

2.1.2 Algorithms of Random Walks

In the subsection 2.1.1, the practical challenges of solving the heat equation defined in the real root images with millions pixels by numerical methods and Monte Carlo simulations are revealed and the probabilistic interpretaion of the heat equation, survival probability, and random walks are introduced. From Eq. 2.6 and Eq. **??**, it is easy to explore that the final goal in this thesis is the aproximation of the heat content, or the survival probability, defined as the intergration of the solution of the heat equation over the whole domain, which is only dependent on the time variable. Moreover, our interest is only related to the time when the first-passage event happens in the stochastic process, that is, the first-passage time. Therefore, instead of trying to approximate the solution of the heat equation and the integration of the solution, two algorithms of random walks are designed in this subsection to mimic particles' first-passage time by $2-$ dimensional fixed-time step Monte Carlo simulations in the real $2-$ dimensional root images for approximating the integration

as expressed in Eq. 2.6 and Eq. **??**.

Lattice Random Walks

Pearson's Random Walks

2.1.3 Output Analysis

The output of the fixed-time step Monte Carlo simulations are particles' first passage time t, which is the number of steps taken by the particles hitting any positions of the target for the first time. Since first-hitting-time models are a sub-class of survival analysis in statistics [**?**], it is straightforwad to use the Kaplan-Meier estimator to estimate the survival function $S(t)$ [**?**] of the numerical simulation, which provide the probability that a particle remains wandering beyound any specified time. Moreover, the pointwise upper and lower confidence interval can also calculatated by the Greenwood's exponential formula [**?**]. In this subsection, the Kaplan-Meier estimator and confidence interval of $S(t)$ are introduced theoretically. However, in practice, the existed Python module, lifeline [**?**], will be used to implement the estimation. After obtaining the estimated survival function of the fixed-time step Monte Carlo simulations, the scaling relationship between t and τ is derived for the valiation of research methodology in the next chapter.

Kaplan-Meier Estimator

The general definition of the survival time is the time starting from a specified point to the occurrence of a given event [**?**], such as death, pregnancy, job loss, etc. Also, the analysis of the group of survival data is called survival analysis [**?**]. In the survival analysis, three kinds of situations will affect the subjects' survival time [**?**]. Firstly, the subjects are uncooperative and refused to continue to participate in the research. Secondly, some subjects do not experience the event before the end of the study, but they would have experienced the event if they keep being observed. Finally, the researchers lose touch with the subjects in the middle of the investigation. In practice, since these subjects have partial information about survival, the scientists will label the above circumstances as censored observations [**?**] instead of ignoring them and decreasing sample size.

In clinical trials or community trials, Kaplan-Meier Estimator [**?**], a non-parametric analysis, is a commonly applied statistical method in the survival analysis for the measurement of the fraction of the survival time after the treatment [**?**] and for generating the corresponding survival curve. It also works well with the mentioned three difficult situations. With various assumptions [**?**] [**?**], the Kaplan-Meier survival curve can be created and provides the probability of surviving in a given length of time while considering the time in many small intervals [**?**].

Let $0 < t_1 < t_2 < ...$ be the distinct increasing observed times, or the number of steps taken by the particle countering the absorbing boundary, in the sample. Let n_i be the number of particles who either have not yet stopped moving up to time t_i or else who are absorbed on the target boundary at time t_i in

the simulations. Let d_i the number of particles hitting the target boundary at time t_i. The Kaplan-Meier or product-limit estimator $\widehat{S}(t)$ of the survival function of the numerical simulation $S(t)$ is [**?**]

$$\widehat{S}(t) = \prod_{i:t_i \leq t} (1 - \frac{d_i}{n_i}) \tag{2.11}$$

Confidence Interval

The upper and lower $(1-\alpha) \times 100\%$ confidence intervals of the survival function $S(t)$ for a fixed time t was firstly proposed and derived by Greenwood in 1926 [**?**],

$$\widehat{S}(t) \pm z_{\alpha/2} \sqrt{\widehat{Var}[\widehat{S}(t)]} \qquad \text{where} \tag{2.12}$$

$$\widehat{Var}[\widehat{S}(t)] = \widehat{S}(t)^2 \sum_{t_i \leq t} \frac{d_i}{n_i(n_i - d_i)} \tag{2.13}$$

Note, z_α is the $\alpha-$th quantile of the normal distribution.

In 1999, Hosmer and Lemeshow [**?**] developed the exponential Greenwood formula based on the earlier works of Kalbfleisch and Prentice [**?**], which provides an asymmetric confidence interval for $S(t)$

$$e^{-e^{c_+(t)}} < S(t) < e^{-e^{c_-(t)}} \qquad \text{where} \tag{2.14}$$

$$c_\pm(t) = log(-log\widehat{S}(t)) \pm z_{\alpha/2}\sqrt{\widehat{V}} \qquad \text{and} \tag{2.15}$$

$$\widehat{V} = \frac{1}{(log\widehat{S}(t))^2} \sum_{t_i \leq t} \frac{d_i}{n_i(n_i - d_i)} \tag{2.16}$$

Note, if $c1 < c2$, there has $e^{-e^{c_2}} < S(t) < e^{-e^{c_1}}$.

Compared with the traditional Greenwood confidence intercal calculation, the exponential Greenwood formula will make sure that the endpoints in Eq. 2.14 lie in $(0, 1)$, while the endpoints in Eq. 2.12 could be negative or larger than 1 [**?**].

Relationship between t and τ

Particles' average one-step displacement $\triangle l$ in the fixed-time step Monte Carlo simulations, LRWs and PRWs, are associated with the time step is δ:

$$\triangle l = 2\sqrt{D\delta} \tag{2.17}$$

where D is the diffusion coefficent.

Eq.(2.15) implies that the time step δ must be designed small enough to make sure that $\triangle l$ is shorter than the smallest geometrical features of the boundaries. Thus, $\triangle l$ should equal or be less than one-pixel

size in the simulations. Furthermore, the δ is regarded as a fundamental bridge between particles' number of steps t and unitless continuous-time τ,

$$\tau = t\delta = \frac{(\triangle l)^2 t}{4D} \tag{2.18}$$

where D is 1.

When running the LRWs in the annulus, $\triangle l$ is always $\frac{1}{100}$ since particle's step length is as same as one-pixel size. However, $\triangle l$ is related to the particle's step length in PRWs. If particle's step length is 0.5, a half of a pixel, then $\triangle l$ equals $\frac{1}{100} \times \frac{1}{2} = \frac{1}{200}$. Similiarly, when the step length in PRWs is 0.1, then $\triangle l$ is $\frac{1}{1000}$.

2.1.4 Sample Size Determination

Based on the law of large number (LLN) [**?**], as the sample size approaches infinity, the sample mean tends to get closer to the true population mean with the high probability. Moreover, the central limit theorem (CLT) [**?**] illustrates how the sampling distribution of the mean changes as a function of the sample size and how much more reliable a large experiment is. On the downside, the increasing number of trials results in the higher cost of performing the simulation, which is the major drawback of the fixed time-step Monte Carlo simulations. Therefore, it is necessary to conduct the minimum number of simulation runs to achieve a desired degree of precision.

General Method

There are five parts of the standard way to determine the sample size in the Monte Carlo simulations. Firstly, simulating with a certain amount of samples. Secondly, repeating the simulation several times with the sample size as the same as the first step. Thirdly, increasing the number of samples and implementing the first two steps again. Fourthly, running a regression analysis of the variability of the sample statistic as a function of sample size. Finally, estimating the sample size that will result in any desired level of convergence by some probabilistic inequalities, including Chebyshev's inequality [**?**], Cantelli's inequality [**?**], Vysochanskij–Petunin inequality [**?**], etc.

Dvoretzky–Kiefer–Wolfowitz (DKW) inequality

Although the sample size estimated by the general method does not depend on the geometries characterized by the random walk models, how long the simulation will run is unknown, and the final calculated sample size will be larger than the really necessary value sometimes. Therefore, an alternative method, named the Dvoretzky–Kiefer–Wolfowitz (DKW) inequality [**?**], is proposed for the sample size determination without simulating random walk models and considering the shape of objects in the images.

Let $F_N(x)$ denote the empirical distribution functions (empirical CDF) for a sample of N real-valued *i.i.d.* random variables, $X_1, ..., X_N$, with continuous cumulative distrbution function (CDF) $F(x)$. The DKW

inequality, as expressed in Eq.(2.18), bounds the probability that the random function $F_N(x)$ differs from the true $F(x)$ by more than a given constant ε [**?**].

$$Pr(\sup_{x\in\mathbb{R}} |F_N(x) - F(x)| > \varepsilon) \leq 2e^{-2N\varepsilon^2} \quad for\ every\ \varepsilon > 0 \tag{2.19}$$

The equally spaced confidence bounds or simultaneous band around the F_N encompassing the entire $F(x)$ can be expressed by

$$F_N(x) - \varepsilon \leq F(x) \leq F_N(x) + \varepsilon \tag{2.20}$$

On the other hand, assume the simultaneous band produced by Eq.(2.19) containing the $F(x)$ at a given confidence level $1 - \alpha$, the interval ε can be calculated by

$$\varepsilon = \sqrt{\frac{\ln \frac{2}{\alpha}}{2N}} \tag{2.21}$$

Given a converge probability α and constant ε, it is straightforward to estimate the sample size N in the fixed-time step Monte Carlo simulations in any images by Eq.(2.20).