Wiener Measure and Donsker's Theorem

Chia-Min Wei

December 30, 2024

1 Random Functions

Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space, $(C, \mathcal{C})$ the space of all continuous functions on $[0,1]$ equipped with the sup norm, and let X map Ω into C. For any $\omega \in \Omega$, $X(\omega)$ is a continuous function in C. Therefore, X is called a random function.

Define a mapping $w : C \times \mathbb{R}_+ \to \mathbb{R}_+$ by

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

Note that fixing any $\delta > 0$, the map $w(\cdot;\delta) : C \to \mathbb{R}_+$ is a continuous mapping.

Theorem 1 (Arzela-Ascolí): A subset A of C is relatively compact if and only if the set is

(i) bounded at one point: $\sup_{x\in A} |x(0)| < \infty$.

(ii) uniformly equicontinuous: For any $\epsilon > 0$, there exists $\delta > 0$ such that for all $x \in A$ and $|t - s| < \delta$,

$$|x(t) - x(s)| < \epsilon.$$

Or, alternatively,

$$\lim_{\delta\to 0} \sup_{x\in A} w(x;\delta) = 0.$$

The following result gives a necessary and sufficient condition for a sequence of probability measures $\{\mu_n\}$ on $(C, \mathcal{C})$ to be tight, and thus relatively compact by Prohorov's Theorem.

Theorem 2: The sequence $\{\mu_n\}$ is tight if and only if these two conditions hold:

(i) For each positive η, there exists an a and an n_0 such that

$$\mu_n(x : |x(0)| \geq a) \leq \eta, \quad n \geq n_0.$$

(ii) For each positive ϵ and η, there exists $0 < \delta < 1$ and an n_0 such that

$$\mu_n(x : w(x;\delta) \geq \epsilon) \leq \eta, \quad n \geq n_0.$$

This can be put in a more compact form:

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \mu_n(x : w_x(\delta) \geq \epsilon) = 0.$$

Proof. Suppose $\{\mu_n\}$ is tight. Given η, choose a compact K such that $\mu_n(K) \geq 1-\eta$. Then by Arzela-Ascolí, there exists a such that $K \subset \{x : |x(0)| < a\}$, and also $\delta > 0$ such that $K \subset \{x : w(x;\delta) < \epsilon\}$. Hence, these two inqualities hold.

Now suppose the two inqualities hold. Fix $\eta > 0$. Since a single probability measure is tight, it satisfies the two inequalities for some a and δ. So, without loss of generality, we may assume that $n_0 = 1$ by increasing a or decreasing δ. Choose a so that the set $B = \{x : |x(0)| \leq a\}$ has probability $\mu_n(B) > 1 - \eta$ for all n. Then for each k, choose δ_k so that $A_k = \{x : w(x, \delta_k) < 1/k\}$ has probability $\mu_n(A_k) > 1 - \eta/2^k$ for all n. Now consider the set $A = B \cap \bigcap_{k \geq 1} A_k$. We have $\mu_n(A) \geq 1 - 2\eta$ for all n. Moreover, A is relatively compact, and thus $\overline{A}$ is compact with $\mu_n(\overline{A}) \geq 1 - \eta$. □

Now for each random function X, let $\mu_X = \mathrm{P}\, X^{-1}$, the distribution over C induced by X. Also, choose any $t_1, ..., t_k \in [0, 1]$, $(X(t_1), ..., X(t_k))$ is a random vector that takes values in $\mathbb{R}^k$.

Let $\{X_n\}$ be a sequence of random functions. We write $X_n \Rightarrow X$ if $\mu_n = \mu_{X_n} \Rightarrow \mu = \mu_X$. The following result associates the weak convergence of random vectors with the weak convergence of random functions.

Theorem 3: $X_n \Rightarrow X$ if and only if

(i) $(X_n(t_1), X_n(t_2), ..., X_n(t_k)) \Rightarrow (X(t_1), X(t_2), ..., X(t_k))$ for all $t_1, ..., t_k \in [0, 1]$.

(ii) $\lim_{\delta \to 0} \limsup_{n \to \infty} \mu_n(x : w_x(\delta) \geq \epsilon) = 0$.

Proof. Suppose (i) and (ii) holds. By Prohorov's Theorem and the fact that the collection of finite dimensional sets is a separating class, it suffices to prove that $\{\mu_n\}$ is tight. Now by (i), the sequence of probability measures $\{\mu_n \pi_0^{-1}\}$ on $\mathbb{R}$ converges weakly to $\mu\pi_0^{-1}$, and therefore condition (i) in Theorem 2 holds. Together with (ii), we have shown that $\{\mu_n\}$ is tight. The opposite direction is obvious. □

Theorem 4: Suppose that $0 = t_0 < t_1 < ... < t_v = 1$ and

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

Then for arbitrary x,

$$w_x(\delta) \leq 3 \max_{1 \leq i \leq v} \sup_{t_{i-1} \leq s \leq t_i} |x(s) - x(t_{i-1})|,$$

and for arbitrary probability measure μ,

$$\mu(x : w_x(\delta) \geq 3\epsilon) \leq \sum_{i=1}^{v} \mu \left(x : \sup_{t_{i-1} \leq s \leq t_i} |x(s) - x(t_{i-1})| \geq \epsilon \right) \tag{1}$$

Remark: Note that it is not required that $t_v - t_{v-1} < \delta$ in this theorem. This is crucial when we use this theorem to prove Lemma 1.

2 Wiener Measure

Notation

A projection π_t defined by $\pi_t(x) = x(t)$ is a random variable from $(C, \mathcal{C}, \mu)$ to $(\mathbb{R}, \mathcal{R})$. Exploiting notation, we will later write $x_t \coloneqq \pi_t$. Think of $t \in [0, 1]$ as time. Then, $\{x_t : t \in [0, 1]\}$ is a stochastic process whose underlying probability space is $(C, \mathcal{C}, \mu)$. Namely, once $x \in C$ is determined, the whole process is determined, and takes the value $\pi_t(x) = x(t)$. For any sequence of times $\{t_k\}$, $\{x_k\} \coloneqq \{x_{t_k}\}$ is a stochastic process. Similarly, once x is determined, the whole sequence $\{x_k\}$ is determined.

2.1 The Wiener Measure

The Wiener measure, denoted by W, is a probability measure on $(C, \mathcal{C})$, that satisfies the two conditions

(i) For each $t \in [0, 1]$,
$$W(x_t \leq \alpha) = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\alpha} e^{-\alpha/2t} \, du.$$

(ii) For any $0 \leq t_1 \leq t_2 \leq ... \leq t_k = 1$, the random variables

$$x_{t_2} - x_{t_1}, x_{t_3} - x_{t_2}, ..., x_{t_k} - x_{t_{k-1}}$$

are independent under W.

If these two conditions hold, for any $t > s$, we have $x_t = x_s + (x_t - x_s)$ where x_s and $x_t - x_s$ are independent. Therefore, by dividing the characteristic function of x_t by the characteristic function of x_s, we get the distribution of $x_t - x_s$.

(iii) For any $t > s$,

$$W(x_t - x_s) = \frac{1}{\sqrt{2\pi(t-s)}} = \int_{-\infty}^{\alpha} e^{-\alpha/2(t-s)}\, du.$$

Theorem 5 (Existence of Wiener Measure): There exists a probability measure on $(C, \mathcal{C})$ that satisfies the above properties.

Remark: Such measure is unique because the collection of finite dimensional sets in C is a separating class.

Now let W denote not only the probability measure on $(C, \mathcal{C})$, but also any random function that has W as its distribution. Write $W_t = W(t) = \pi_t \circ W$. Then the properties (i)-(iii) can be rewritten as

(i) For each $t \in [0, 1]$,

$$W_t \sim \mathcal{N}(0, t).$$

(ii) For any $0 \le t_1 \le t_2 \le ... \le t_k = 1$,

$$W_{t_1} - W_{t_2}, W_{t_2} - W_{t_1}, ..., W_{t_k} - W_{t_{k-1}}$$

are independent.

(iii) For any $t \ge s$,

$$W_t - W_s \sim \mathcal{N}(0, t - s).$$

Or putting in another way, for any $t_1 \le t_2 \le ... \le t_k$, $(W_0, W_1, ..., W_k)$ is jointly normal with distribution $(N_1, N_1 + N_2, ..., N_1 + ... + N_k)$ where $N_j \sim \mathcal{N}(0, t_j - t_{j-1})$ and N_j's are independent.

2.2 Construction of Wiener Measure

Our goal is to construct a sequence of random functions $\{Y^n\}$ that is tight and has the described finite dimensional properties in the limit: for any $t_1 < t_2 < t_3 < ... < t_k$,

$$(Y^n_{t_1}, ..., Y^n_{t_k}) \Rightarrow (N_1, N_1 + N_2, ..., N_1 + ... + N_k),$$

where $N_j \sim \mathcal{N}(0, t_j - t_{j-1})$ and N_j's are independent.

We start with a sequence of i.i.d. random variables $\xi_1, ..., \xi_n$ with mean 0 and finite variance σ on the same probability space. Let $S_n = \xi_1 + ... + \xi_n$, and define Y^n in the following way:

$$Y^n(t) = \frac{1}{\sigma\sqrt{n}} S_{\lfloor nt \rfloor} + (nt - \lfloor nt \rfloor) \frac{1}{\sigma\sqrt{n}} \xi_{\lfloor nt \rfloor + 1} \tag{2}$$

for all $t \in [0, 1]$. The idea is simple. Fixing n, first we partition $[0, 1]$ into $n - 1$ equal segments with n endpoints $0 < 1/n < 2/n < ... < (n-1)/n < 1$. Then we set

$$Y^n\left(\frac{j}{n}\right) = \frac{1}{\sigma\sqrt{n}} S_j.$$

Next, for any t between the endpoints $\frac{j-1}{n}$ and $\frac{j}{n}$, say $t = (j + a)/n$, where $a < 1$, we take the convex combination of the values at the endpoints:

$$Y^n(t) = aY^n\left(\frac{j+1}{n}\right) + (1-a)Y^n\left(\frac{j}{n}\right).$$

The rightmost term in Equation 2 converges to 0. On the other hand, since $\lfloor nt \rfloor/(nt) \to 1$, by Lindeberg's CLT,

$$\frac{S_{\lfloor nt \rfloor}}{\sqrt{n}\sigma} = \sqrt{t} \frac{S_{\lfloor nt \rfloor}}{\sqrt{nt}\sigma} \Rightarrow \sqrt{t}N,$$

where N is the standard normal distribution. We conclude that for all $t \in [0, 1]$,

$$Y_t^n = Y^n(t) \Rightarrow N(0, t).$$

Similarly, for any $t > s$, when considering the asymptotic distribution of $(Y_s^n, Y_t^n - Y_s^n)$, we only have to consider the asymptotic distribution of

$$\left(\frac{1}{\sigma\sqrt{n}} S_{\lfloor ns \rfloor}, \frac{1}{\sigma\sqrt{n}} (S_{\lfloor nt \rfloor} - S_{\lfloor ns \rfloor})\right),$$

which again by Lindeberg's CLT is

$$(N_1, N_2)$$

where $N_1 \sim \mathcal{N}(0, t)$, $N_2 \sim \mathcal{N}(0, t - s)$, and N_1, N_2 are independent. By the continuous mapping theorem, we conclude that

$$(Y_s^n, Y_t^n) \Rightarrow (N_1, N_1 + N_2).$$

Using the same method (but with more notation), we can prove that for any $t_1 < t_2 < t_3 <$

... $< t_k$,

$$(Y^n_{t_1}, Y^n_{t_2}, ..., Y^n_{t_k}) \Rightarrow (N_1, N_1 + N_2, ..., N_1 + ... + N_k),$$

where $N_1 \sim \mathcal{N}(0, t_1)$ and $N_k \sim \mathcal{N}(0, t_k - t_{k-1})$.

Now all we need is to prove that $\{Y^n\}$ defined by Equation 2 is tight. We will make use of the following lemma:

> **Lemma 1:** Let $\{\xi_n\}$ be stationary (fixing j, $(\xi_k, ..., \xi_{k+j})$ has the same distribution for all k) and suppose that Y^n is defined by Equation 2. If
>
> $$\lim_{\lambda \to \infty} \limsup_{n \to \infty} \lambda^2 \, \mathrm{P}(\max_{k \le n} |S_k| \ge \lambda \sigma \sqrt{n}) = 0, \tag{3}$$
>
> then $\{Y^n\}$ is tight.

Proof. We make use of Theorem 2. Since $Y^n_0 = 0$ for all n, condition (i) of Theorem 2 holds. Now condition (ii) of Theorem 2 translates into

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \mathrm{P}(w(X^n; \delta) \ge \epsilon) = 0.$$

For any n and δ, we shall find a suitable way of choosing $m, v \in \mathbb{N}$ and set $t_i = i(m/n)$ for $0 \le i \le v - 1$ so that we can use Equation 1 in Theorem 4. The requirements are (i) $t_i - t_{i-1} > \delta$ for $1 \le i \le m - 1$ and (ii) $t_v = 1$. This means that we need (i) $m \ge n\delta$ and (ii) $m(v-1) < n \le mv$. Hence, we set

$$m = \lceil n\delta \rceil, \quad v = \left\lceil \frac{n}{m} \right\rceil.$$

Write $m_i = m \times i$ for $0 \le i \le v - 1$ and $m_v = n$.

$$\begin{aligned}
\mathrm{P}(w(X^n; \delta) \ge 3\epsilon) &\le \sum_{i=1}^{v} \mathrm{P}\left(\sup_{t_{i-1} \le s \le t_i} |X^n(s) - X^n(t_{i-1})| \ge \epsilon \right) \\
&\le \sum_{i=1}^{v} \mathrm{P}\left(\max_{m_{i-1} \le k \le m_i} \frac{|S_k - S_{m_{i-1}}|}{\sigma\sqrt{n}} \ge \epsilon \right) \\
&= v \, \mathrm{P}\left(\max_{k \le m} |S_k| \ge \epsilon\sigma\sqrt{n} \right).
\end{aligned}$$

The first line holds by Theorem 4. The second line holds by the construction of X^n: for $s \in [0, 1]$ that is not an endpoint, $X^n(s)$ is defined to be the convex combination of the endpoints nearby. The third line holds from the fact that $X_1, ..., X_n$ are i.i.d. For large n,

$$v \longrightarrow \frac{1}{\delta} < \frac{2}{\delta}, \quad \frac{n}{m} \longrightarrow \frac{1}{\delta} > \frac{1}{2\delta}.$$

For any $\epsilon > 0$, write

$$\lambda = \frac{\epsilon}{\sqrt{2\delta}}.$$

Then we have

$$\begin{aligned} v\,\mathrm{P}\left(\max_{k\le m}|S_k| \ge \epsilon\sigma\sqrt{n}\right) &\le \frac{2}{\delta}\,\mathrm{P}\left(\max_{k\le m}|S_k| \ge \frac{\epsilon}{\sqrt{2\delta}}\sigma\sqrt{m}\right) \\ &= \frac{4\lambda^2}{\epsilon^2}\,\mathrm{P}\left(\max_{k\le m}|S_k| \ge \lambda\sigma\sqrt{m}\right). \end{aligned}$$

Finally,

$$\begin{aligned} \lim_{\delta\to 0}\limsup_{n\to\infty}\mathrm{P}(w(X^n;\delta)\ge\epsilon) &\le \lim_{\delta\to 0}\limsup_{m\to\infty}\frac{4\lambda^2}{\epsilon^2}\,\mathrm{P}\left(\max_{k\le m}|S_k| \ge \lambda\sigma\sqrt{m}\right) \\ &= \frac{4}{\epsilon^2}\lim_{\lambda\to\infty}\limsup_{m\to\infty}\lambda^2\,\mathrm{P}\left(\max_{k\le m}|S_k| \ge \lambda\sigma\sqrt{m}\right) = 0 \end{aligned}$$

where the last equality follows from the assumption. □

By our construction, ξ_n's are i.i.d., and hence the first condition of the lemma holds. Now we can choose any $\{\xi_n\}$ that is convenient to check that the second condition Equation 3 holds. Since ξ_n's are i.i.d., by Etamadi's Inequality, we have

$$\mathrm{P}(\max_{k\le n}|S_k| > \alpha) \le 3\max_{k\le n}\mathrm{P}(|S_k| \ge \alpha/3).$$

Hence, it suffices to choose $\{\xi_n\}$ such that

$$\lim_{\lambda\to\infty}\limsup_{n\to\infty}\lambda^2\max_{k\le n}\mathrm{P}(|S_k|\ge\lambda\sigma\sqrt{n}) = 0. \tag{4}$$

Choose ξ_n's to be $\mathcal{N}(0,1)$. Then

$$\max_{k\le n}\mathrm{P}(|S_k|\ge\lambda\sigma\sqrt{n}) = \mathrm{P}\left(\left|\frac{S_n}{\sqrt{n}}\right| \ge \lambda\right) = 2(1-\Phi(\lambda)).$$

But we know that

$$\lim_{\lambda\to\infty}\lambda^2(1-\Phi(\lambda)) = 0,$$

and therefore Equation 4 holds. We have thus proved the existence of the Wiener measure.

3 Donsker's Theorem

Theorem 6 (Donsker): If $\xi_1, ..., \xi_n$ are i.i.d. with mean 0 and finite variance σ^2, and if Y^n is the random function defined by Equation 2, then $Y^n \Rightarrow W$.

Proof. In the construction of Wiener measure, we have showed that the finite dimensional distribution of Y^n converges to the finite dimensional distribution of W. Hence, it suffices to prove that $\{Y^n\}$ is tight. We have showed that under the additional assumption ξ_n's are normally distributed, then $\{Y^n\}$ is tight. For general $\{\xi_n\}$'s, when giving an upper bound to $\max_{k\leq n} \mathrm{P}(|S_k| \geq \lambda\sigma\sqrt{n})$, we can use the Central Limit Theorem for large k's and Chebyshev's Inequality for small k's. Fix $\lambda > 0$. By Central Limit Theorem, there exists k_λ such that for all $k_\lambda < k < n$,

$$\mathrm{P}(|S_k| \geq \lambda\sigma\sqrt{n}) \leq \mathrm{P}(|S_k| \geq \lambda\sigma\sqrt{k}) = \mathrm{P}\left(\left|\frac{S_k}{\sqrt{k}\sigma}\right| \geq \lambda\right) \leq 3(1-\Phi(\lambda)).$$

For $k < k_\lambda$, by Chebyshev's Inequality,

$$\mathrm{P}(|S_k| \geq \lambda\sigma\sqrt{n}) \leq \frac{\mathrm{Var}(S_k)}{n\sigma^2\lambda^2} \leq \frac{k_\lambda}{n\lambda^2}.$$

Therefore,

$$\limsup_{n\to\infty} \lambda^2 \max_{k\leq n} \mathrm{P}(|S_k| \leq \lambda\sigma\sqrt{n}) \leq \limsup_{n\to\infty}\left(\max\left(\frac{k_\lambda}{n}, 2\lambda^2\Phi(\lambda)\right)\right) = 3\lambda^2(1-\Phi(\lambda)),$$

and since

$$\lim_{\lambda\to\infty} \lambda^2(1-\Phi(\lambda)) = 0,$$

Equation 4 holds. We conclude that $\{Y^n\}$ is tight. □

3.1 An application of the Donsker's Theorem

Let ξ_n's be i.i.d. and write $S_n = \xi_1 + ... + \xi_n$ and $S_0 = 0$. In this section, we use Donsker's Theorem to find the limiting distribution of

$$M_n = \max_{0\leq i\leq n} S_n.$$

Let $\{Y^n\}$ be defined by Equation 2. Then clearly,

$$\sup_{t\in[0,1]} Y^n(t) = \frac{M_n}{\sigma\sqrt{n}}.$$

Therefore, by Donsker's Theorem and the Continuous Mapping Theorem (because the function $\pi^* : C \to \mathbb{R}$ defined by $\pi^*(x) = \sup_{t\in[0,1]} x(t)$ is a continuous function on C),

$$\frac{M_n}{\sigma\sqrt{n}} \Rightarrow \sup_{t\in[0,1]} W(t).$$

Therefore, we have to find the distribution of $\sup_{t\in[0,1]} W(t)$.

Our strategy is to find a sequence of i.i.d. ξ_i's that we know the asymptotic distribution of $M_n/\sigma\sqrt{n} = \sup_{t\in[0,1]} Y^n(t)$. Then by Donsker's Theorem and Continuous Mapping Theorem, this asymptotic distribution is exactly the distribution of $\sup_{t\in[0,1]} W(t)$. We choose ξ's to take values with 1 and -1 with probability $1/2$ for both. Setting $S_0 = 0$. Therefore, S_n is the position of a symmetric random walk after the nth step with initial position 0. Note that when $a < 0$, since $S_0 = 0$, $\mathrm{P}(M_n \geq a) = 1$. We now prove that for any $a \geq 0$,

$$\mathrm{P}(M_n \geq a) = 2\,\mathrm{P}(S_n > a) + \mathrm{P}(S_n = a). \tag{5}$$

When $a = 0$, $\mathrm{P}(M_n \geq a) = 1$. Also, $\mathrm{P}(S_n > 0) = \mathrm{P}(S_n < 0)$. Therefore, we have

$$\mathrm{P}(M_n \geq 0) = 1 = \mathrm{P}(S_n > 0) + \mathrm{P}(S_n < 0) + \mathrm{P}(S_n = 0) = 2\,\mathrm{P}(S_n > 0) + \mathrm{P}(S_n = 0).$$

Now assume $a > 0$.

$$\mathrm{P}(M_n \geq a) - \mathrm{P}(S_n = a) = \mathrm{P}(M_n \geq a, S_n > a) + \mathrm{P}(M_n \geq a, S_n < a).$$

But $\mathrm{P}(M_n \geq a, S_n > a) = \mathrm{P}(S_n > a)$, so it suffices to prove that

$$\mathrm{P}(M_n \geq a, S_n < a) = \mathrm{P}(M_n \geq a, S_n > a).$$

There are 2^n possible paths each with the same probability. If we can prove that the number of paths in the event $\{M_n \geq a, S_n < a\}$ and $\{M_n \geq a, S_n > a\}$ are the same, then the equality holds. Given a path in $\{M_n \geq a, S_n > a\}$, match it with the path obtained by reflecting through a all the partial sums after the one that first achieves a. This describes a one-to-one correspondence between paths in the two events. In Figure 1, the solid path in the event $\{M_n \geq a, S_n > a\}$ is matched with the dashed path in $\{M_n \geq a, S_n < a\}$.

Now take $a = \sqrt{n}\sigma\alpha = \sqrt{n}\alpha$ ($\mathrm{Var}(\xi_i) = 1$) where $\alpha \geq 0$. In Equation 5, the right-most term converges to 0. And so by CLT,

$$\mathrm{P}(M_n \geq \sqrt{n}\alpha) = 2\,\mathrm{P}(S_n > \sqrt{n}\alpha) \longrightarrow 2\,\mathrm{P}(N > \alpha)$$

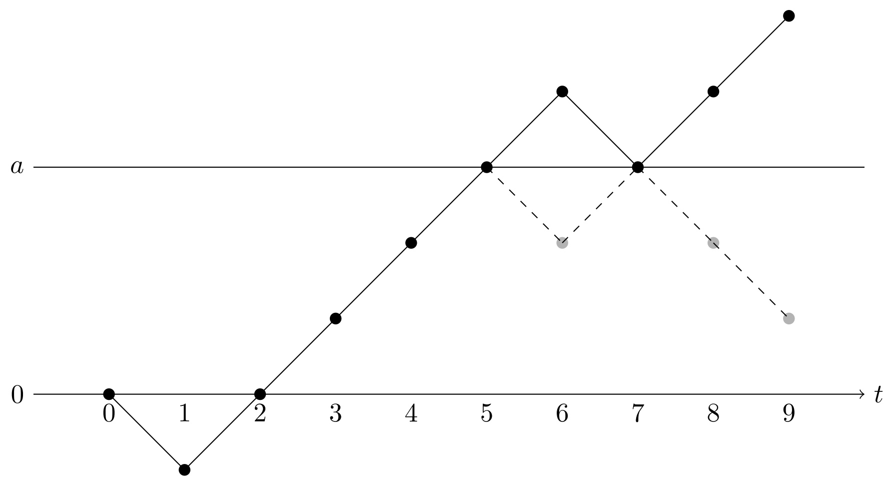

Figure 1: Reflection Principle

where N is standard normal. Finally, we can conclude that for each $\alpha \geq 0$,

$$P\left(\sup_{t\in[0,1]} W_t \leq \alpha\right) = P(-\alpha < N < \alpha) = 2\,P(0 < N < \alpha).$$

And when $\alpha < 0$,

$$P\left(\sup_{t\in[0,1]} W_t \leq \alpha\right) = 0,$$

since by definition, $W_0 \equiv 0$.

Theorem 7: Let $\xi_1, ..., \xi_k$ be i.i.d. with mean 0 and finite variance σ^2. Let $S_0 = 0$, $S_k = \xi_1 + ... + \xi_k$ for $k \geq 1$. Define

$$M_n = \sup_{0\leq k\leq n} S_k.$$

Then for all $t \geq 0$,

$$P\left(\frac{M_n}{\sqrt{n}\sigma} \leq \alpha\right) \longrightarrow P\left(\sup_{t\in[0,1]} W_t \leq \alpha\right) = 2\,P(0 < N < \alpha),$$

where N is the standard normal. That is, the asymptotic distribution of $M_n/\sqrt{n}\sigma$ follows the folded normal law.