7 Randomness and determinism

Recall the definition of the random walk: it is a discrete-time stochastic signal \( X = (X_t)_{t \in \mathbb{Z}_+} \) with the deterministic initial condition \( X_0 = 0 \) and the update rule

\[
X_{t+1} = X_t + U_t, \quad t = 0, 1, 2, \ldots
\]

where \( U = (U_t)_{t \in \mathbb{Z}} \) is an i.i.d. stochastic signal. Let \( \mu = \mathbb{E}[U_0] \) and \( \sigma^2 = \text{Var}[U_0] \). Then

\[
\mathbb{E}[X_t] = \mathbb{E}[U_0 + \ldots + U_{t-1}] = t\mu
\]

and

\[
\text{Var}[X_t] = \text{Var}[U_0 + \ldots + U_{t-1}] = \sum_{s=0}^{t-1} \text{Var}[U_s] = t\sigma^2,
\]

where we have used the fact that \( U_0, \ldots, U_{t-1} \) are independent. Thus, both the mean and the variance of \( X_t \) grow linearly with \( t \), which means that, if \( \mu \neq 0 \), then the random walk will drift farther and farther away from the origin as time goes by: at time \( t \), with high probability it will be somewhere in the interval \([t\mu - \sqrt{t}\sigma, t\mu + \sqrt{t}\sigma]\). If the increments of the walk are zero-mean, then the walk will stay near the origin on average, but will take longer and longer excursions as \( t \) increases. Therefore, we are justified in saying that the amount of randomness in \( X_t \) increases with \( t \)—as \( t \to \infty \), the value of \( X_t \) will become less and less predictable.

On the other hand, consider the average displacement of the random walk at time \( t \):

\[
\bar{X}_t \triangleq \frac{X_t}{t} = \frac{U_0 + \ldots + U_{t-1}}{t}.
\]

Then \( \mathbb{E}[\bar{X}_t] = \frac{1}{t} \mathbb{E}[X_t] = \mu \) and \( \text{Var}[\bar{X}_t] = \frac{1}{t^2} \text{Var}[X_t] = \frac{\sigma^2}{t} \). We notice two things:

1. The expected average displacement is constant and equal to \( \mu \), which makes sense: \( \mu \) is the average displacement per time step.

2. The variance of the expected average displacement decays as \( \frac{1}{t} \).

Therefore, if we observe the random walk for a long enough time, then we will see that it will tend to spend a great deal of time around the point \( x = \mu \). In fact, as we will make precise later, as \( t \to \infty \), the average displacement \( \bar{X}_t \) will be in the interval \([\mu - \frac{\sigma}{\sqrt{t}}, \mu + \frac{\sigma}{\sqrt{t}}]\) with overwhelming probability. This suggests that the operation of averaging has the effect of reducing the fluctuations around the mean, and in fact drives them to zero as \( t \to \infty \). It is convenient to subtract off the mean \( \mu \) and to focus on the random variables

\[
S_t \triangleq \bar{X}_t - \mu = \frac{U_0 + \ldots + U_{t-1}}{t} - \mu.
\]

Then \( \mathbb{E}[S_t] = 0 \) and \( \text{Var}[S_t] = \frac{\sigma^2}{t} \). Since any random variable with zero variance is deterministic and equal to its mean, we see that, in the limit as \( t \to \infty \), the random variables \( S_t \) will become deterministic: \( S_t \to 0 \). This is the Law of Large Numbers (LLN).
Now, if we want to take a better look at the fluctuations of $S_t$ around zero, we should scale $S_t$ in such a way that the variance of the scaled quantity remains constant as $t$ increases. Since $S_t = \frac{X_t - t\mu}{t}$, we see that multiplying $S_t$ by $\sqrt{t}$ will have the desired effect:

$$\text{Var}[\sqrt{t}S_t] = t\text{Var}[S_t] = t \cdot \frac{\sigma^2}{t} = \sigma^2.$$ 

With this in mind, let us define

$$Z_t \triangleq \frac{\sqrt{t}S_t}{\sigma} = \frac{U_0 + \ldots + U_{t-1} - t\mu}{\sqrt{t}\sigma^2}.$$ 

Then $E[Z_t] = 0$ and $\text{Var}[Z_t] = \frac{1}{\sigma^2}\text{Var}[S_t] = 1$. This scaling helps us “zoom in” on the fluctuations of $S_t$ around 0. As we will see shortly, at this scale the fluctuations are Gaussian, with zero mean and unit variance. That is, as $t \to \infty$, the distribution of $Z_t$ will approach $N(0, 1)$. This is the Central Limit Theorem (CLT).

### 7.1 The LLN, the CLT, and stability of linear systems

Before proving the LLN and the CLT, it is instructive to look at these results through the lens of linear systems. We have two stochastic signals, $S = (S_t)_{t \in \mathbb{N}}$ and $Z = (Z_t)_{t \in \mathbb{N}}$, that are given by linear transformations of the i.i.d. stochastic signal $U$. In particular,

$$S_{t+1} = \frac{U_0 + \ldots + U_t}{t+1} - \mu$$

$$= \frac{U_0 + \ldots + U_{t-1} + U_t}{t+1} - \mu$$

$$= \frac{t(S_t + \mu)}{t+1} + \frac{U_t}{t+1} - \mu$$

$$= \frac{t}{t+1} S_t + \frac{U_t - \mu}{t+1}.$$ 

Introducing the centered version of $U_t$, $V_t \triangleq U_t - \mu$, we can write so we can write

$$S_{t+1} = f_t(S_t, V_t), \quad \text{where } f_t(s, v) \triangleq \frac{t}{t+1}s + \frac{v}{t+1}. \tag{7.1}$$

Similarly,

$$Z_{t+1} = \frac{\sqrt{t+1}}{\sigma} S_{t+1}$$

$$= \frac{\sqrt{t+1}}{\sigma} \left( \frac{t}{t+1} S_t + \frac{U_t - \mu}{t+1} \right)$$

$$= \frac{\sqrt{t+1}}{\sigma} \left( \frac{t}{t+1} \frac{\sigma Z_t}{\sqrt{t}} + \frac{U_t - \mu}{\sqrt{t}} \right)$$

$$= \frac{t}{t+1} Z_t + \frac{U_t - \mu}{\sigma \sqrt{t+1}},$$
which allows us to write
\[ Z_{t+1} = g_t(Z_t, V_t), \]
where \( g_t(z, u) \triangleq \sqrt{\frac{t}{t+1} z + \frac{v}{\sigma \sqrt{t+1}}}. \) (7.2)

Noting that \( V_t \) is independent of \( S_t \) and \( Z_t \), we see from (7.1) and (7.2) that both \( S \) and \( Z \) are discrete-time, continuous-state, time-inhomogeneous Markov chains, that \( S_{t+1} \) is a linear function of \( S_t \) and \( V_t \), and that \( Z_{t+1} \) is a linear function of \( Z_t \) and \( V_t \). Moreover, note that \( f_t(0,0) = g_t(0,0) = 0 \), so we can think about \( s = 0 \) and \( z = 0 \) as “equilibrium points” of these two time-varying linear systems. We can now restate the LLN and the CLT as follows:

1. As \( t \to \infty \), \( S_t \) will converge to the equilibrium point \( s = 0 \).
2. As \( t \to \infty \), the distribution of the fluctuations of \( Z_t \) around the equilibrium point \( z = 0 \) will approach that of a zero-mean, unit-variance Gaussian random variable.

### 7.2 Proving the LLN and the CLT

We now sketch the proofs of the LLN and the CLT by characterizing the limiting behavior of the probability distributions of \( S_t \) and \( Z_t \). To that end, we will look at their characteristic functions. Our goal is to show that

\[ \lim_{t \to \infty} \Phi_{S_t}(u) = 1 \] (7.3)

and

\[ \lim_{t \to \infty} \Phi_{Z_t}(u) = e^{-u^2/2}. \] (7.4)

That is, as \( t \to \infty \), the characteristic functions of \( S_t \) will converge to the characteristic function of the deterministic random variable taking the value 0, while those of \( Z_t \) will converge to the characteristic function of a \( N(0,1) \) random variable. Since the distribution of a random variable is uniquely determined by its characteristic function, Eq. (7.3) gives the Law of Large Numbers, while Eq. (7.4) gives the Central Limit Theorem.

We will first express everything in terms of the characteristic function \( \Phi(u) = E[e^{iuU_0}] \) of \( U_0 \). Using the fact that the \( U_t \)’s are i.i.d., we have

\[
\Phi_{S_t}(u) = E[e^{iuS_t}] \\
= E \left[ \exp \left( iu \left( \frac{U_0 + \ldots + U_{t-1}}{t} - \mu \right) \right) \right] \\
= e^{-iu\mu} E \left[ \exp \left( \frac{iu}{t} U_0 + \ldots + \frac{iu}{t} U_{t-1} \right) \right] \\
= e^{-iu\mu} E[e^{i(u/t)U_0} \ldots e^{i(u/t)U_{t-1}}] \\
= e^{-iu\mu} E[e^{i(u/t)U_0}] \ldots E[e^{i(u/t)U_{t-1}}] \\
= e^{-iu\mu} \left( \Phi \left( \frac{u}{t} \right) \right)^t.
\] (7.5)
and

\[
\Phi_{Z_t}(u) = E[e^{iu\sqrt{tS_t}/\sigma}]
\]
\[
= \Phi_{S_t} \left( \frac{u \sqrt{t}}{\sigma} \right)
\]
\[
= e^{-iu\sqrt{t}/\sigma} \left( \Phi \left( \frac{u}{\sqrt{t}\sigma} \right) \right)^t.
\] (7.6)

Now we will investigate the limits as \( t \to \infty \). To that end, we will use first- and second-order Taylor approximations of the characteristic function \( \Phi(u) \) around \( u = 0 \). Recall that the first-order Taylor approximation of a differentiable function \( f \) around the point \( u = 0 \) is given by

\[
f(u) = f(0) + f'(0)u + R_1(u),
\]

where the remainder term \( R_1(u) \) has the property that \( \lim_{u \to 0} \frac{R_1(u)}{u} = 0 \). Similarly, the second-order Taylor approximation of a twice-differentiable function \( f \) around \( u = 0 \) is given by

\[
f(u) = f(0) + f'(0)u + \frac{1}{2}f''(0)u^2 + R_2(u),
\]

where the remainder \( R_2(u) \) is such that \( \lim_{u \to 0} \frac{R_2(u)}{u^2} = 0 \). In the special case of \( f \) being the characteristic function of some random variable \( Z \), i.e., \( f(u) = E[e^{iuZ}] \), we have

\[
f(0) = E[e^{iuZ}] \bigg|_{u=0} = 1,
\]

and the first two derivatives at \( u = 0 \) are

\[
f'(0) = \frac{d}{du} E[e^{iuZ}] \bigg|_{u=0} = iE[Ze^{iuZ}] \bigg|_{u=0} = iE[Z]
\]

and

\[
f''(0) = \frac{d}{du} E[Ze^{iuZ}] \bigg|_{u=0} = -E[Z^2] = E[Z]^2 - \text{Var}[Z].
\]

Now let us examine the term involving \( \Phi \) in Eq. (7.5). Using the first-order Taylor approximation of \( \Phi(u/t) \) around 0, we have

\[
\left( \Phi \left( \frac{u}{t} \right) \right)^t = \left( \Phi(0) + \Phi'(0) \frac{u}{t} + R_1 \left( \frac{u}{t} \right) \right)^t
\]
\[
= \left( 1 + \frac{iu\mu}{t} + R_1 \left( \frac{u}{t} \right) \right)^t
\]
\[
= \left( 1 + \frac{1}{t} (iu\mu + \xi_t) \right)^t,
\]
where we have defined $\xi_t \triangleq tR_1(u/t)$. Since $R_1$ is the remainder term in the first-order Taylor approximation, we have $\lim_{t \to \infty} \xi_t = 0$ (recall that $u$ is fixed). Now we will use the following result: If $(a_t)_{t \in \mathbb{N}}$ is any sequence of complex numbers, such that the limit $a = \lim_{t \to \infty} a_t$ exists, then

$$\lim_{t \to \infty} \left( 1 + \frac{a_t}{t} \right)^t = e^a. \quad (7.7)$$

Applying (7.7) to the sequence $a_t = iu\mu + \xi_t$, we get

$$\lim_{t \to \infty} \left( 1 + \frac{1}{t} (iu\mu + \xi_t) \right)^t = e^{iu\mu},$$

and therefore

$$\lim_{t \to \infty} \Phi_{S_t}(u) = e^{-iu\mu} \lim_{t \to \infty} \left( \Phi \left( \frac{u}{t} \right) \right)^t = e^{-iu\mu} e^{iu\mu} = 1.$$

Next, we turn to (7.6). Writing

$$\Phi(u) = \mathbf{E}[e^{iuU_0}] = \mathbf{E}[e^{iu(V_0 + \mu)}] = e^{iu\mu} \Psi(u),$$

where $\Psi(u) \triangleq \mathbf{E}[e^{iuV_0}]$ is the characteristic function of $V_0 = U_0 - \mu$, we can express the right-hand side of (7.6) as

$$e^{-iu\sqrt{t}/\sigma} \left( \Phi \left( \frac{u}{\sqrt{t} \sigma} \right) \right)^t = e^{-iu\sqrt{t}/\sigma} e^{iu/\sqrt{t} \sigma} \Psi \left( \frac{u}{\sqrt{t} \sigma} \right)^t = \left( \Psi \left( \frac{u}{\sqrt{t} \sigma} \right) \right)^t.$$

Using the second-order Taylor approximation of $\Psi(u/\sqrt{t} \sigma)$ at 0 in the above expression, we have

$$\left( \frac{u}{\sqrt{t} \sigma} \right)^t = \left( \Psi(0) + \frac{\Psi'(0)u}{\sqrt{t} \sigma} + \frac{\Psi''(0)}{2} \left( \frac{u}{\sqrt{t} \sigma} \right)^2 + R_2 \left( \frac{u}{\sqrt{t} \sigma} \right)^t \right)^t$$

$$= \left( 1 - \frac{\sigma^2}{2} \left( \frac{u}{\sqrt{t} \sigma} \right)^2 + R_2 \left( \frac{u}{\sqrt{t} \sigma} \right)^t \right)^t$$

$$= \left( 1 - \frac{1}{t} \left( \frac{u^2}{2} + \eta_t \right) \right)^t,$$

where we have defined $\eta_t \triangleq tR_2(u/\sqrt{t} \sigma)$. Since $R_2$ is the remainder term in the second-order Taylor approximation, we have $\lim_{t \to \infty} \eta_t = 0$. Therefore, using (7.7) with $a_t = -\left( \frac{u^2}{2} + \eta_t \right)$, we get

$$\lim_{t \to \infty} \left( 1 - \frac{1}{t} \left( \frac{u^2}{2} + \eta_t \right) \right)^t = e^{-u^2/2}.$$

Consequently,

$$\lim_{t \to \infty} \Phi_{Z_t}(u) = \lim_{t \to \infty} e^{-iu\sqrt{t}/\sigma} \left( \Phi \left( \frac{u}{\sqrt{t} \sigma} \right) \right)^t = \lim_{t \to \infty} \left( \Psi \left( \frac{u}{\sqrt{t} \sigma} \right) \right)^t = e^{-u^2/2}.$$
7.3 Variance reduction by averaging

Informally, the Law of Large Numbers says that, if we average a large number of independent random variables \( U_0, U_1, \ldots, U_{t-1} \) with common mean \( \mu \), then the resulting quantity \( \bar{X}_t \triangleq \frac{1}{t}(U_0 + \ldots + U_{t-1}) \) will be nearly constant (and equal to \( \mu \)). Moreover, the Central Limit Theorem says that, provided all the \( U_t \)'s have the same finite variance \( \sigma^2 \), then, for all sufficiently large \( t \), the rescaled average \( \sqrt{t} \cdot \bar{X}_t \) will resemble a \( N(\mu, \sigma^2) \) random variable. These two fundamental results of probability theory have many important consequences, and we will discuss them in what follows.

7.3.1 The Monte Carlo method

As we have discussed earlier, there are many cases where randomness can be beneficial. One such instance is the problem of numerical integration. Suppose that we wish to compute the definite integral

\[
I = \int_a^b g(w) \, dw,
\]

where \( g \) is some function of interest, and where \( -\infty \leq a < b \leq +\infty \). We assume that \( g(w) \) is easy to evaluate for any given \( w \in [a,b] \), but computing the integral in closed form is not possible. An ingenious idea, which has its origins in the Manhattan Project during World War II, is as follows: Pick a well-behaved pdf \( f \) supported on the interval \([a,b]\), i.e., \( f(w) > 0 \) for \( w \in [a,b] \) and \( f(w) = 0 \) for \( w \notin [a,b] \) and write

\[
I = \int_a^b f(w) \frac{g(w)}{f(w)} \, dw.
\]  

(7.8)

For example, if \([a,b]\) is a finite interval, then we can take \( f \) to be the pdf of a \( U(a,b) \) random variable, in which case \( f(w) = \frac{1}{b-a} \) for \( w \in [a,b] \) and 0 otherwise; if \([a,b] = \mathbb{R}\), we can take the Gaussian pdf with mean 0 and variance 1. Now, if we define the function \( h(w) \triangleq \frac{g(w)}{f(w)} \) for all \( w \in [a,b] \), then Eq. (7.8) shows that the value \( I \) of the integral is equal to the expectation \( E[h(W)] \) with \( W \sim f \):

\[
I = E[h(W)] = E\left[\frac{g(W)}{f(W)}\right], \quad W \sim f.
\]

(7.9)

Now let us make two additional assumptions:

- We can easily generate i.i.d. samples \( W_0, W_1, \ldots \) with pdf \( f \).
- Given any point \( w \in [a,b] \), it is easy to compute the value \( h(w) = g(w)/f(w) \).

Then we can consider the following randomized procedure for approximating \( I \): pick a sufficiently large \( t \), generate random samples \( W_0, W_1, \ldots, W_{t-1} \) i.i.d. \( f \), and compute

\[
\hat{I}_t \triangleq \frac{1}{t} \sum_{s=0}^{t-1} h(W_s) = \frac{1}{t} \sum_{s=0}^{t-1} \frac{g(W_s)}{f(W_s)}.
\]

(7.10)
This is known as the Monte Carlo method\(^1\), and \(\widehat{I}_t\) is referred to as the Monte Carlo estimate of \(I\).

To get an idea of how good of an estimate \(\widehat{I}_t\) is, we will use the LLN and the CLT. First, note that, if we define the random variable \(U \triangleq h(W) = \frac{g(W)}{f(W)}\), then (7.10) can be rewritten as

\[
\widehat{I}_t \triangleq \frac{1}{t} \sum_{s=0}^{t-1} U_s.
\]

Since \(W_0, W_1, \ldots\) are i.i.d., so are \(U_0, U_1, \ldots\), and moreover

\[
E[U_0] = E[h(W_0)] = \int_a^b f(w) \frac{g(w)}{f(w)} \, dw = I.
\]

Therefore, by the LLN, \(\widehat{I}_t\) will converge to \(I\) as \(t \to \infty\). In other words, the more samples from the pdf \(f\) we generate, the better our Monte Carlo approximation will be. On the other hand, \(\widehat{I}_t - I\) is small for any finite \(t\), but we know that

\[
E[(\widehat{I}_t - I)^2] = \text{Var}[\widehat{I}_t] = \frac{1}{t} \text{Var}[U],
\]

so, provided \(\text{Var}[U] = \text{Var}[g(W)/h(W)]\) is small, the absolute error \(|\widehat{I}_t - I|\) will be small on average. With the help of the CLT, we can say even more — when \(t\) is sufficiently large, the probability distribution of the quantity

\[
Z_t = \frac{\widehat{I}_t - I}{\sqrt{t \text{Var}[U]}}
\]

will be approximately normal with zero mean and unit variance. In particular, if we introduce the so-called \(Q\)-function

\[
Q(z) \triangleq \frac{1}{\sqrt{2\pi}} \int_z^\infty e^{-x^2/2} \, dx,
\]

then the CLT says, roughly, that

\[
P[\widehat{I}_t - I \geq a\sqrt{t \text{Var}[U]}] \approx Q(a) \quad \text{and} \quad P[\widehat{I}_t - I \leq -a\sqrt{t \text{Var}[U]}] \approx Q(a)
\]

for any \(a > 0\) and all sufficiently large \(t\). As we will see shortly, \(Q(a) \leq e^{-a^2/2}\), and therefore, for all sufficiently large \(t\) and for all \(a > 0\),

\[
P[|\widehat{I}_t - I| \geq a\sqrt{t \text{Var}[U]}] = P[\{\widehat{I}_t - I \geq a\sqrt{t \text{Var}[U]}\} \cup \{\widehat{I}_t - I \leq -a\sqrt{t \text{Var}[U]}\}] \\
\leq P[\widehat{I}_t - I \geq a\sqrt{t \text{Var}[U]}] + P[\widehat{I}_t - I \leq -a\sqrt{t \text{Var}[U]}] \\
\approx 2e^{-a^2/2}.
\]

Of course, the variance of \(U\) is determined by the function \(g\) and on the pdf \(f\), and it may not be easy to compute it exactly.

---

\(^1\)The name “Monte Carlo,” which is a reference to the famous Monte Carlo Casino in Monaco, was used as a code name at the Los Alamos Laboratory during World War II.