6. Central Limit Theorems

We will use characteristic functions to find conditions under which the rescaled sum
\[ \frac{X_1 + \cdots + X_n - n \bar{E} X_i}{\sqrt{n \text{Var}(X_i)}} \]
converges in distribution to a standard normal random variable as \( n \to \infty \).

The proofs will rely on some facts about the complex exponential and logarithmic functions. Once those are established, the central limit theorems follow relatively easily.

Introduction

Let \( X_i, X_j, \ldots \) be independent and identically distributed.

In the Strong Law of Large Numbers, we assumed that the \( X_i \) were integrable and showed that the sample average \( \frac{X_1 + \cdots + X_n}{n} \) converges almost surely to \( \bar{E} X_i \).

Now we want to learn more about the distribution of \( X_1 + \cdots + X_n \) as \( n \to \infty \).

Assuming the \( X_i \) are integrable, \( X_1 + \cdots + X_n \) has mean \( n \bar{E} X_i \).

Assuming the \( X_i \) are square integrable, \( X_1 + \cdots + X_n \) has variance \( n \text{Var}(X_i) \).

This being the case, it seems that the distributions of \( X_1 + \cdots + X_n \) will not be tight, and so \( X_1 + \cdots + X_n \) will not converge in distribution as \( n \to \infty \).

We can adjust for this by centering and scaling \( X_1 + \cdots + X_n \).

Let
\[ Z_n = \frac{X_1 + \cdots + X_n - n \bar{E} X_i}{\sqrt{n \text{Var}(X_i)}} \]

It is easy to verify that \( Z_n \) has mean 0 and variance 1 for all \( n \), so Chebyshev's inequality tells us that \( \text{Pr}(|Z_n| > M) \leq \frac{1}{M^2} \), from which we see that \( (Z_n) \) is tight.

We calculate the characteristic function of \( Z_n \) this way:

\[
Y_n(t) = E e^{itZ_n} = E e^{it \frac{X_1 + \cdots + X_n - n \bar{E} X_i}{\sqrt{n \text{Var}(X_i)}}} \\
= E e^{i \frac{t}{\sqrt{n \text{Var}(X_i)}} (X_1 - \bar{E} X_i) \cdots (X_n - \bar{E} X_i)} = \varphi \left( \frac{t}{\sqrt{n \text{Var}(X_i)}} \right)^n,
\]

where \( \varphi \) is the characteristic function of \( X_i - \bar{E} X_i \).
The limiting behavior of \( q_n(t) \) thus comes down to the behavior of the characteristic function \( q \) of \( X, -EX \), near 0. We will see that, because \( X, -EX \), has mean 0 and finite variance, \( q(t) \) is roughly parabolic for \( t \) near 0:

\[
q(t) \approx 1 - ct^2 \quad \text{for } t \text{ near 0.}
\]

We will calculate the constant \( c \) and evaluate \( q \left( \frac{t}{\sqrt{n \cdot \text{Var}(X_i)}} \right)^n \) as \( n \to \infty \) with the help of some results concerning the complex exponential function \( z \to e^z \).

Finally, once we understand the distribution of \( Z_n \), we will be able to do computations concerning the distribution of \( X, + \cdots + X_n \), because \( X, + \cdots + X_n \) is a simple affine function of \( Z_n \):

\[
X, + \cdots + X_n = n \cdot EX_1 + \sqrt{n \cdot \text{Var}(X_1)} \cdot Z_n.
\]
Facts about the complex exponential function

For each \( z \) in \( \mathbb{C} \), \( \exp(z) = e^z \) is defined by \( \exp(z) = e^z = \sum_{k=0}^{\infty} \frac{z^k}{k!} \).

The series converges absolutely.

We would like to approximate \( e^z \) by a polynomial in \( z \) of degree \( n \).

Define the remainder function \( R_n \) by \( R_n(0) = 0 \) and for \( z \neq 0 \) by

\[
e^z = \sum_{k=0}^{n} \frac{z^k}{k!} + R_n(z) \cdot |z|^n
\]

First we show a bound on \( R_n(z) \) for \( z \) near \( 0 \):

\[
|R_n(z) \cdot |z|^n| = \left| \sum_{k=n+1}^{\infty} \frac{|z|^k}{k!} \right| \leq \sum_{k=n+1}^{\infty} \frac{|z|^k}{k!} \leq \sum_{k=0}^{\infty} \frac{|z|^{n+k}}{(n+k)!} \leq |z|^{n+1} \sum_{k=0}^{\infty} \frac{|z|^k}{k!} = |z|^{n+1} e^{|z|}
\]

Thus, \( |R_n(z)| \leq |z| e^{|z|} \) for all \( z \in \mathbb{C} \).

Clearly, as \( z \to 0 \), \( R_n(z) \to 0 \) as well. This is essentially what Taylor's Theorem with remainder term does for you, but here we have shown it directly. Note that \( R_n \) is continuous at \( 0 \).

Now we consider \( R_n(ib) \) where \( b \in \mathbb{R} \) as \( b \to \pm \infty \).

\[
R_n(ib) \cdot |ib|^n = e^{ib} - \sum_{k=0}^{n} \frac{(ib)^k}{k!}
\]

\[
|R_n(ib)| \cdot |ib|^n \leq |e^{ib}| + \sum_{k=0}^{n} \left| \frac{(ib)^k}{k!} \right| \leq 1 + 1 + |b| + \frac{|b|^2}{2} + \ldots + \frac{|b|^n}{n!}
\]

As \( b \to \pm \infty \), the right-hand side clearly converges to \( \frac{1}{n!} \).

Thus, the function \( b \mapsto R_n(ib) \) is continuous (from its definition), equals \( 0 \) at \( b=0 \), and is bounded by some number \( R_n \) (since \( b \mapsto R_n(ib) \) is bounded on each finite interval and since, as \( b \to \pm \infty \), it is bounded by a function with a finite limit.)
Theorem 2.4.1 Central Limit Theorem

Let \( X_1, X_2, \ldots \) be independent, identically distributed real valued random variables with finite mean \( \mu \) and finite, non-zero variance \( \sigma^2 \).

Then the sequence

\[
Z_n = \frac{X_1 + \cdots + X_n - n \cdot \mu}{\sigma \sqrt{n}}, \quad n = 1, 2, \ldots
\]

converges in distribution to a standard normal random variable \( Z \).

Remark

Note that \( E(Z_n) = 0 \) and \( Var(Z_n) = Var(\frac{X_1 + \cdots + X_n}{\sigma \sqrt{n}}) \)

\[= \frac{1}{\sigma^2 n} \cdot n \cdot Var(X_1)\]

\[= 1,\]

the same mean and variance as the limiting random variable \( Z \).

The result says that, in addition to the mean and the variance, the rest of the distribution of \( Z_n \) also converges, to normal.

Proof

Let \( Y_n = X_n - n \cdot \mu \cdot \mu \). Then \( Y_1, Y_2, \ldots \) are i.i.d. with mean zero and variance \( \sigma^2 \), and \( Z_n = \frac{1}{\sigma \sqrt{n}} (Y_1 + \cdots + Y_n) \).

Let \( \phi \) be the characteristic function of \( Y_1 \), and use the previous result on the approximation of \( e^{it} \) using \( n = 2 \):

\[
\phi(t) = E(e^{itY_1}) = 1 + \frac{(itY_1)^1}{1!} + \frac{(itY_1)^2}{2!} + R_2(itY_1) \cdot |itY_1|^2
\]

\[= E(1 + itEY_1 - \frac{t^2}{2} EY_1^2 + t^2 E R_2(itY_1) Y_1^2),\]

by linearity and integrability, the last term being integrable because \( |R_2| \) is bounded by \( \sigma \) and \( \sigma \cdot \sigma \cdot \sigma \cdot \sigma = \sigma^2 \leq \infty \). But \( \sigma = 0 \), so

\[
\phi(t) = 1 - \frac{t^2}{2} \cdot \frac{\sigma^2}{\sigma^2} + t^2 \cdot E R_2(itY_1) Y_1^2.
\]

This is essentially a Taylor series expansion of \( \phi(t) \) to second degree.
Let \( \Psi_n \) be the characteristic function of \( Z_n \) and let \( t \in \mathbb{R} \). Then,

\[
\Psi_n(t) = \mathbb{E} \exp \left( i \frac{Y_1 + \ldots + Y_n}{\sqrt{n}} t \right) = \left[ \mathbb{E} \exp \left( i Y_i \frac{t}{\sqrt{n}} \right) \right]^n
\]

since \( Y_1, \ldots, Y_n \) are iid.

\[
= \varphi \left( \frac{t}{\sqrt{n}} \right)^n = \mathbb{E} \exp \left( i Y_i \frac{t}{\sqrt{n}} \right)^n
\]

by definition of \( \varphi \).

\[
= \left[ 1 - \frac{t^2}{\sigma^2 n} + \frac{t^2}{\sigma^2 n} \cdot \mathbb{E} R_2 \left( i Y_i \frac{t}{\sqrt{n}} \right) Y_i^2 \right]^n
\]

where \( C_n = -\frac{t^2}{2} + \frac{t^2}{\sigma^2} \cdot \mathbb{E} R_2 \left( i Y_i \frac{t}{\sqrt{n}} \right) Y_i^2 \).

Note that, as \( n \to \infty \), the argument \( i Y_i \frac{t}{\sqrt{n}} \) of \( R_2 \) goes to 0.

Now \( |R_2(i b)| \) is bounded by \( r_2 \), and \( r_2 Y_i^2 \) is integrable since \( \sigma^2 < \infty \). Thus, by the Dominated Convergence Theorem,

\[
\lim_{n \to \infty} C_n = -\frac{t^2}{2}.
\]

We will show below that, then,

\[
\Psi_n(t) = \left[ 1 + \frac{C_n}{n} \right]^n \to e^{-\frac{t^2}{2}} \text{ as } n \to \infty.
\]

The limiting function \( e^{-\frac{t^2}{2}} \) is continuous at 0, so by the Continuity Theorem, \( Z_n \) converges in distribution to a random variable having characteristic function \( e^{-\frac{t^2}{2}} \).

By Exercise 21 and the Uniqueness Theorem, the limiting random variable has the standard normal distribution \( \mathcal{N}(0,1) \), which has density \( \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \) with respect to the Lebesgue measure on \( \mathbb{R} \). \( \square \)
Application of the Central Limit Theorem

Example

You are a geologist.
Your job is to measure the distance between mountain peaks.
These distances can change over time due to volcanic activity and the movement of tectonic plates.
The distances must be measured carefully.

Call the distance between peak A and peak B $\mu$. You don’t know $\mu$.
The measurements you make are not all the same.
They are influenced by variations in temperatures between the peaks, wind hitting your instruments, small electrical variations in your instruments, etc.
However, you have used these instruments enough times that you do know the amount of variability to expect.
In order to have reasonably small numbers to work with, let’s say that the standard deviation of the measurements is $\sigma = 1$ foot.

You are going to make 100 measurements of the distance and compute their average, call it $A$.
What should you report for the value of $\mu$?
Can you say, “I believe $\mu$ is in the interval [(A-\varepsilon, A+\varepsilon)]”?
What would you use for $\varepsilon$?

We can model the situation this way.
Let $X_1, X_2, \ldots$ be the successive measurements you will make.
Assuming they are made under roughly the same conditions (all at the same time of day, for example, all with the instruments properly warmed up), we can assume the $X_i$ are identically distributed.
Assuming that the causes of variability themselves change rapidly over time, we may assume the $X_i$ are independent.

Hopefully the instruments are designed so they do not make a systematic error, by, say, always adding 500 ft. to the distance.
It is reasonable to assume that $EX$, the actual distance, or else just try to figure out what $\mu = EX$, is.
The Central Limit Theorem tells us that, if we make \( n \) measurements,
then \( \frac{X_1 + \ldots + X_n - \mu}{\sigma \sqrt{n}} \) has mean 0, variance 1,
and is approximately normal as \( n \to \infty \). That is, in terms of the distribution function,

\[
\lim_{n \to \infty} P \left( \frac{X_1 + \ldots + X_n - \mu}{\sigma \sqrt{n}} \leq z \right) = \frac{2}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{x^2}{2}} \, dx = \Phi(z)
\]

Let us rewrite this so that it looks like a statement about the sample mean, \( \bar{X}_n \).

\[
\lim_{n \to \infty} P \left( \frac{X_1 + \ldots + X_n - \mu}{\sigma \sqrt{n}} \leq \frac{\bar{X}_n - \mu}{\sigma \sqrt{n}} \right) = \Phi(z)
\]

\[
\lim_{n \to \infty} P \left( \bar{X}_n - \mu \leq \frac{\sigma \bar{X}_n}{\sqrt{n}} \right) = \Phi(z).
\]

Subtracting,

\[
\lim_{n \to \infty} P \left( -\frac{\sigma}{\sqrt{n}} \leq \bar{X}_n - \mu \leq \frac{\sigma}{\sqrt{n}} \right) = \Phi(z) - \Phi(-z)
\]

Even though we do not know the distribution of \( X_1, \ldots, X_n \), for large \( n \), we basically know the distribution of \( \bar{X}_n \).

**Example**

Suppose \( \sigma = 1 \) ft., \( n = 100 \), \( z = 1 \).

Then

\[
P \left( -\frac{1}{10} \leq \bar{X}_{100} - \mu \leq \frac{1}{10} \right) \approx \Phi(1) - \Phi(-1) = 0.6827
\]

So, most of the time, the computed sample mean \( \bar{X}_{100} \) is within \( \frac{1}{10} \) ft. of the actual mean.

Rewriting,

\[
P \left( \mu \in [\bar{X}_{100} - \frac{1}{10}, \bar{X}_{100} + \frac{1}{10}] \right) \approx 0.6827.
\]

Thus, if we say "the distance is \( \bar{X}_{100} \pm \frac{1}{10} \), we will be right most of the time.

Using \( z = 3 \), we get

\[
P \left( \mu \in [\bar{X}_{100} - \frac{3}{10}, \bar{X}_{100} + \frac{3}{10}] \right) \approx \Phi(3) - \Phi(-3) = 0.9973,
\]

so this interval will cover the true value with very high probability, but the interval is wider and so is less informative.

The Central Limit Theorem allows us to quantify the trade-off between confidence (the probability) and precision (the size of the interval).

**Note:** \( P(\mu \in [\bar{X}_{100} - 1, \bar{X}_{100} + 1]) \approx \Phi(10) - \Phi(-10) \approx 1 \).
Probability mass function of original random variable

Error bound $10.2126n^{-0.5}$. Actual error $0.06969n^{-0.50045}$

Probability mass function, $n = 32$

Probability mass function, $n = 256$

Probability mass function, $n = 2048$

Difference of cumulative distribution functions

Difference of cumulative distribution functions

Difference of cumulative distribution functions
Error bound $11.8045n^{-0.5}$. Actual error $0.20413n^{-0.49983}$
Probability mass function of original random variable

Error bound 5.0026n^{−0.5}. Actual error 0.88193n^{−0.49988}

Difference of cumulative distribution functions

Probability mass function, n = 32

Probability mass function, n = 256

Probability mass function, n = 2048
Error bound $3.3906n^{-0.5}$ Actual error $0.20629n^{-0.5}$
Probability mass function of original random variable

Error bound $8.6565n^{-0.5}$. Actual error $2.6004n^{-0.49687}$

Probability mass function, $n = 32$

Probability mass function, $n = 256$

Probability mass function, $n = 2048$

Difference of cumulative distribution functions

Difference of cumulative distribution functions

Difference of cumulative distribution functions
2.4d Rates of convergence (Berry - Esseen)

Let \( X_1, X_2, \ldots \) be iid, \( \mathbb{E} X_i = 0, \sigma^2 = \mathbb{E} X_i^2 < \infty. \)

Let \( Z_n = \frac{X_1 + \ldots + X_n}{\sigma \sqrt{n}} \) be the normalized sum, and let \( Z \sim N(0,1). \)

Then \( Z_n \Rightarrow Z, \) meaning that for all bounded continuous \( g: \mathbb{R} \to \mathbb{R}, \)

\[
\mathbb{E} g(Z_n) \to \mathbb{E} g(Z).
\]

Equivalently, if \( F_n \) is the distribution function of \( Z_n \) and \( \Phi \) is the df of \( Z, \)

\[
F_n(x) \to \Phi(x) \quad \text{as} \quad n \to \infty, \quad \text{for all} \quad x.
\]

Recalling the first discussion of Chapter 2, when \( \sigma = 1, \)

\[
\mathbb{P}(\frac{\sum_{i=1}^{n} X_i}{\sqrt{n}} \in [-0.1, 0.1]) = \mathbb{P}(\frac{\sum_{i=1}^{n} X_i}{\sqrt{n}} \in [-1, 1]) \approx \Phi(1) - \Phi(-1) \approx 0.6827.
\]

Since \( \mathbb{P}(\frac{\sum_{i=1}^{n} X_i}{\sqrt{n}} \in [-1, 1]) \to \Phi(1) - \Phi(-1). \)

However, we don't yet know that \( n=100 \) is large enough!

**Theorem 2.4.9** If \( X_1, X_2, \ldots \) are iid, \( \mathbb{E} X_i = 0, \sigma^2 = \mathbb{E} X_i^2 < \infty, \) and \( \rho = \mathbb{E}|X_i|^3 < \infty, \)

then for all \( x, \) all \( n \geq 2, \ldots, \)

\[
|F_n(x) - \Phi(x)| \leq \frac{3\rho}{\sigma^2 \sqrt{n}}.
\]

In the case of our example, \( \sigma = 1, n = 100, \) so

\[
F_{100}(x) = \Phi(x) \pm \frac{3\rho}{10}.
\]

\[
\mathbb{P}(\frac{\sum_{i=1}^{100} X_i}{\sqrt{100}} \in [-1, 1]) = F_{100}(1) - F_{100}(-1) = 0.6827 \pm \frac{3\rho}{5} \quad \text{(due)}
\]

Is this any use?

**Not really!** \( \rho = \mathbb{E}|X|^3 \) will always be positive \((x^3)^{\frac{1}{3}} \)

\[
\geq (\mathbb{E}X)^{\frac{1}{3}}, \quad \text{Jensen's inequality}
\]

\[
\geq \sigma.
\]

So \( \rho \) will not help! Besides, if we don't know \( F_n, \) how do we know \( \rho? \)

If \( X_i \) is symmetric, it probably won't be so bad...

One can replace \( 3\rho \) by \( 0.8\rho \) with a harder proof, giving

\[
\mathbb{P}(\frac{\sum_{i=1}^{100} X_i}{\sqrt{100}} \in [-1, 1]) = 0.6827 \pm \frac{1.6\rho}{10} = 0.6827 \pm 0.16\rho.
\]

\[
\mathbb{P}(\frac{\sum_{i=1}^{1000} X_i}{\sqrt{1000}} \in [-1, 1]) = 0.6827 \pm \frac{1.6\rho}{100} = 0.6827 \pm 0.016\rho.
\]
Theorem

\[ |F_n(x) - \Phi(x)| \leq \frac{3p}{\sigma^2 \sqrt{n}} \]

You can replace 3 by 0.8.

Proof

Let \( \Psi \) denote the characteristic function of \( X_i \).

If \( X_i \) is discrete, \( \Psi \) will not be integrable (a sum of \( e^{i\alpha} \) terms).

Replace \( Z_n \) by \( Z_n + P_L \), where \( P_L \) has the Polya density, which is very spread out, but which has characteristic function

\[ \omega_L(t) = \max(0, \frac{1}{L} - \frac{|t|}{L}) \]

Then \( Z_n + P_L \) has characteristic function \( \Psi \cdot \omega_L \), which will be integrable.

Thus, the first idea is to "spread out" the possible values of \( Z_n \). Then, you undo this by letting \( L \to \infty \) later in the proof.
Outline of proof

First, we cast the difference $F_n(x) - G_n(x)$ in terms of the corresponding characteristic functions.

**Lemma 2.4.11** Let $F$ and $G$ be distribution functions with the same mean and characteristic functions $\Phi$ and $\Psi$. Suppose $\Phi$ and $\Psi$ are integrable.

Then

$$F(x) - G(x) = \frac{1}{2\pi} \int_\mathbb{R} e^{itx} \frac{\Phi(t) - \Psi(t)}{it} \, dt,$$

for all $x$ in $\mathbb{R}$.

**Proof** By Exercise 17, $F$ and $G$ have bounded continuous densities $f$ and $g$,

$$f(x) = \frac{1}{2\pi} \int_\mathbb{R} e^{-ity} \Phi(t) \, dt, \quad g(x) = \frac{1}{2\pi} \int_\mathbb{R} e^{-ity} \Psi(t) \, dt.$$

Fix $a \in \mathbb{R}$. Subtract $g$ from $f$ and integrate from $a$ to $x$:

$$\int_a^x (f(y) - g(y)) \, dy = \frac{1}{2\pi} \int_\mathbb{R} e^{-ity} (\Phi(t) - \Psi(t)) \, dy \cdot \int_\mathbb{R} e^{itx} \, dt - \frac{1}{it} \left[ e^{itx} - e^{ita} \right].$$

The singularity at $t = 0$ is canceled by $e^{itx} - e^{ita}$ as $t \to 0$.

However, we would like to split this into two integrals.

Because $F$ and $G$ have the same mean, $\Phi(t) = \Psi(t)$, as well as $\Phi(t) = \Psi(t) = 1$, so

$$\frac{\Phi(t) - \Psi(t)}{it} \to 0 \text{ as } t \to 0. \quad \text{(the derivative is zero there!)}$$

Thus,

$$(F(x) - G(x)) - (F(a) - G(a)) = \frac{1}{2\pi} \int_\mathbb{R} (\Phi(t) - \Psi(t)) \, dt \cdot \int_\mathbb{R} e^{itx} \, dt + \frac{1}{2\pi} \int_\mathbb{R} e^{itx} \, dt \cdot \frac{\Phi(t) - \Psi(t)}{it} - e^{ita}.$$

As $a \to -\infty$, the LHS goes to $F(x) - G(x)$.

The RHS is the limit of the inverse transform of $\frac{\Phi(t) - \Psi(t)}{it}$,

an integrable function. $a \to -\infty$ makes $e^{ita}$ oscillate rapidly.

The Riemann-Lebesgue Lemma says this limit is 0.

Exercise 4.4 in the Appendix, p. 462.
Second, we need to deal with two facts:

(i) The characteristic function \( \Phi_n \) of \( Z_n \) may not be integrable.
That would require \( X_i \) to be a continuous random variable,
but \( X_i \) may have discrete components!

(ii) We can only bound \( |\Phi_n(t) - \Phi(t)| \) for \( t \) near 0.
That is where we have a power series expansion.

As it happens, we can deal with both at once!

(i) If we add a continuous random variable \( P_L \) to \( Z_n \), the sum will always be a continuous random variable.
If \( P_L \) takes values near 0, the distribution of \( Z_n \) should be
almost the same as that of \( Z_n + P_L \).

(ii) The characteristic function \( \Phi_n \) of \( Z_n \) will be the product of
\( \Phi_n \) with the characteristic function \( \Phi_L \) of \( P_L \).

The Polya density:
\[
h_L(x) = \frac{1 - \cos(Lx)}{\pi Lx^2}.
\]

Is symmetric, concentrates near 0 as \( L \to \infty \),
but has no mean or variance (Cauchy-like).
Its characteristic function is
\[
\Phi_L(t) = \max(0, 1 - |t|^L).
\]

The characteristic function of \( Z_n \) is then 0 outside of \([-L, L]\),
so it will be integrable.
Lemma 2.4.10, p. 127

Let $F$ and $G$ be distribution functions with $G'(x) < \infty$ for all $x$.
Let $F_L$ and $G_L$ be the distribution functions of the sums with $P_L$.

Then for all $x$,

$$|F(x) - G(x)| \leq \frac{24\lambda}{nL} + 2\sup_{x \in \mathbb{R}} |F_L(x) - G_L(x)|.$$

The point is that the difference between $F$ and $G$ can't be much larger than the difference between $F_L$ and $G_L$. 
In order to use the lemma, fix \( n \), set \( F_n = F_n \) and \( G = \Phi \). Then

\[
\lambda = \lambda_n(\alpha) \approx 0.3989 \sqrt{n} < \frac{2}{5}.
\]

Fix \( L \) and write \( F_n^L, \Phi^L; \psi_n^L, \psi^L \) for the ch. functions

\[
\left| F_n(x) - \Phi(x) \right| \leq \frac{24 \cdot \frac{2}{5}}{\pi \cdot \frac{\sqrt{n} \cdot \pi}{4 \cdot 3 \cdot 5}} + 2 \sup_{x \in \mathbb{R}} \left| F_n(x) - \Phi^L(x) \right|
\]

\[
\leq \frac{36 \rho}{5 \pi \sqrt{n}} + 2 \cdot \frac{1}{2\pi} \sup_{x \in \mathbb{R}} \left| \int_{\mathbb{R}} - e^{-itx} \frac{\psi_n^L(t) - \psi^L(t)}{it} \, dt \right|
\]

\[
= \frac{36 \rho}{5 \pi \sqrt{n}} + \frac{1}{\pi} \sup_{x \in \mathbb{R}} \left| \int_{-L}^{L} - e^{-itx} \frac{\psi_n(t) - \psi(t)}{it} \left( 1 - \frac{L}{L} \right) \, dt \right|
\]

See the following paper for the bound on \( |\psi_n(t) - \psi(t)| \)

\[
\leq \frac{36 \rho}{5 \pi \sqrt{n}} + \frac{1}{\pi} \int_{-L}^{L} \frac{|\psi_n(t) - \psi(t)|}{L} \, dt
\]

\[
= \frac{\rho}{\sqrt{n}} \cdot \frac{1}{\pi} \left[ \frac{36}{5} + \frac{1}{8 \pi} \int_{-L}^{L} e^{4 \pi t^2} \, dt + \frac{1}{8 \pi \sqrt{n}} \int_{-L}^{L} e^{\frac{4 \pi t^2}{\sqrt{n}}} \, dt \right]
\]

\[
\leq \frac{\rho}{\sqrt{n}} \cdot \frac{1}{\pi} \left[ \frac{36}{5} + \frac{1}{8 \pi} \sqrt{4\pi} + \frac{2}{\sqrt{10}} \right]
\]

Assuming \( n \geq 10 \), (for \( n < 9 \) the theorem is trivially true),

\[
\frac{1}{\sqrt{n}} \leq \frac{1}{\sqrt{10}}
\]

Also, \( \rho > 1 \), \( \frac{1}{\rho} \leq 1 \),

\[
\left| F_n(x) - \Phi(x) \right| \leq \frac{\rho}{\sqrt{n}} \cdot \frac{1}{\pi} \left[ \frac{36}{5} + \frac{\sqrt{4\pi}}{3} + \frac{2}{\sqrt{10}} \right] = \frac{\rho}{\sqrt{n}} \cdot 2.05927
\]

\[
< \frac{3\rho}{\sqrt{n}}
\]
Recall \( \Psi_n(t) = \phi(\frac{t}{\tau})^n \)

Set \( \tau \approx \frac{\delta}{\sqrt{n}} \).

We would like to bound \( \Psi_n(t) - \Psi(t) \):

\[
\left| \Psi_n(t) - \Psi(t) \right| = \left| \phi(\tau)^n - e^{-\frac{t^2}{\tau^2}} \right|
\]

\[
\approx \left| \phi(\tau)^n - (e^{-\frac{t^2}{\delta^2}})^n \right| (t^2 \approx \delta^2 t^2)
\]

It's difficult to bound the difference of powers.

Note that, for \( \alpha, \beta \in \mathbb{C} \) with \( |\alpha|, |\beta| \leq \gamma \),

\[
|\alpha^n - \beta^n| = |\alpha - \beta|^n|1 + \alpha^{n-1}\beta + \alpha^{n-2}\beta^2 + \cdots + \beta^{n-1}| \leq \sum_{m=0}^{n-1} |\alpha - \beta|^n|\alpha^{n-m-1}\beta^m| \leq n \gamma^{n-m-1} \gamma^m |\alpha - \beta|
\]

\[
= n |\alpha - \beta| \gamma^{n-1}
\]

Thus,

\[
\left| \Psi_n(t) - \Psi(t) \right| \leq n \left| \phi(\tau) - e^{-\frac{t^2}{\delta^2}} \right| \gamma^{n-1}
\]

where \( \gamma \) is an upper bound on \( |\phi(\tau)| \) and \( e^{-\frac{t^2}{\delta^2}} \).

These functions are quadratic, opening down, near \( t = 0 \).

We can get a bound on them for \( t \) near \( 0 \).
Third, we realize that, although $Y_n$ (the ch. f. of $Z_n$) converges to $Y$ (the ch. f. of $Z$) pointwise, we cannot say that this convergence is \textbf{uniform}, so we cannot directly bound \[ \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{itx} \left( Y_n(t) - Y(t) \right) \, dt. \]

However, we could get a bound on the difference over $t$ in $[-L, L]$.

- For $n \leq q$, $\frac{3P}{n^{5/4}} \geq 1$, so the theorem is trivially true.
- Without loss of generality, we may assume $\sigma = 1$, by considering

\[ Z_n = \frac{X_1 + \cdots + X_n}{\sqrt{n}}, \quad X_1 = \frac{X_1}{\sqrt{n}}, \]

so $X_1 = \frac{X_1}{\sqrt{n}}$. But then $p' = \mathbb{E} |X_1|^3 = \frac{\mathbb{E} |X|^3}{\sigma^3}$, and all will be well.

- Recall that $e^{ia} = 1 + ia + \frac{(ia)^2}{2} + \cdots$, where $\frac{S(k)}{a^k} \to 0$ as $a \to 0$.

\[ |S(a)| = |e^{ia} - 1 - ia - \frac{(ia)^2}{2}| \leq 2 + |a| + |a|^2 \]

Thus $\frac{|S(a)|}{a^3}$ is bounded, goes to 0 at 0.

Durrett Lemma 2.3.6, p. 101 says (n=2 case)

\[ \frac{|S(a)|}{a^3} \leq \min \left( \frac{|a|}{4}, 1 \right) \]

This is the infamous $f$ from last class.

Anyway, $|S(a)| \leq \frac{|a|^3}{\sigma}$. Writing $q$ for the ch. f. of $X_1$, and remembering that $\mathbb{E} X_1 = 0$ and $\sigma = 1$,

\[ |q(t)| \leq \left| e^{itX_1} - 1 - itX_1 - \frac{(itX_1)^2}{2} \right| \leq \mathbb{E} \left| e^{itX_1} - 1 - itX_1 - \frac{(itX_1)^2}{2} \right| \leq \frac{1}{\sigma} \mathbb{E} |tx|^3 = \frac{\rho |t|^3}{\sigma} \]

Then

\[ |q(t)| \leq |q(t)| - 1 + \frac{1}{2} t^2 | + |1 - \frac{1}{2} t^2 | \leq \frac{\rho |t|^3}{\sigma} + |t|^2 | + |t|^2 | \leq 1 - \frac{5}{18} |t|^2 \quad \text{provided} \quad |t| \leq \frac{4}{3} \rho \]

\[ \leq \exp \left( -\frac{5}{18} |t|^3 \right) \]
Now think of \( \alpha = \varphi (\tau) \), \( |\alpha| \leq \exp \left(-\frac{5}{18} \tau^2 \right) \) \( |\tau| \leq \min \left(\frac{4}{3} \frac{r}{p} \sqrt{2} \right) \),

\[ \beta = \exp \left(-\frac{5}{4} \tau^2 \right) \]

\[ \gamma = \exp \left(-\frac{5}{18} \tau^2 \right), \] the upper bound.

We get, provided \( |\tau| \leq \min \left(\frac{4}{3} \frac{r}{p} \sqrt{2} \right) \),

\[ |\psi_n(t) - \psi(t)| \leq n \left| x - p \right| \varphi^{n-1} \]

\[ \left( \frac{n=10}{\frac{5}{18} \frac{9}{10} = \frac{1}{4}} \right) \]

\[ \tau^{(n-1)} = \frac{t_{(n-1)}}{\tau} \]

\[ \geq \frac{9 + \tau^2}{10} \]

\[ \varphi^{(n-1)} \]

\[ \leq n e^{-\frac{1}{4} \tau^2} \left[ \varphi (\tau) - \frac{1}{\alpha^2} (1 - \frac{1}{\alpha^2}) + (1 - \frac{1}{\alpha^2}) - e^{-\frac{1}{4} \tau^2} \right] \]

\[ \leq n e^{-\frac{1}{4} \tau^2} \left[ \frac{\rho}{\alpha} \frac{(\alpha |t|)^3}{\alpha} + x^2 \frac{\alpha^2}{\alpha^2} \right] \]

\[ \leq n e^{-\frac{1}{4} \tau^2} \left[ \frac{\rho}{\alpha} \frac{(\alpha |t|)^3}{\alpha} + x^2 \right] \]

alternating series

\[ = n e^{-\frac{1}{4} \tau^2} \left[ \frac{\rho}{\alpha} \frac{t^3}{\alpha^3} + \frac{1}{\alpha^2} \frac{t^4}{\alpha^4} \right] \]

\[ = e^{-\frac{1}{4} \tau^2} \left[ \frac{\rho}{\alpha} \frac{t^3}{\alpha^3} + \frac{1}{\alpha^2} \frac{t^4}{\alpha^4} \right] \]

provided that \( |t| \leq \sqrt{n} \min \left(\frac{4}{3} \frac{r}{p} \sqrt{2} \right) \).

Thus, we choose \( L = \sqrt{n} \min \left(\frac{4}{3} \frac{r}{p} \sqrt{2} \right) \).

As \( n \to \infty \), the random variable \( \psi_L \) that we add gets more and more concentrated near 0.
Graph of the complex exponential function

Write $z = a + i b$ and represent the graph of

$$(a, b) \rightarrow e^{a + i b} = e^a (\cos(b) + i \sin(b)) = (e^a \cos(b), e^a \sin(b))$$

in a graphical way.

The line with constant $a$ are mapped to circles with radius $e^a$.
The only point that is not hit by $\exp$ is the origin, $O$.
Using values of $b$ above $\pi$ or below $-\pi$ maps points onto the same circles,
so we see that $\exp$ is not one to one.

Still, we can define an inverse of $\exp$ if we restrict the domain to $a \in \mathbb{R}$, $-\pi < b \leq \pi$.

The complex natural logarithm

We define $\log$ to be the inverse of $\exp$ when the domain of $\exp$ is restricted as above. This is called the "principal branch" of $\log$; one sometimes thinks of $\log$ as a "multiply-valued function," but not here.

Note that $\log(0)$ is not defined, and $\log(-1) = i\pi$. 
Taylor series for \( \log (1 + z) \)

As in the case of the real-number logarithm, \( \frac{1}{z} \log (z) = \frac{1}{z} \).

Let \( f(z) = \log (1 + z) \).

Then \( f'(z) = \frac{1}{1 + z} \); \quad f'(0) = 1 \)

\( f''(z) = -2(1 + z)^{-2} \); \quad f''(0) = -1 \)

\( f'''(z) = 2(1 + z)^{-3} \); \quad f'''(0) = 2 \)

\( f^{(4)}(z) = -3(1 + z)^{-4} \); \quad f^{(4)}(0) = -3 \)

\( f^{(n)}(z) = (-1)^{n+1}(n-1)! (1 + z)^{-n}; \quad f^{(n)}(0) = (-1)^{n+1}(n-1)! \)

This explains, but does not fully justify, this Taylor series:

\[
\log (1 + z) = z - \frac{z^2}{2} + \frac{z^3}{3} - \frac{z^4}{4} + \cdots, \quad |z| < 1.
\]

We need \( |z| < 1 \) to avoid \( z = -1 \), \( \log (0) \).

**Quadratic bound**

Define \( h(z) \) for \( z \neq -1 \) implicitly by \( \log (1 + z) = z + h(z) \).

Then

\[
|h(z)| = |\log (1 + z) - z| \\
= \left| -\frac{z^2}{2} + \frac{z^3}{3} - \cdots \right| \\
\leq \frac{1}{2} \sum_{n=2}^{\infty} \frac{|z|^n}{n} \\
\leq \frac{1}{2} \cdot \frac{2}{n=2} |z|^n \\
= \frac{1}{2} \left| \frac{|z|^2}{1-|z|} \right| \quad \text{geometric series} \\
\leq \frac{1}{2} \left| \frac{|z|^2}{1-\frac{1}{2}} \right| \quad \text{for } |z| < \frac{1}{2}, \quad 1-|z| > \frac{1}{2} \\
= |z|^2
\]

This is about all we will need to know about the complex logarithm.
Theorem 2.4.2, p. 112

Suppose \((c_n) \subset \mathbb{C}\) satisfies \(c_n \rightarrow c\) as \(n \rightarrow \infty\).

Then \((1 + \frac{c_n}{n})^n \rightarrow \exp(c)\) as \(n \rightarrow \infty\).

Proof (using complex exponential and log)

\[
(1 + \frac{c_n}{n})^n = \exp\left(\log\left(1 + \frac{c_n}{n}\right)^n\right) \\
= \exp\left(n \cdot \log\left(1 + \frac{c_n}{n}\right)\right) \\
= \exp\left(n \cdot \left(\frac{c_n}{n} + h\left(\frac{c_n}{n}\right)\right)\right) \\
= \exp\left(c_n + n \cdot h\left(\frac{c_n}{n}\right)\right).
\]

But for \(n\) large enough, \(|\frac{c_n}{n}| \leq \frac{1}{2}\), so

\[
|n \cdot h\left(\frac{c_n}{n}\right)| \leq n \cdot \frac{|c_n|}{n} = \frac{|c_n|^2}{n},
\]

which goes to 0 as \(n \rightarrow \infty\).

Thus, \(c_n + n \cdot h\left(\frac{c_n}{n}\right) \rightarrow c\) as \(n \rightarrow \infty\), and so

\((1 + \frac{c_n}{n})^n \rightarrow \exp(c)\). \qed
Central Limit Theorem for triangular arrays

To what extent must the distributions of the summands in \( X_1 + \ldots + X_n \) be the same? When the \( X_i \) are i.i.d. with variance \( \sigma^2 \), let

\[
Z_n = \frac{X_1 + \ldots + X_n - n \mu}{\sqrt{n}}. 
\]

Then \( Z_n \) has mean 0 and variance 1.

\[ \sigma Z_n = \frac{X_1 - EV}{\sqrt{V_n}} + \ldots + \frac{X_n - EV}{\sqrt{V_n}} \] is a sum of \( n \) independent r.v.'s with variance \( \frac{\sigma^2}{n} \), mean 0.

Think of this as the sum of \( n \) independent r.v.'s whose variance sums to \( \sigma^2 \).

That's just about all we need.

Imagine a triangular array of this form:

\[ \begin{aligned}
X_{1,1} & \\
X_{2,1} & X_{2,2} \\
\vdots & \\
X_{n,1} & \ldots & X_{m,m} & \ldots & X_{n,n}
\end{aligned} \]

Where, in each row, the r.v.'s are independent and their variances add up to approx. \( \sigma^2 \).

Set \( Z_n = X_{n,1} + \ldots + X_{n,n} \), let \( Z \sim N(0,1) \).

Theorem 2.4.5, p. 116

Lindeberg-Feller Central Limit Theorem

Suppose \( X_{n,1}, \ldots, X_{n,n} \) are independent and have mean 0.

Suppose (i) \( \lim_{n \to \infty} \frac{\sum_{m=1}^{n} E[X_{n,m}^2]}{n} = \sigma^2 > 0 \)

(ii) For all \( \varepsilon > 0 \),

\[
\lim_{n \to \infty} \sum_{m=1}^{n} E[|X_{n,m}|^2 \mathbb{1}_{|X_{n,m}| \geq \varepsilon}] = 0.
\]

Then \( Z_n \to \sigma Z \) as \( n \to \infty \).

Remark. \( Z_n \) is the sum of a large number of independent random variables.

\( Z_n \) has mean 0 and variance approaching \( \sigma^2 \).

Then the distribution of \( Z_n \) is approximately normal.

- This generalizes our previous CLT.

- Condition (i) forces the variances of \( X_{n,m} \) to sum to \( \sigma^2 \).

Condition (ii) prevents all the variance being concentrated in a few of the \( X_{n,m} \).

It forces the \( X_{n,m} \) to take on small values.

Otherwise we could have, say, \( X_{n,1} \sim \text{Uniform}([-a, a]) \) for all \( n \) and \( X_{n,2} = \ldots = X_{n,n} = 0 \) to satisfy (i).

- Condition (ii) says that values \( \leq \varepsilon \) eventually account for all the variance of the row sum, rather than a few large r.v.'s doing it.
Proof

Let \( \phi_{nm} \) be the ch. function of \( X_{nm} \).
Let \( \phi_n \) be the ch. function of \( Z_n \).

Fix \( t \in \mathbb{R} \). By independence,

\[
\phi_n(t) = \phi_{n1}(t) \cdot \phi_{n2}(t) \cdots \phi_{nn}(t) = \prod_{m=1}^{n} \phi_{nm}(t)
\]

Now for \( n \) large enough, all the second moments of \( X_{nm} \) are finite since their sum converges to \( \sigma^2 \) by (i).

Thus,
\[
\phi_{nm}(t) = 1 - \frac{1}{2} \sigma_{nm}^2 t^2 + R_{nm}(t)
\]

where, from the Taylor series expansion of exp, we have
\[
|R_{nm}(t)| \leq E|R_z(it X_{nm})| |X_{nm}|^2 \cdot t^2
\]

where \( |R_z| \) is bounded by \( r_z \) and \( \lim_{t \to 0} R_z(it) = 0 \).

To multiply these \( \phi_{nm} \) together, it helps to use exponentials:

\[
\phi_{nm}(t) = \exp \left( \ln \left( 1 - \frac{1}{2} \sigma_{nm}^2 t^2 + R_{nm}(t) \right) \right)
\]

\[
= \exp \left( \ln \left( 1 + Z_{nm} \right) \right), \quad Z_{nm} = -\frac{1}{2} \sigma_{nm}^2 t^2 + R_{nm}(t)
\]

\[
= \exp \left( Z_{nm} + h(Z_{nm}) \right), \quad h(z) = \ln(1+z) - z.
\]

Then
\[
\phi_n(t) = \prod_{m=1}^{n} \phi_{nm}(t)
\]

\[
= \exp \left( \sum_{m=1}^{n} Z_{nm} + \sum_{m=1}^{n} h(Z_{nm}) \right)
\]

We will find the limits of these sums as \( n \to \infty \) are \( -\frac{1}{2} \sigma^2 t \) and \( 0 \), respectively.
Fix $\varepsilon > 0$, $n$, and $m$. Then,

\[ \sum_{m=1}^{\infty} |R_{nm}(t)| \leq \sum_{m=1}^{\infty} E|R_n(tX_{nm})| t^2 \varepsilon^2 = \sum_{m=1}^{\infty} \max_{\omega \in \mathcal{F}} |R_n(t\omega)| \cdot t^2 \cdot E|X_{nm}|^2 1_{|X_{nm}| \leq \varepsilon} + \sum_{m=1}^{\infty} \max_{\omega \in \mathcal{F}} |R_n(t\omega)| \cdot t^2 \cdot E|X_{nm}|^2 1_{|X_{nm}| > \varepsilon} + \varepsilon t^2 \cdot E|X_{nm}|^2 1_{|X_{nm}| > \varepsilon} \]

The first sum converges to $\sigma^2$, the second converges to 0, by assumption (ii).

Thus,

\[ \limsup_{n \to \infty} \sum_{m=1}^{n} |R_{nm}(t)| \leq \max_{\omega \in \mathcal{F}} \{ R_n(t\omega) \cdot t^2 \cdot \sigma^2 + \varepsilon t^2 \cdot 0 \} \]

As $\varepsilon \to 0$, the right hand side goes to 0 because $R_n(t\omega)$ is continuous at 0. Thus, \( \lim_{n \to \infty} \sum_{m=1}^{n} |R_{nm}(t)| = 0 \).

\[ \lim_{n \to \infty} \sum_{m=1}^{n} Z_{nm} = -\frac{1}{2} \sigma^2 t^2 \]

\[ \sum_{m=1}^{n} Z_{nm} = \sum_{m=1}^{n} \sum_{m=1}^{n} -\frac{1}{2} \sigma^2 t^2 + \sum_{m=1}^{n} R_{nm}(t) \]

\[ \to -\frac{1}{2} \sigma^2 t^2 + 0 \]

by (2) and assumption (i).

\[ \lim_{n \to \infty} \sup_{1 \leq m \leq n} \sigma_{nm}^2 = 0 \]

The variances of $X_{nm}$ go uniformly to 0.

Fix $\varepsilon > 0$, $m$, $n$.

\[ \sigma_{nm}^2 = E X_{nm}^2 = E X_{nm}^2 1_{|X_{nm}| \leq \varepsilon} + E X_{nm}^2 1_{|X_{nm}| > \varepsilon} \]

\[ \leq \varepsilon^2 + \sum_{m=1}^{n} E X_{nm}^2 1_{|X_{nm}| > \varepsilon} \]

\[ \sup_{1 \leq m \leq n} \sigma_{nm}^2 \leq \varepsilon^2 + \sum_{m=1}^{n} E X_{nm}^2 1_{|X_{nm}| > \varepsilon} \]

As $n \to \infty$, the right side goes to $\varepsilon^2 + 0$ by assumption (ii).

Letting $\varepsilon \to 0$, \( \lim_{n \to \infty} \sup_{1 \leq m \leq n} \sigma_{nm}^2 = 0 \).
\[ \lim_{n \to \infty} \sum_{m=1}^{n} h(Z_{nm}) = 0 \]

For \( n \) large enough, \( |Z_{nm}| = \left| 1 - \frac{1}{2} \sigma_{n}^{2} t^2 + R_{nm}(t) \right| \leq \frac{1}{2} \) for all \( m \) by 2 and 4.

\[ \sum_{m=1}^{n} h(Z_{nm}) \leq \sum_{m=1}^{n} |Z_{nm}| = \left( \sup_{1 \leq m \leq n} |Z_{nm}| \right) \cdot \sum_{m=1}^{n} |Z_{nm}| \]

\[ = \left( \sup_{1 \leq m \leq n} \left| 1 - \frac{1}{2} \sigma_{n}^{2} t^2 + R_{nm}(t) \right| \right) \cdot \sum_{m=1}^{n} \left| 1 - \frac{1}{2} \sigma_{n}^{2} t^2 + R_{nm}(t) \right| \]

\[ \leq \left( \sup_{1 \leq m \leq n} \frac{1}{2} t^2 \sigma_{n}^{2} + \sum_{m=1}^{n} |R_{nm}(t)| \right) \cdot \left( \sum_{m=1}^{n} \frac{1}{2} t^2 \sigma_{n}^{2} + \sum_{m=1}^{n} |R_{nm}(t)| \right), \]

The right hand side converges to \((0 + 0) \cdot (\frac{1}{2} t^2 \sigma^2 + 0)\) as \( n \to \infty \).

Thus, \( \lim_{n \to \infty} \sum_{m=1}^{n} h(Z_{nm}) = 0 \).

\( \square \)

Thus, the characteristic function \( \Psi_n \) of \( Z_n \) converges pointwise to \( e^{-\frac{1}{2} \sigma^2 t^2} \), the characteristic function of the \( N(0, \sigma^2) \) distribution.

By the continuity theorem, \( Z_n \to eZ \).
Comparison of limit theorems

CLT
\[ S_n = X_1 + \ldots + X_n \]
\[ E S_n = n \cdot E X_1 \]
\[ Var(S_n) = n \cdot Var(X_1) \]
both grow linearly
Subtract the mean, divide by \( \sqrt{Var(S_n)} \):
\[ Z_n = \frac{X_1 + \ldots + X_n - n \cdot E X_1}{\sigma \sqrt{n}} \rightarrow Z. \]

LF CLT
Means are already O, rescaling is already being done.
\[ Z_n = X_{n,1} + \ldots + X_{n,n} \]
\[ Var(Z_n) = \sum_{k=1}^{n} Var(X_{n,k}) \rightarrow \text{constant}. \]

Poisson convergence setting
\( X_{n,k} \) are Bernoulli:
\[ P(X_{n,k} = 1) = \mu_{nk} \]
\[ P(X_{n,k} = 0) = 1 - \mu_{nk} \]
- means are close to 0
\[ E X_{n,k} = \mu_{nk} \]
- variances are small
\[ Var(X_{n,k}) = \mu_{nk} (1 - \mu_{nk}) \]
- however, r.v.'s take values 0 and 1, not values getting closer to 0.
2.6 Poisson convergence

The Lindeberg-Feller CLT had a triangular array of random variables, independent across each row, sum of variances converging to \( \sigma^2 \), and satisfying

\[
\sum_{m=1}^{n} \mathbb{E} X_{nm}^2 1_{(X_{nm} < \epsilon)} \rightarrow \sigma^2, \quad \text{as } n \to \infty
\]

so that all of the variance was accounted for by values smaller than \( \epsilon \).

Think of row \( n \) as \( \frac{X_1}{\sqrt{n}}, \frac{X_2}{\sqrt{n}}, \ldots, \frac{X_n}{\sqrt{n}} \), as in our first CLT.

The rescaling makes these random variables small.

In other situations, the \( X_{nm} \) are Bernoulli distributed (0-1 valued random variables), and the row sum

\[
S_n = X_{n1} + \cdots + X_{nn}
\]

is the number of times something happens.

In the case that the mean and variance of \( S_n \) grow linearly, we can standardize \( S_n \) and use the CLT.

There are important cases, however, in which the mean (and variance) of \( S_n \) stay constant, so we don't want to rescale.

We will use independence and characteristic functions to show that \( S_n \) converges in distribution to a Poisson random variable.

Example

A long piece of fiber optic cable will have a number of small flaws or defects. The number of flaws is random. We can argue that the number of flaws should have a Poisson distribution.

Fix \( n \) and imagine dividing the cable into \( n \) segments of equal length.
Let \( X_{nm} \) equal 1 if there is a flaw in segment \( m \), 0 otherwise.
It is fairly reasonable to think that the \( X_{nm} \) are independent, at least for certain types of flaws.

Now \( S_n = X_{n1} + \cdots + X_{nn} \) undercounts the total number of flaws, since there could be two or more flaws in any one segment. However, as \( n \to \infty \), \( S_n \) converges to the total number of flaws.
Theorem 2.6.1, p. 137

For each \( n \), suppose \( X_1, \ldots, X_n \) are independent Bernoulli random variables with success probabilities \( p_1, \ldots, p_n \); \( P(X_{nm} = 1) = p_m \). Let \( S_n = X_1 + \cdots + X_n \nmspace{3mu} \)

Suppose \( (i) \quad E S_n = \sum_{m=1}^{n} p_m \to \lambda \) as \( n \to \infty \)

\( \nmspace{3mu} \)

\( (ii) \quad \max_{1 \leq m \leq n} p_m \to 0 \) as \( n \to \infty \).

Then \( S_n \Rightarrow S \) where \( S \sim \text{Poisson}(\lambda) \).

Proof: A straightforward computation.

Recall that \( \ln(1 + z) = z + o(z) \) and \( |o(z)| \leq |z|^2 \) for \( |z| < \frac{1}{2} \).

Fix \( t \in \mathbb{R} \).

\[
E e^{itS_n} = \prod_{m=1}^{n} E e^{itX_m} = \prod_{m=1}^{n} \left[ p_m e^{it} + (1 - p_m) e^0 \right] = \prod_{m=1}^{n} \exp \left( \ln \left( p_m (e^{it} - 1) + 1 \right) \right) = \exp \left( \sum_{m=1}^{n} \left( p_m (e^{it} - 1) + h(p_m(e^{it} - 1)) \right) \right)
\]

Claim \( \lim_{n \to \infty} \sum_{m=1}^{n} h(p_m(e^{it} - 1)) = 0 \).

Fix \( n \) large enough that \( \max_{1 \leq m \leq n} |p_m(e^{it} - 1)| < \frac{1}{2} \).

\[
0 \leq \left| \sum_{m=1}^{n} h(p_m(e^{it} - 1)) \right| \leq \sum_{m=1}^{n} |h(p_m(e^{it} - 1))| \leq \sum_{m=1}^{n} p_m^2 |e^{it} - 1|^2 \quad \text{because } |p_m(e^{it} - 1)| < \frac{1}{2}
\]

\[
\leq 4 \sum_{m=1}^{n} p_m^2 \leq 4 \left( \max_{1 \leq m \leq n} p_m \right) \sum_{m=1}^{n} p_m
\]

Use the Squeeze Law. The RHS has limit \( 4 \cdot 0 \cdot \lambda = 0 \).

Thus, \( \lim_{n \to \infty} E e^{itS_n} = \exp (\lambda(e^{it} - 1)) \)

which is continuous at \( t = 0 \). This is the characteristic function of the Poisson distribution with parameter \( \lambda \). By the Continuity Theorem, \( S_n \) converges in distribution to a \( \text{Poisson}(\lambda) \) variable. \( \Box \)
Alternative hypotheses

Note that \[ \text{Var}(S_n) = \sum_{m=1}^{n} \text{Var}(X_{nm}) \]
\[ = \sum_{m=1}^{n} \mathbb{E} X_{nm}^2 - (\mathbb{E} X_{nm})^2 \]
\[ = \sum_{m=1}^{n} p_{nm} - \left( \sum_{m=1}^{n} p_{nm} \right)^2, \]

which converges to \( \lambda \) by assumptions (i) and (ii).

Show: \( \mathbb{E} S_n \to \lambda \) and \( \text{Var}(S_n) \to \lambda \),
then (i) and (ii) hold.

Solution \[ \max_{1 \leq m \leq n} p_{nm} \leq \sum_{m=1}^{n} p_{nm}, \] which goes to \( 0 \).

Non-Bernoulli case

Theorem 2.6.7, p. 145

Suppose \( X_{nm}, 1 \leq m \leq n \) are independent random variables taking values in \( \{0, 1, 2, \ldots \} \). Let \( p_{nm} = \mathbb{P}(X_{nm} = 1) \) and \( E_{nm} = \mathbb{P}(X_{nm} > 1) \). Suppose that, as \( n \to \infty \),

(i) \[ \sum_{m=1}^{n} p_{nm} \to \lambda \]

(ii) \[ \max_{1 \leq m \leq n} p_{nm} \to 0 \]

(iii) \[ \sum_{m=1}^{n} E_{nm} \to 0 \]

Then \( S_n = X_{n1} + \ldots + X_{nn} \) converges in distribution to \( \text{Poisson}(\lambda) \).
Thinning a Poisson random variable

Example  Think again of fiber-optic cable. In producing the cable, there are occasional flaws, and some of these flaws are fatal in the sense that the cable won't work. Let $X_1, X_2, \ldots$ be Bernoulli random variables, where $X_n = 1$ if the $n$th flaw is fatal and $X_n = 0$ if not. Let $N$ be the number of flaws in a certain length of cable. The number, $F$, of fatal flaws in that cable is then

$$ F = X_1 + X_2 + \cdots + X_N. $$

We will now see conditions under which $F$ is again Poisson.

Proposition  Let $X_1, X_2, \ldots$ be Bernoulli with success probability $p$, let $N$ be Poisson ($\lambda$), and suppose that $N, X_1, X_2, \ldots$ are independent. Let $F = X_1 + \cdots + X_N$ ($F = 0$ if $N = 0$). Then $F \sim$ Poisson ($\lambda p$).

Proof  We simply compute the characteristic function of $F$. Let $t \in \mathbb{R}$.

$$ E e^{itF} = E e^{it(X_1 + \cdots + X_N)} = E \sum_{n=0}^{\infty} \mathbf{1}_{\{N = n\}} e^{it(X_1 + \cdots + X_n)} = \sum_{n=0}^{\infty} E \mathbf{1}_{\{N = n\}} e^{it(X_1 + \cdots + X_n)} = \sum_{n=0}^{\infty} E \mathbf{1}_{\{N = n\}} (E e^{itX_1})^n = \sum_{n=0}^{\infty} E \mathbf{1}_{\{N = n\}} (e^{it} + (1-p)) \lambda^n e^{-\lambda} \frac{\lambda^n}{n!} (1 + p(e^{it} - 1))^n = e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} ((1 + p(e^{it} - 1))^n \frac{(\lambda(1 + p(e^{it} - 1)))^n}{n!} = e^{-\lambda} e^{\lambda (1 + p(e^{it} - 1))} = e^{\lambda p(e^{it} - 1)}, $$

which is the characteristic function of the Poisson ($\lambda p$) distribution. By the Uniqueness Theorem, $F$ has the Poisson ($\lambda p$) distribution. \(\Box\)
Rate of convergence to Poisson

**Example** Let $S_n \sim \text{Binomial}(n, p)$.

Then $E S_n = np$ and $\text{Var}(S_n) = np(1-p)$.

If we allow $p$ to vary with $n$, $p_n = \frac{\lambda}{n}$, then $E S_n = \lambda$ and $\text{Var}(S_n) = \lambda \left(1 - \frac{\lambda}{n}\right) \to \lambda$.

Thus, $S_n \Rightarrow S$ where $S \sim \text{Poisson}(\lambda)$.

Durrett p. 141 does an explicit calculation which bounds the largest possible distance between the distributions of $S_n$ and $S$:

$$\sup_{A \in \mathcal{Z}} |P(S_n \in A) - P(S \in A)| \leq \frac{3}{4e\mu}$$

Our goal now is to find such a bound in general.

(The same can be done for fixed $p$, $\frac{S_n - np}{\sqrt{np(1-p)}}$ converging to $N(0, 1)$.)
Total variation distance

Let $C$ be a countable set and $\mu$ and $\nu$ two probability measures on it. (discrete $\sigma$-algebra)

Then $\mu$ is essentially a sequence $\mu(c_1), \mu(c_2), \ldots$

The total variation distance between $\mu$ and $\nu$ is defined to be

$$\|\mu - \nu\| = 2 \sup_{A \subseteq C} |\mu(A) - \nu(A)|$$

It is useful to write this a different way.

Note that, for $A \subseteq C$,

$$2 |\mu(A) - \nu(A)| = |\mu(A) - \nu(A)| + |\mu(A) - 1 + 1 - \nu(A)|$$

$$= |\mu(A) - \nu(A)| + |\mu(A^c) - \nu(A^c)|$$

$$= |\sum_{c \in A} \mu(c) - \nu(c)| + |\sum_{c \in A^c} \mu(c) - \nu(c)|$$

$$\leq \sum_{c \in C} |\mu(c) - \nu(c)|,$$

what you might call the $L^1$ distance between $\mu$ and $\nu$.

So

$$\|\mu - \nu\| \leq \sum_{c \in C} |\mu(c) - \nu(c)|.$$

Let $A = \{ c : \mu(c) \geq \nu(c) \}$.

Then

$$|\sum_{c \in A} \mu(c) - \nu(c)| = \sum |\mu(c) - \nu(c)|$$

$$|\sum_{c \in A^c} \mu(c) - \nu(c)| = \sum |\mu(c) - \nu(c)|,$$

all signs are the same.

So

$$2 |\mu(A) - \nu(A)| = \sum_{c \in C} |\mu(c) - \nu(c)|,$$

in this case.

Thus,

$$\|\mu - \nu\| = \sum_{c \in C} |\mu(c) - \nu(c)|.$$

This makes it clear that total variation distance is a metric.
Do Exercise 2.6.1, p. 139 on your own.

Exercise Ex. 2.6.2, p. 189.

Show that \( \| \mu - \nu \| \leq 2\delta \) if and only if there are random variables \( X \) and \( Y \) with distributions \( \mu \) and \( \nu \) so that \( \Pr(X \neq Y) < \delta \).
Theorem 2.6.1  

For each \( n \), let \( X_1, \ldots, X_n \) be independent Bernoulli, success probs \( p_1, \ldots, p_n \). Let \( S_n = X_1 + \cdots + X_n \). Let \( \lambda_n = E(S_n) \).

Suppose \( \lambda_n \to \lambda \in (0, \infty) \) and \( \text{Var}(S_n) \to \lambda \).

Let \( \mu_n \) be the distribution of \( S_n \) and \( \nu_n \) be the Poisson \( (\lambda_n) \) distribution.

Then \( \| \mu_n - \nu_n \| \leq 2 \sum_{m=1}^{n} p_m^2 \).

Since this goes to 0 and \( \nu_n \) converges weakly to Poisson \( (\lambda) \), \( \mu_n \) converges weakly to Poisson \( (\lambda) \).

Proof.  

We will need a few lemmas, but here is the basic idea.

Let \( \mu_{nm} \) be the Bernoulli \( (p_{nm}) \), \( \nu_{nm} \) be the Poisson \( (p_{nm}) \).

Then \( \mu_n = \mu_{n1} \ast \mu_{n2} \ast \cdots \ast \mu_{nn} \), the convolution, which corresponds to adding the random variables.

\[
\| \mu_n - \nu_n \| = \| \mu_{n1} \ast \cdots \ast \mu_{nn} - \nu_{n1} \ast \cdots \ast \nu_{nn} \| \\
\leq \sum_{m=1}^{n} \| \mu_{nm} - \nu_{nm} \| \quad \text{* Lemma} \\
\leq 2 \sum_{m=1}^{n} p_m^2 \quad \text{* Lemma}
\]

We have seen that this goes to 0.

Also, \( \| \mu_n - \nu_n \| \leq \| \mu_n - \nu_{n1} \| + \| \nu_{n1} - \nu_n \| \), both of which go to 0.

\[\square\]

**Convolution**

Let \( X_1, X_2 \) be \( \mathbb{Z} \)-valued, \( \alpha \)-disjoint \( \mu_1, \mu_2 \).

Let \( \mu_1 \ast \mu_2 \) denote the dist'n of \( X_1 + X_2 \).

\[
(\mu_1 \ast \mu_2)(x) = \mathbb{P}(X_1 + X_2 = x) \\
= \sum_{y} \mathbb{P}(X_1 + X_2 = x, X_2 = y) \\
= \sum_{y} \mathbb{P}(X_1 = x - y, X_2 = y) \\
= \sum_{y} \mu_1(x-y) \mu_2(y).
\]
Lemmas 6.3, 6.4

Let \( \mu_1, \mu_2, \nu_1, \nu_2 \) be prob. measures on \( \mathbb{X} \).

Then \( \| \mu_1 \times \mu_2 - \nu_1 \times \nu_2 \| \leq \| \mu_1 - \nu_1 \| + \| \mu_2 - \nu_2 \| \).

Proof

\[
\| \mu_1 \times \mu_2 - \nu_1 \times \nu_2 \| = \sum_x |(\mu_1 \times \mu_2)(x) - (\nu_1 \times \nu_2)(x)|
\]

\[
= \sum_x \left| \sum_y \mu_1(x,y) \mu_2(y) - \nu_1(x,y) \nu_2(y) \right|
\]

\[
\leq \sum_x \sum_y | \mu_1(x,y) \mu_2(y) - \nu_1(x,y) \nu_2(y) |
\]

\[
eq \sum_j \sum_{\varepsilon} | \mu_1(z) \mu_2(\varepsilon) - \nu_1(z) \nu_2(\varepsilon) |
\]

\[
\leq \sum_j \sum_{\varepsilon} | \mu_1(z) \mu_2(\varepsilon) - \nu_1(z) \mu_2(\varepsilon) |

+ \sum_j \sum_{\varepsilon} | \mu_1(z) \nu_2(\varepsilon) - \nu_1(z) \nu_2(\varepsilon) |
\]

\[
= \sum_j \mu_2(\varepsilon) \sum_{\varepsilon} | \mu_1(z) - \nu_1(z) | + \cdots
\]

\[
\leq \| \mu_1 - \nu_1 \| + \| \mu_2 - \nu_2 \|.\]