Probability I
Central Limit Theorem

Daniel Yu

October 8, 2024

Contents

1 Review

Recall
Let X_1 be any random variable with $E[X_1] = \mu$ and $Var(X_1) = \sigma^2$ and $\{X_i\}$ be an iid sequence. Then,

$$Z_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \text{ the average of the first n experiments.}$$

Theorem 1. The weak law of large numbers states: $Z_n \to \mu$ in probability or equivalently, $\forall \varepsilon > 0$:

$$\lim_{n\to\infty} P[| Z_n - \mu |> \varepsilon] = 0.$$

or equivalently,

$$\lim_{n\to\infty} P[| \sum_{i=1}^{n} X_i - n\mu |> \varepsilon n] = 0.$$

what this is saying is that there is emergent determinism, the probability of everything outside of this epsilon, will have eventually have probability 0!

Proof. First calculate the variance and expected value then apply chebyslev.

$$\begin{aligned} E[Z_n] &= E[\frac{1}{n}\sum_{i=1}^{n} X_i] \\ &= \frac{1}{n}\sum_{i=1}^{n} E[X_i] \\ &= \frac{nE[X_i]}{n} \\ &= \mu. \end{aligned}$$

Then,

$$\begin{aligned} Var(Z_n) &= Var(\frac{1}{n}\sum_{i=1}^{n} X_i) \\ &= \frac{1}{n}^2 Var(\sum_{i=1}^{n} X_i) \\ &\text{since } X_i \text{ are iid} \\ &= \frac{1}{n}^2 \sum_{i=1}^{n} Var(X_i) \\ &= \frac{1}{n}^2 n \cdot \sigma^2 \\ &= \frac{\sigma^2}{n}. \end{aligned}$$

Apply chebyslev's inequality.

$$\begin{aligned} P[| Z_n - \mu |> \varepsilon] &\leq \frac{Var(Z_n)}{\varepsilon^2} \\ &= \frac{\sigma^2}{\varepsilon^2 n} \\ &\to 0 \text{ as } n \to \infty. \end{aligned}$$

□

Note. Big Idea

What is a random variable with variance equal to 0? It is a deterministic random variable. In fact, in the proof above for $\{X_i\}$ we only use iid to show taht the variance goes to 0. More generally, as long as $\{Y_i\}$ to be any sequence of random variables such that

1. $E[Y_n] \to \mu$
2. $E[Y_n] \to 0$

Then we have that $Y_n \to \mu$ in probability. Question, does the other direction hold? I.e. does $Y_n \to \mu$ for some sequence $\{Y_i\}$ imply that $E[Y_n] \to \mu$ and $Var(Y_n) \to 0$. The answer is no!

Consider $P[A_n = 0] = 1 - \frac{1}{n}, P[A_n = n] = \frac{1}{n}$ so $A_n \to 0$ in probability. The $A_n \to 0$ in probability. However, $E[A_n] = 0 \cdot (1 - \frac{1}{n}) + n \cdot \frac{1}{n} = 1$ and

$$\begin{aligned} Var(A_n) &= E[A_n^2] - E[A_n]^2 \\ &= \left(n^2 \cdot \frac{1}{n} + 0^2 \cdot (1 - \frac{1}{n})\right) - 1^2 \\ &= n - 1 \\ &\to \infty. \end{aligned}$$

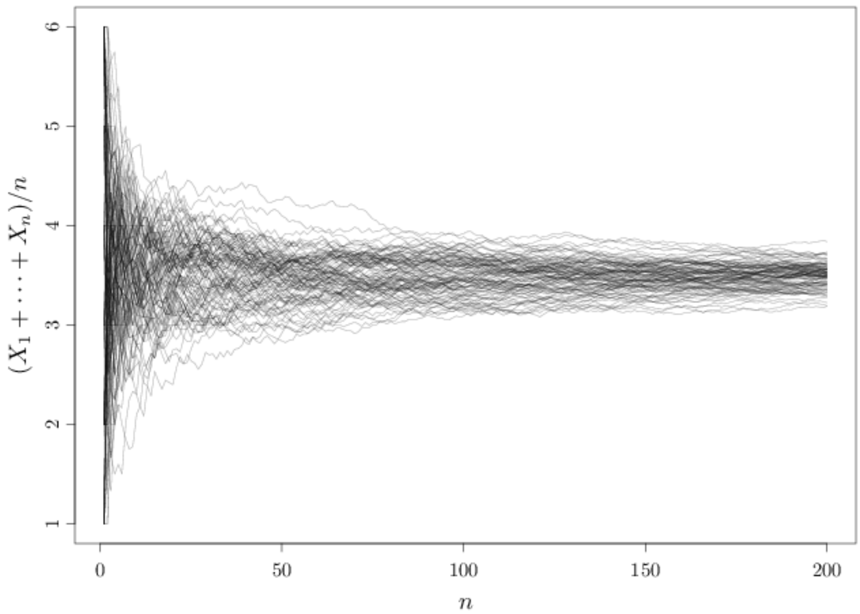

Figure 1: Weak Law of Large Numbers

2 Strongest Statement of Weak Law of Large Numbers and its limitations

Can we do better? in case of averages of idd sequences?

Without loss of generality, let $E[X_1] = 0$ where X_i are iid. What can we say about?

$$P[|\sum_{i=1}^{n} X_i - 0 \cdot \mu \,|] > \varepsilon n^{\frac{2}{3}}.$$

Let's try applying chevslevs inequality again!

$$P[|\sum_{i=1}^{n} X_i |> \varepsilon n^{\frac{2}{3}}] \leq \frac{Var(\sum_{i=1}^{n} X_i)}{\varepsilon^2 n^{\frac{2}{3}}}$$
$$= \frac{n\sigma^2}{\varepsilon^2 n^{\frac{4}{3}}}$$
$$= \frac{\sigma^2}{\varepsilon^2 n^{\frac{1}{3}}}$$
$$\to 0 \text{ as } n \to \infty.$$

So in general, what α should we choose so that

$$P[|\sum_{i=1}^{n} X_i - n\mu |> \varepsilon n^{\alpha}] \to 0?.$$

Clearly, this works as long as $\alpha > \frac{1}{2}$ by chevslev's inequality!

$$P[|\sum_{i=1}^{n} X_i - n\mu |> \varepsilon n^{\alpha}] \leq \frac{\sigma^2 n}{\varepsilon^2 n^2}$$
$$= \frac{\sigma^2}{\varepsilon^2}\frac{1}{n^{2\alpha-1}}.$$

As long as $2 \cdot \alpha - 1 > 0$ then the limit is 0.

Definition 1. Thus, with $\{X_i\}$ iid. sequence and $E[X_1] = \mu$ and $Var(X_1) = \sigma^2$, then as long as $\alpha > \frac{1}{2}$,

$$P[|\sum_{i=1}^{n} X_i - n\mu |> \varepsilon n^{\alpha}] \to 0.$$

Remark. What happens whe $\alpha = \frac{1}{2}$? The weak law of large numbers breaks down ($P[|\sum_{i=1}^{n} X_i - n\mu |> \varepsilon n^{\frac{1}{2}}] \leq \frac{\sigma^2}{\varepsilon^2}$. This bound is trivial oftentimes because it is greater than 1 or extremely loose).

As it turns out,

$$\frac{|\sum_{i=1}^{n} X_i - \mu n|}{\sqrt{n}} \not\to c.$$

in probability, because it is NO LONGER deterministic, it is random!!!

Remark. Our notion of continuity breaks down, we need a new **topology**

3 Convergence in distribution

Definition 2. Define convergence in distribution as follows. Let $\{Y_n\}$ be a sequence of random variables and Y be a random variable. Then $Y_n \to Y$ meaning Y_n converges to Y in distribution if $\forall E \in \mathbb{R}$,

$$\lim_{n\to\infty} P[Y_n \leq t] = P[Y \leq t].$$

If Y is discrete, it is sufficient to check

$$\lim_{n\to\infty} P[Y_n = x] = P[Y = x].$$

Note. Example

1. $P[A_n = 1] = \frac{1}{2} + \frac{1}{n+1}$

2. $P[A_n = -1] = \frac{1}{2} - \frac{1}{n+1}$

Then $A_n \to A$ where A is defined as:

$$P[A = 1] = \frac{1}{2}$$
$$P[A = -1] = \frac{1}{2}.$$

Note. We would like
Let S_n to be defined as the set $\{\frac{-n+1}{n}, \frac{-n+2}{n}, \frac{-n+3}{n}, \ldots, \frac{n-1}{n}, \frac{n-1}{n}\}$. Clearly as $n \to \infty$, the $S_n \to Uniform(-1, 1)$ the uniform distribution over $[-1, 1]$. However, how can we should this?

Proof. We need to compute $P[S_n \leq t]$.

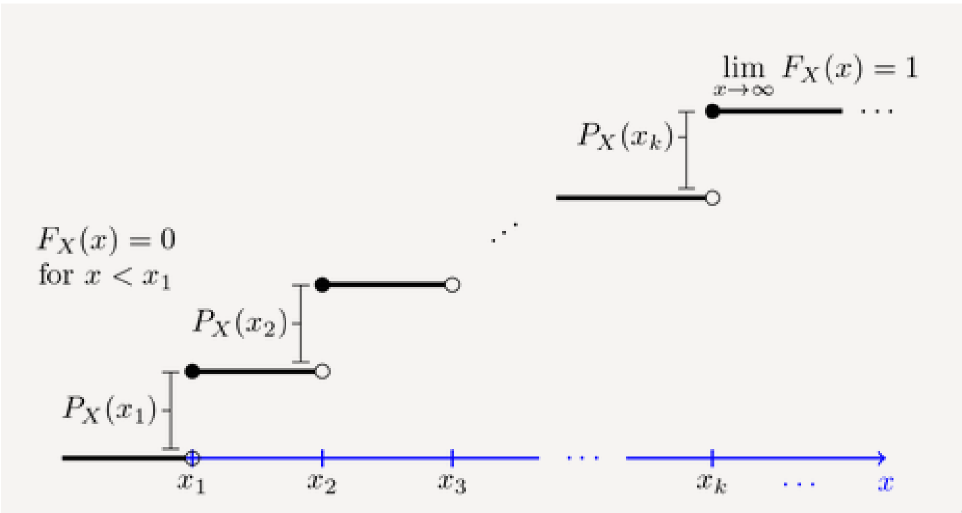

Figure 2: Graph of Probability, where $x_i = \frac{n-i}{n}$, $x_1 = -1$, and $x_k = 1$

We can see that the number of intervals is $2n - 1$.

$$P[S_n \leq 0] = \frac{n+1}{2n-1}.$$

and,

$$P[S_n \leq t] = \frac{n + 1 + \lceil nt \rceil}{2n-1} \forall t \in [-1, 1].$$

Thus,

$$\lim_{n\to\infty} P[S_n \leq t] = \frac{1}{2} + \frac{t}{2} = P[Uniform(-1, 1) \leq t].$$

□

Note. Example
Take $B_n \sim Bin(n, \frac{2}{n})$. Define Z as:

$$\begin{cases} P[Z = 0] = \lim_{n\to\infty} P[B_n = 0] \\ P[Z = 1] = \lim_{n\to\infty} P[B_n = 1] \\ P[Z = 2] = \lim_{n\to\infty} P[B_n = 2] \\ \ldots \\ P[Z = n] = \lim_{n\to\infty} P[B_n = n] \end{cases}.$$

Then, we would like $B_n \to Z$.

Proof.

$$\begin{aligned}
\lim_{n\to\infty} P[B_n = 0] &= \lim_{n\to\infty}\left(1-\frac{2}{n}\right)^n \\
&= e^{-2} \\
&= P[Z=0] \\
P[Z=1] &= \lim_{n\to\infty}\left(\binom{n}{1}\frac{n}{2}^1\left(1-\frac{n}{2}\right)^{n-1}\right) \\
&= 2\cdot e^{-2} \\
&\dots \\
P[Z=k] &= \lim_{n\to\infty}\binom{n}{k}\frac{n}{2}^k\left(1-\frac{n}{2}\right)^{n-k} \\
&= \lim_{n\to\infty}\frac{n\cdot(n-1)\cdot\ldots\cdot(n-k+1)}{k!}\frac{2^k}{n^k}(1-\frac{2}{n})^{n-k} \\
&= \frac{2^k}{k!}\lim_{n\to\infty}(1-\frac{2}{n})^k\cdot\lim_{n\to\infty}\frac{n\cdot(n-1)\cdot\ldots\cdot(n-k+1)}{n^k} \\
&= \frac{e^{-2}\cdot 2^k}{k!} \\
&\sim Poisson(2).
\end{aligned}$$

So $Z\sim Poiss(2)$. In general, with $B_n\sim Bin(n,\frac{\lambda}{n})$, then $B_n\to Z$ for $Z\sim Possion(\lambda)$ □

Theorem 2. What can we say about the sums of iid random variables? X_1 is a random variable with $E[X_1]=\mu<\infty$, $Var(X_1)=\sigma^2<\infty$, $\{X_i\}$ iid.

$$\frac{(\sum_{i=1}^n X_i - n\mu)}{\sigma\sqrt{n}} \Rightarrow N(0,1).$$

Recall that $N(0,1)$ is the random variable with pdf $\frac{1}{\sqrt{2\pi}}e^{\frac{-X^2}{2}}$. This is the **central limit theorem**. Equivalently, this means as $n\to\infty$:

$$P[a\le\frac{(\sum_{i=1}^n X_i - n\mu)}{\sigma\sqrt{n}}\le b]\to P[a\le N(0,1)\le b]=\int_a^b\frac{1}{\sqrt{2\pi}}e^{\frac{-X^2}{2}}dx.$$

Lemma 1. A rough approximation is (keyword rough):

$$\sum_{i=1}^n X_i\approx n\cdot\mu+\sqrt{n}\cdot\sigma\cdot N(0,1)+\ldots.$$

where the other terms grow slower than $\sqrt{n}$

Note. Important Example
Consider W_n the random walk (remember that this is a sum of independent walks across timsteps).

1. What is $P[W_{10000000000}\ge 0]$? $P[W_{10000000000}\ge 0]\approx\frac{1}{2}$ by symmetry and $E[W_{1000000000}]=0$
2. What about $P[W_{1000000}\ge 2000]$? $P[W_{1000000}\ge 2000]=P[\frac{\sum_{i=1}^{1000000}W_i - 0\cdot 1000000}{1000}\ge\frac{2000}{1000}]=P[\frac{\sum_{i=1}^{1000000}W_i}{1000}\ge 2]$
3. Ex: X is a random variable with $P[X_1=2]=\frac{1}{3}, P[X_1=4]\frac{1}{2}, P[X_1=8]=\frac{1}{6}$. Approximate $P[\sum_{i=1}^{10000}X_i\le 39700]$ using central limit theorem. I expect $\sum_{i=1}^{10000}X_i\approx 10,000E[X]$

$$E[X_1]=2\cdot\frac{1}{3}+4\cdot\frac{1}{2}+8\cdot\frac{1}{6}=4.$$

and,

$$\sum_{i=1}^{10000} X_i \approx 10,000 \cdot E[X] \approx 40,000.$$

Now we rescale using the central limit theorem:

$$P[\sum_{i=1}^{10000} X_i \leq 39700] \Rightarrow P[\sum_{i=1}^{10000} X_i - 40000 \leq -300].$$

Now, compute the variance:

$$Var(X_1) = E[X_1^2] - E[X_1]^2 = 20 - 16 = 4.$$

So,

$$\sigma \cdot \sqrt{n} = \sqrt{Var(X_1)} \cdot \sqrt{n} = 2 \cdot 100 = 200.$$

and,

$$P[\sum_{i=1}^{10000} X_i - 40000 \leq -300] \Rightarrow P[\frac{\sum_{i=1}^{10000} X_i - 40000}{200} \leq -1.5] \approx P[N(0,1) < -1.5] \approx .0668.$$

using the calculation.

4. now consider the sequence of random variables

$$H_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} = \frac{\sum_{i=1}^{n}(X_i - \mu)}{\sigma\sqrt{n}}.$$

What is the expected value of H_n?

$$E[H_n] = \frac{1}{\sigma\sqrt{n}}[\sum_{i=1}^{n}(E[X_i]) - n\mu] = 0.$$

What is $E[H_n^2]$?

$$\begin{aligned}
E[H_n^2] &= \frac{E[(\sum_{i=1}^{n}(X_i - \mu))^2]}{\sigma^2 n} \\
&= \frac{1}{\sigma^2 n} E[\sum_{i=1,j=1}^{n} (X_i - \mu)(X_j - \mu)] \text{ squaring = all possible pairs} \\
&= \frac{1}{\sigma^2 n} \sum_{i=1,j=1}^{n} E[(X_i - \mu)(X_j - \mu)] \\
&\text{this is the Covariance!} \\
&= \frac{1}{\sigma^2 n} \sum_{i=1,j=1}^{n} Cov(X_i, X_j) \\
&\text{when i = j, same RV so cov = var, otherwise covariance of two independent RV is 0!} \\
&= \frac{1}{\sigma^2 n} \left(\sum_{i=j} Var(X_i) + \sum_{i \neq j} 0 \right) \\
&= 1.
\end{aligned}$$

What about $E[H_n^3]$?

$$
\begin{aligned}
E[H_n^3] &= \frac{1}{\sigma^3 n^{\frac{3}{2}}} \sum_{i,j=1}^{n} E[(X_i - \mu)(X_j - \mu)(X_k - \mu)] \\
&\text{ consider case where } i \neq j \neq k \text{, since } X_i, X_j, X_k \text{ independent} \\
E[(X_i - \mu)(X_j - \mu)(X_k - \mu)] &= E[(X_i - \mu)] \cdot E[(X_j - \mu)] \cdot E[(X_k - \mu)] \\
&= 0 \\
&\text{consider case } i = j \neq k \\
E[(X_i - \mu)^2 (X_k - \mu)] &= E[(X_i - \mu)^2] E[(X_k - \mu)] \\
&= 0 \\
&\text{ via symmetry this applies to } i \neq j = k \text{ and } i = k \neq j \\
&\text{consider case } i = j = k \\
E[(X_i - \mu)^3] &= ?.
\end{aligned}
$$

So,

$$
\begin{aligned}
E[H_n^3] &= \frac{1}{\sigma^3 n^{\frac{3}{2}}} \sum_{i,j=1}^{n} E[(X_i - \mu)(X_j - \mu)(X_k - \mu)] \\
&= \frac{E[(X_i - \mu)^3]}{\sigma^3 \sqrt{n}} && \text{as } n \to \infty \\
&\to 0.
\end{aligned}
$$

where $E[(X_i - \mu)^3]$ is known as the third central moment. What about $E[H_n^4]$?

$$
\begin{aligned}
E[H_n^4] &= \frac{1}{\sigma^4 n^2} \sum_{i,j,k,l} E[(X_i - \mu)(X_j - \mu)(X_k - \mu)(X_l - \mu)] \\
&\text{we know any pairing of } i, j, k, l \text{ with has at least "one on its own" its own contributes 0} \\
&= \frac{1}{\sigma^4 n^2} [3 \cdot n(n-1) \cdot E[(X_1 - \mu)^2 (X_2 - \mu)^2 + nE[(X_1 - \mu)^4]] \\
&\text{since } x_1, x_2 \text{ independent} \\
E[(X_1 - \mu)^2] E[(X_2 - \mu)^2] &= E[(X_1 - \mu)^2] \cdot E[(X_2 - \mu)^2] \\
&= \sigma^4 \\
E[H_n^4] &= \frac{\sigma^4 n(n-1)3}{\sigma^4 n} + \frac{1}{\sigma^4 n} E[(X_1 - \mu)^4] \\
&\text{as } n \to \infty \\
&\to 3.
\end{aligned}
$$

The punchline is:

Moment	H_n	$N(0,1)$
$E[H_n^2]$	1	1
$E[H_n^3]$	$\to 0$	0
$E[H_n^4]$	$\to 3$	3

Table 1: Comparison of Moments: H_n vs $N(0,1)$

4 Moment Generating Function

We need a new tool!

Definition 3. Given a random variable X, the moment generating function of X is:

$$M_x(t) = E[e^{tX}].$$

why is this called a moment generating function?

1. $M_X(0) = E[e^{0X}] = 1 = E[X^0]$
2. $\frac{d}{dt} M_X(t) = E[\frac{d}{dt}\left(e^{tX}\right)] = E[Xe^{tX}]$
$$\frac{d}{dt} M_X(0) = E[X].$$
3. $\frac{d^2}{dt^2}(M_X(t)) = E[X^2 e^{tX}]$
$$\frac{d^2}{dt^2} M_X\,(0) = E[X^2].$$

Another way of viewing this is that the coefficients of the taylor series expansion of M_X at 0 is precisely the moments of X.

Lemma 2. $M_{X+Y}(t) = M_X(t) M_Y(t)$

Proof.

$$\begin{aligned} M_{X+Y}(t) &= E[e^{t(X+Y)}] \\ &= E[e^{tX} e^{tY}] \\ &= E[e^{tX}] E[e^{tY}] \\ &= M_X(t) M_Y(t). \end{aligned}$$

when X, Y are independent ☐

Lemma 3. If c constant,

$$M_{cX}(t) = M_X(ct).$$

Proof.

$$\begin{aligned} M_{cX}(t) &= E[e^{ctX}] \\ &= E[e^{(ct)X}] \\ &= M_X(ct). \end{aligned}$$

☐

Note. Let $X = \begin{cases} 1, P = \frac{1}{2} \\ 0, P = \frac{1}{2} \end{cases}$

$$\begin{aligned} M_X(t) &= E[e^{tX}] \\ &= e^{1 \cdot t} \cdot \frac{1}{2} + e^{-1 \cdot t} \cdot \frac{1}{2} \\ &= \frac{e^t + e^{-t}}{2} \\ &= cosh(t). \end{aligned}$$

$$\frac{d^n}{dt^n} M_X(t) = \begin{cases} \frac{e^t - e^{-t}}{2}, & \text{n is odd} \\ \frac{e^t + e^{-t}}{2}, & \text{n is even} \end{cases} .$$

$$E[X^n] = \begin{cases} 0, & \text{if n is even} \\ 1, & \text{if n is odd} \end{cases} .$$

Note. Let $Y \sim \exp(\lambda)$ so the the pdf of Y is $\lambda e^{\lambda S}, S \geq 0$.

$$\begin{aligned} M_Y(t) &= E[e^{tY}] \\ &= \int_{-\infty}^{\infty} e^{tS} \cdot \lambda e^{-\lambda S} dS \\ &= \lambda \int_0^{\infty} e^{(t-\lambda)S} dS \\ &= \lambda \cdot \lim_{b \to \infty} \int_0^b e^{(t-\lambda)S} dS \\ &= \lambda \cdot \lim_{b \to \infty} [\frac{e^{(t-\lambda)S}}{t-\lambda}]_0^b \\ &= \lambda \lim_{b \to \infty} (\frac{e^{(t-\lambda)b}}{t-\lambda} - \frac{1}{t-\lambda}). \end{aligned}$$

If

$$\begin{cases} t > \lambda, M_Y(t) = \frac{\lambda}{\lambda - t} \\ t \leq \lambda, M_Y(t) = \infty \end{cases} .$$

We can see that:

$$\frac{d}{dt}^n M_Y(t) = \frac{n! \cdot \lambda}{(\lambda - t)^{n+1}} E[(\exp(\lambda))^n] = \frac{n!}{\lambda^n}.$$

Theorem 3. $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-X^2}{2}} dx = I$ i.e. a normal distribution converges.

Proof.

$$\begin{aligned} I^2 &= \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{X^2}{2}} dx \right) \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{Y^2}{2}} dy \right) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{\frac{-(x^2+y^2)}{2}} dxdy \\ &\text{switch to polar} \\ &= \int_0^{2\pi} \int_0^{\infty} \frac{1}{2\pi} e^{\frac{-(x^2+y^2)}{2}} rdrd\theta \\ &= \int_0^{2\pi} \int_0^{\infty} \frac{1}{2\pi} e^{\frac{-r^2}{2}} rdrd\theta \\ &\text{let } u = e^{\frac{-r^2}{2}} \\ &= \int_0^{2\pi} \int_0^{\infty} \frac{1}{2\pi} e^{-u} dud\theta \\ &= \int_0^{2\pi} \frac{1}{2\pi} d\theta \\ &= 1. \end{aligned}$$

Then,

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-X^2}{2}} dx = I = 1.$$

□

Proposition 1. $Z \sim N(0,1)$ what is $M_Z(t) = E[e^{tZ}]$?

Proof. Since Z is normally distributed with mean 0 and variance 1, we can express the expectation as an integral:

$$M_Z(t) = \int_{-\infty}^{\infty} e^{tz} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz$$

This simplifies to:

$$M_Z(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\left(tz - \frac{z^2}{2}\right)}\, dz$$

Now, complete the square in the exponent. The expression inside the exponential is:

$$tz - \frac{z^2}{2} = -\frac{1}{2}\left(z^2 - 2tz\right) = -\frac{1}{2}\left((z-t)^2 - t^2\right)$$

Thus, the integral becomes:

$$M_Z(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\left(-\frac{(z-t)^2}{2} + \frac{t^2}{2}\right)}\, dz$$

Factor out the term that does not depend on z:

$$M_Z(t) = \frac{1}{\sqrt{2\pi}} e^{\left(\frac{t^2}{2}\right)} \int_{-\infty}^{\infty} e^{\left(-\frac{(z-t)^2}{2}\right)}\, dz$$

Since the remaining integral is just the integral of a Gaussian distribution by the previous thereom we have:

$$M_Z(t) = \exp\left(\frac{t^2}{2}\right)$$

Thus, the moment generating function of $Z \sim N(0,1)$ is:

$$M_Z(t) = e^{\left(\frac{t^2}{2}\right)}$$

□

Theorem 4.

1. If $M_X(t) = M_Y(t)$ then X, Y have the same distribution
2. Let $\{X_n\}$ be a sequence of random variables. If $M_{X_n}(t) \to M_Y(t)$ as $n \to \infty$ then $X_n \to Y$ in distirubtion.

Note. Example Let $z_1 + z_2$ be two independent is $N(0,1)$ what is the distribution of $\frac{z_1+z_2}{\sqrt{2}}$.

Proof.

$$Y = \frac{z_1 + z_2}{\sqrt{2}}$$

The moment generating function (MGF) of Y is:

$$M_Y(t) = E\left[e^{tY}\right] = E\left[e^{t\frac{z_1+z_2}{\sqrt{2}}}\right] = E\left[e^{\frac{tz_1}{\sqrt{2}}}\right] \cdot E\left[e^{\frac{tz_2}{\sqrt{2}}}\right]$$

Since z_1 and z_2 are independent $N(0,1)$, their MGF is:

$$M_{z_i}(t) = \exp\left(\frac{t^2}{2}\right)$$

Substituting $t \to \frac{t}{\sqrt{2}}$:

$$M_Y(t) = \exp\left(\frac{t^2}{4}\right) \cdot \exp\left(\frac{t^2}{4}\right) = \exp\left(\frac{t^2}{2}\right)$$

Thus, $M_Y(t)$ is the MGF of $N(0,1)$, so:

$$Y \sim N(0,1)$$

$\square$

Lemma 4. Then $H_n = \frac{\sum_{i=1}^n X_i}{\sqrt{n}} \to N(0,1)$ in distribution in the special case $X_i = \pm 1$ with equal probability and $\{X_i\}$ are iid.

Proof.

$$\begin{aligned} M_{H_n}(t) &= M_{\frac{\sum_{i=1}^n X_i}{\sqrt{n}}}(t) \\ &= \Pi_{i=1}^n M_{\frac{X_i}{\sqrt{n}}}(t) \\ &= \Pi_{i=1}^n M_{X_i}(\frac{t}{\sqrt{n}}) \\ &\text{since } X_i \text{ are iid} \\ &= (M_{X_1}(\frac{t}{\sqrt{n}}))^n. \end{aligned}$$

In our case when $X_1 = \pm 1$ with probability $\frac{1}{2}$, then:

$$\begin{aligned} M_{X_1}(t) &= E[e^{tX_1}] \\ &= \frac{e^t}{2} + \frac{e^{-t}}{2} \\ &= \cosh(t). \end{aligned}$$

Recall that $\cosh(t)$:

$$\begin{aligned} \cosh(t) &= \frac{1}{2}\left(\sum_{n=0}^{\infty} \frac{t^n}{n!}\right) + \frac{1}{2}\sum_{n=0}^{\infty} \frac{(-t)^n}{n!} \\ &= \sum_{n=0}^{\infty} \frac{t^{2n}}{(2n)!}. \end{aligned}$$

Then,

$$\begin{aligned} M_{H_n}(t) &= (\cosh(\frac{t}{\sqrt{n}}))^n \\ &= [1 + \frac{\left(\frac{t}{\sqrt{n}}\right)^2}{2} + \frac{\left(\frac{t}{\sqrt{n}}\right)^4}{4!} + \ldots]^n \end{aligned}$$

.

Take the limit as $n \to \infty$!

$$\cosh(\frac{t}{\sqrt{n}}) \approx [1 + \frac{t^2}{2n} + 0].$$

since the rest of the terms goes to 0 and are much smaller Let $L = \lim_{n=\infty}(1 + \frac{t^2}{2n})$.

$$\begin{aligned} \ln(L) &= \ln[\lim_{n\to\infty}(1 + \frac{t^2}{2n})^n] \\ &= \lim_{n\to\infty} n \cdot \ln(1 + \frac{t^2}{2n})^n \\ &= . \end{aligned}$$

So $\cosh(\frac{t}{\sqrt{n}})^n \to e^{\frac{t^2}{2}} = N(0,1)$! $\square$