

## CONFIDENCE INTERVALS

MAKSIM LEVENTAL

Suppose  $x_1, x_2, \dots, x_n$  are all quantities drawn, without replacement, a “random” sample, from some population, the same population, which itself has an  $\mathcal{N}(\mu, \sigma^2)$  distribution. If the population size is large enough it’s safe to make the approximation that each draw does not affect the population distribution (despite non-replacement) and so effectively each of the  $x_i$  is a draw from an  $\mathcal{N}(\mu, \sigma^2)$  distribution, i.e. each  $x_i$  is actually the random variable  $X_i \sim \mathcal{N}(\mu, \sigma^2)$ . What can we say about the distribution of the random variable  $\bar{X} \equiv \frac{1}{n} \sum_{i=1}^n X_i$ , i.e. the distribution of the “sample mean”? Furthermore what can we say about the distribution of the sample variance  $S^2 \equiv \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ ?

**Theorem 1.**

- I.  $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$
- II.  $\bar{X}$  and  $S^2$  are independent.
- III.  $\frac{(n-1)}{\sigma^2} S^2 \sim \chi_{n-1}^2$

What does this really say? It says that  $\bar{X}$  is distributed with the same distribution and the **same mean as the population you drew from** and that its variance is inversely proportional to the variance of the population by a factor equal to the sample size. That is to say, if we could somehow figure out the mean of  $\bar{X}$ , we would then immediately know the mean of the population. Of course we can’t (philosophically speaking) but we’ll do something else involving the mean that’ll be almost as useful.

**Proof I.** First we prove that the  $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$ : the MGF of  $X_i$

$$MGF_{X_i}(t) = e^{\mu t + \frac{1}{2} \sigma^2 t^2}$$

and therefore by the convolution theorem

$$MGF_{X_1 + \dots + X_n}(t) = \left( e^{\mu t + \frac{1}{2} \sigma^2 t^2} \right)^n = e^{(\mu n)t + \frac{1}{2} (\sqrt{n} \sigma)^2 t^2}$$

Hence  $\sum X_i \sim \mathcal{N}(n\mu, n\sigma^2)$ . Then since the normal distribution is a scale family  $(1/n) \sum X_i \sim \mathcal{N}\left(\frac{1}{n}n\mu, \left(\frac{1}{n}\right)^2 n\sigma^2\right) \sim \mathcal{N}(\mu, \sigma^2/n)$ .  
■

Let’s stop here for a moment see what just this alone buys us. Suppose you had some population for which you knew the variance  $\sigma^2$  but you didn’t know the mean (absurd but suspend disbelief): could you make a guess of the mean and then qualify your guess? Well since we know that  $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$  we know, by

“standardizing”  $\bar{X}$ , that

$$P\left(-1.96 < \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} < 1.96\right) \geq .95$$

After a little algebra (and flipping two inequalities) we have that

$$P\left(\bar{X} - 1.96\sqrt{\frac{\sigma^2}{n}} < \mu < \bar{X} + 1.96\sqrt{\frac{\sigma^2}{n}}\right) \geq .95$$

So with probability .95 if we measure  $\bar{X}$  the interval

$$\left(\bar{X} - 1.96\sqrt{\frac{\sigma^2}{n}}, \bar{X} + 1.96\sqrt{\frac{\sigma^2}{n}}\right)$$

will cover the true mean of the population (remember we said we somehow knew  $\sigma^2$  and  $n$  is just the sample size). For this reason this is called a 95% confidence interval for  $\mu$ . This is also called computing a “z test statistic” since a standard normal random variable is typically represented by a  $Z$ .

Back to the proof of theorem.

**Proof II.** Note that by location-scale it suffices to prove  $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$  for  $\mu = 0$  and  $\sigma^2 = 1$ . To prove independence of  $\bar{X}$  and  $S^2$  we show that they’re functions of independent random vectors. First we write  $S^2$  in terms of  $n - 1$  “deviations”

$$\begin{aligned} S^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \\ &= \frac{1}{n-1} \left( (X_1 - \bar{X})^2 + \sum_{i=2}^n (X_i - \bar{X})^2 \right) \\ &= \frac{1}{n-1} \left( \left[ \sum_{i=2}^n (X_i - \bar{X}) \right]^2 + \sum_{i=2}^n (X_i - \bar{X})^2 \right) \end{aligned}$$

Now solve for the quantity  $(X_1 - \bar{X})$  using  $\sum_{i=1}^n (X_i - \bar{X}) = 0$  and factor out a  $-1$  (which gets killed in the  $\square^2$ ).

$$S^2 = \frac{1}{n-1} \left( \left[ \sum_{i=2}^n (X_i - \bar{X}) \right]^2 + \sum_{i=2}^n (X_i - \bar{X})^2 \right)$$

So  $S^2 = g((X_2 - \bar{X}, X_3 - \bar{X}, \dots, X_n - \bar{X}))$  where  $g$  is clearly continuous. Now we show that the random vector  $(X_2 - \bar{X}, X_3 - \bar{X}, \dots, X_n - \bar{X})$  is independent of  $\bar{X}$ , and by continuity  $S^2$  will be independent of  $\bar{X}$ . The joint pdf of  $(X_1, X_2, \dots, X_n)$  is

$$\begin{aligned} f_X(x_1, \dots, x_n) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_i^2} \\ &= \left( \frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{1}{2} \sum_{i=1}^n x_i^2} \end{aligned}$$

Now let

$$\begin{pmatrix} 1/n & 1/n & \cdots & 1/n \\ -1/n & 1-1/n & -1/n & -1/n \\ \vdots & -1/n & \ddots & -1/n \\ -1/n & -1/n & -1/n & 1-1/n \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

i.e.

$$\begin{aligned} y_1 &= \bar{x} = \frac{1}{n}x_1 + \frac{1}{n}x_2 + \cdots + \frac{1}{n}x_n \\ y_2 &= x_2 - \bar{x} \\ &\vdots \\ y_n &= x_n - \bar{x} \end{aligned}$$

and therefore solve the first equation for  $x_1$  in terms of the  $y_i$  by substituting in  $x_2, x_3, \dots, x_n$

$$\begin{pmatrix} 1 & -1 & \cdots & -1 \\ 1 & 1 & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

Then the Jacobian of the transformation

$$|J| = \left| \begin{pmatrix} 1 & -1 & \cdots & -1 \\ 1 & 1 & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \right| = \left| \begin{pmatrix} n & 0 & \cdots & 0 \\ 1 & 1 & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \right| = n$$

using the invariance of the determinant under linear combinations of rows. Hence

$$\begin{aligned} f_Y(y_1, \dots, y_n) &= f_X\left(y_1 - \sum_{i=2}^n y_i, y_1 + y_2, \dots, y_1 + y_n\right) n \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\left((y_1 - \sum_{i=2}^n y_i)^2 + \sum_{i=2}^n (y_1 + y_i)^2\right)} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\left(y_1^2 - 2y_1 \sum_{i=2}^n y_i + \left(\sum_{i=2}^n y_i\right)^2 + \sum_{i=2}^n y_1^2 + 2y_1 \sum_{i=2}^n y_i + \sum_{i=2}^n y_i^2\right)} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\left(y_1^2 + \sum_{i=2}^n y_1^2 + \sum_{i=2}^n y_i^2 + \left(\sum_{i=2}^n y_i\right)^2\right)} \\ &= \left(\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}ny_1^2}\right) \left(e^{-\frac{1}{2}\left(\sum_{i=2}^n y_i^2 + \left(\sum_{i=2}^n y_i\right)^2\right)}\right) \end{aligned}$$

and so  $Y_1 \equiv \bar{X}$  is independent of  $(Y_2, Y_3, \dots, Y_n) \equiv (X_2 - \bar{X}, X_3 - \bar{X}, \dots, X_n - \bar{X})$   
and so  $\bar{X}$  is independent of  $S^2 = g((X_2 - \bar{X}, X_3 - \bar{X}, \dots, X_n - \bar{X}))$ .

■

This is kind of amazing in and of itself, that  $S^2 = h(\bar{X})$  is statistically speaking independent of  $\bar{X}$  but it's also useful and why will be apparent after the completion of the proof, i.e. establishing that  $\frac{(n-1)}{\sigma^2} S^2 \sim \chi_{n-1}^2$ .

**Proof III.** There are two ways to do this. First

$$\begin{aligned}
 \sum_{i=1}^n (X_i)^2 &= \sum_{i=1}^n ((X_i - \bar{X}) + \bar{X})^2 \\
 &= \sum_{i=1}^n (X_i - \bar{X})^2 + 2(\bar{X})^2 \sum_{i=1}^n (X_i - \bar{X}) + \sum_{i=1}^n (\bar{X})^2 \\
 &= \sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^n (\bar{X})^2 \\
 &= (n-1)S^2 + n(\bar{X})^2
 \end{aligned}$$

and by almost exactly the same manipulation (the “cross” term will be a little different but it’s irrelevant since it goes to 0)

$$\sum_{i=1}^n (X_i - \mu)^2 = (n-1)S^2 + n(\bar{X} - \mu)^2$$

but then (if you’re allowed to do this????)

$$\begin{aligned}
 \sum_{i=1}^n \left( \frac{X_i - \mu}{\sigma} \right)^2 &= \frac{1}{\sigma^2} \left( (n-1)S^2 + n(\bar{X} - \mu)^2 \right) \\
 &= \frac{(n-1)}{\sigma^2} S^2 + \left( \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2
 \end{aligned}$$

Now again by convolution theorem and independence of  $S^2$  and  $\bar{X}$  we have

$$MGF_{\sum_{i=1}^n \left( \frac{X_i - \mu}{\sigma} \right)^2} = MGF_{\frac{(n-1)}{\sigma^2} S^2} MGF_{\left( \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2}$$

But  $\sum_{i=1}^n \left( \frac{X_i - \mu}{\sigma} \right)^2 \sim \chi_{n-1}^2$  and  $\left( \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2 \sim \chi_1^2$  so

$$\left( \frac{1}{1-2t} \right)^{n/2} = MGF_{\frac{(n-1)}{\sigma^2} S^2} \left( \frac{1}{1-2t} \right)^{1/2}$$

and so

$$\begin{aligned}
 MGF_{\frac{(n-1)}{\sigma^2} S^2} &= \left( \frac{1}{1-2t} \right)^{n/2} \bigg/ \left( \frac{1}{1-2t} \right)^{1/2} \\
 &= \left( \frac{1}{1-2t} \right)^{\frac{n-1}{2}}
 \end{aligned}$$

and hence  $\frac{(n-1)}{\sigma^2} S^2 \sim \chi_{n-1}^2$ . But doesn't this directly imply that  $\sum_{i=1}^n \left( \frac{X_i - \mu}{\sigma} \right)^2$  and  $n \left( \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2$  doesn't it?

$$\begin{aligned} \frac{(n-1)}{\sigma^2} S^2 &= \frac{1}{\sigma^2} \left( \sum_{i=1}^n \left( \frac{X_i - \mu}{\sigma} \right)^2 - \left( \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2 \right) \\ &= \sum_{i=1}^n \left( \frac{X_i - \mu}{\sigma} \right)^2 - n \left( \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2 \end{aligned}$$

The second way is to use induction with  $\bar{X}_k$  and  $S_k^2$  being the sample mean and variance for a sample of size  $k$ . The base case

$$\begin{aligned} S_2^2 &= \frac{1}{2-1} \left( \left( X_1 - \frac{1}{2}(X_1 + X_2) \right)^2 + \left( X_2 - \frac{1}{2}(X_1 + X_2) \right)^2 \right) \\ &= \frac{1}{2} (X_1 - X_2)^2 \end{aligned}$$

and if  $X_1 \sim X_2 \sim \mathcal{N}(\mu, \sigma^2)$  then  $X_1 - X_2 \sim \mathcal{N}(0, 2\sigma^2)$  and  $X_1 - X_2 / (\sqrt{2}\sigma) \sim \mathcal{N}(0, 1)$  so

$$\begin{aligned} \frac{(2-1)}{\sigma^2} S_2^2 &= \frac{1}{2\sigma^2} (X_1 - X_2)^2 \\ &= \left( \frac{X_1 - X_2}{\sqrt{2}\sigma} \right)^2 \end{aligned}$$

hence  $\frac{(2-1)}{\sigma^2} S^2 \sim \chi_2^2$ . To finish the induction we have to prove that  $\frac{(n-1)}{\sigma^2} S_n^2 \sim \chi_{n-1}^2$  if  $\frac{(n-2)}{\sigma^2} S_{n-1}^2 \sim \chi_{n-2}^2$  (remember the subscript on the  $S^2$  is the subscript necessary for induction). Now the induction step requires

$$\frac{(n-1)}{\sigma^2} S_n^2 = \frac{(n-2)}{\sigma^2} S_{n-1}^2 + \left( \frac{n-1}{\sigma^2 n} \right) (X_n - \bar{X}_{n-1})^2$$

which is true because of a very length algebraic manipulation. Now observe that the random vector  $(X_n, \bar{X}_{n-1})$  is independent of  $S_{n-1}^2$  since as before  $\bar{X}_{n-1}$  is independent of  $S_{n-1}^2$  and  $X_n$  is definitely independent of  $S_{n-1}^2$  since  $X_n$  is “in the future”. And finally since  $X_n - \bar{X}_{n-1} \sim \mathcal{N}(0, n\sigma^2/(n-1))$  and by the induction hypothesis

$$\frac{(n-1)}{\sigma^2} S_n^2 = \chi_{n-2}^2 + \chi_1^2 = \chi_{n-1}^2$$

■

Now what does *this* buy us? Well as mentioned, in general you have no idea what  $\sigma^2$  for a population is so computing

$$P \left( \bar{X} - 1.96 \sqrt{\frac{\sigma^2}{n}} < \mu < \bar{X} + 1.96 \sqrt{\frac{\sigma^2}{n}} \right)$$

i.e.

$$P \left( -1.96 < \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} < 1.96 \right)$$

is impossible. But what about

$$P\left(a < \frac{\bar{X} - \mu}{\sqrt{S^2/n}} < b\right)$$

Well

$$T = \frac{\bar{X} - \mu}{\sqrt{S^2/n}} = \frac{(\bar{X} - \mu) \frac{\sqrt{n}}{\sigma}}{\sqrt{\frac{(n-1)S^2/\sigma^2}{(n-1)}}} = \frac{Z}{\sqrt{V/v}}$$

where  $v = n - 1$ ,  $Z \sim \mathcal{N}(0, 1)$ , and  $V \sim \chi_{n-1}^2$  and  $V$  is independent of  $Z$  since yet again  $S^2$  is independent of  $\bar{X}$ . This is the exact form of a Student's  $t$  distribution with  $v = n - 1$  degrees of freedom, i.e.  $T \sim T(v) \sim T(n - 1)$ . The density is

$$f_T(t) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\Gamma\left(\frac{v}{2}\right)} \left(1 + \frac{t^2}{v}\right)^{-\frac{v+1}{2}}$$

and so now we can compute “ $t$  test statistics”. Let  $a$  be such that  $P(-a < T < a) = .90$  then

$$P\left(-a < \frac{\bar{X} - \mu}{\sqrt{S^2/n}} < a\right) = P\left(\bar{X} - a\sqrt{\frac{S^2}{n}} < \mu < \bar{X} + a\sqrt{\frac{S^2}{n}}\right) = .90$$

and so if we draw from our population the interval  $\left(\bar{X} - a\sqrt{\frac{S^2}{n}}, \bar{X} + a\sqrt{\frac{S^2}{n}}\right)$  will “cover” the true mean with probability .9, i.e. it is a 90% confidence interval for  $\mu$ . Quoting wikipedia:

“Therefore, if we find the mean of a set of observations that we can reasonably expect to have a normal distribution, we can use the  $t$ -distribution to examine whether the confidence limits on that mean include some theoretically predicted value – such as the value predicted on a null hypothesis.”

### Central Limit theorem

The central limit theorem is magical.

**Theorem 2.** If  $X_i$  are all i.i.d with  $E(X_i) = \mu$  and  $\text{Var}(X_i) = \sigma^2$  and let

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)$$

the sampling distribution and

$$T_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$$

Then  $\lim_{n \rightarrow \infty} F_{T_n}(t) = F_Z(t)$  where  $Z \sim \mathcal{N}(0, 1)$ . That is to say  $T_n$  converges in distribution to a standard normal.

Think about this. Regardless of the random variables (no matter what they themselves are distributed as), as long as their mean and variance exist, after centering, their properly normalized sum will converge to a normal distribution.

The proof will use the characteristic function (fourier transform) and so we need a couple of lemmas. The first one has to do with the “smoothness” of characteristic functions

**Lemma 1.** For real  $t$

$$\left| e^{it} - \sum_{k=0}^n \frac{(it)^k}{k!} \right| \leq \min \left\{ \frac{|t|^{n+1}}{(n+1)!}, \frac{2|t|^n}{n!} \right\}$$

Afterward this result will be used to show that if  $E(X) = 0$  and  $\text{Var}(X) = \sigma^2 < \infty$  then its characteristic function  $\phi_X(t)$ , to second order, is

$$\phi(t) = 1 - \frac{t^2 \sigma^2}{2} + o(t^2)$$

**Proof Lemma 1.** Clearly this has something to do with the Taylor series for  $e^{it}$ . So expand  $e^{it}$

$$e^{it} = \sum_{k=0}^n \frac{(it)^k}{k!} + \frac{i^{n+1}}{(n)!} \int_0^t (t-s)^n e^{is} ds$$

Then

$$\left| e^{it} - \sum_{k=0}^n \frac{(it)^k}{k!} \right| = \left| \frac{i^{n+1}}{(n)!} \int_0^t (t-s)^n e^{is} ds \right|$$

We estimate

$$\left| \frac{i^{n+1}}{(n)!} \int_0^t (t-s)^n e^{is} ds \right|$$

in two different ways; one will be better when  $|t|$  is small and the other when  $|t|$  is large. First the obvious thing is

$$\begin{aligned} \left| e^{it} - \sum_{k=0}^n \frac{(it)^k}{k!} \right| &= \left| \frac{i^{n+1}}{(n)!} \int_0^t (t-s)^n e^{is} ds \right| \\ &= \frac{1}{(n)!} \left| \int_0^t (t-s)^n e^{is} ds \right| \\ &\leq \frac{1}{(n)!} \int_0^t |(t-s)^n e^{is}| ds \\ &= \frac{1}{(n)!} \int_0^t |(t-s)^n| ds \\ &\leq \frac{|t|^{n+1}}{(n+1)!} \end{aligned}$$

Alternatively since by integration by parts

$$\int_0^t (t-s)^{n-1} e^{is} ds = \frac{i}{n} \int_0^t (t-s)^n e^{is} ds + \frac{t^n}{n}$$

i.e.

$$\begin{aligned}\frac{i}{n} \int_0^t (t-s)^n e^{is} ds &= \int_0^t (t-s)^{n-1} e^{is} ds - \frac{t^n}{n} \\ \int_0^t (t-s)^n e^{is} ds &= \frac{n}{i} \left( \int_0^t (t-s)^{n-1} e^{is} ds - \frac{t^n}{n} \right)\end{aligned}$$

and so it's the case that

$$\begin{aligned}\left| e^{it} - \sum_{k=0}^n \frac{(it)^k}{k!} \right| &= \left| \frac{i^{n+1}}{(n)!} \int_0^t (t-s)^n e^{is} ds \right| \\ &= \left| \frac{i^{n+1}}{(n)!} \frac{n}{i} \left( \int_0^t (t-s)^{n-1} e^{is} ds - \frac{t^n}{n} \right) \right| \\ &= \frac{1}{(n-1)!} \left| \left( \int_0^t (t-s)^{n-1} e^{is} ds - \frac{t^n}{n} \right) \right| \\ &\leq \frac{1}{(n-1)!} \left( \frac{|t|^n}{n} + \frac{|t|^n}{n} \right) = \frac{2|t|^n}{n!}\end{aligned}$$

■

Note that for large  $|t|$  the tighter estimate is  $\frac{2|t|^n}{n!}$  (i.e. it's smaller than  $\frac{|t|^{n+1}}{(n+1)!}$ ) and for small  $|t|$  the tighter estimate is  $\frac{|t|^{n+1}}{(n+1)!}$ .

It's a short hop skip and a jump away from the result of Lemma 1 to this

**Lemma 2.** If  $E(|X|^n) < \infty$  then its characteristic function is  $n$  times continuously differentiable with derivative

$$\phi^{(k)}(t) = i^k E(X^k e^{itX})$$

and its Taylor series expansion is bounded

$$\left| \phi(t) - \sum_{k=0}^n E(X^k) \frac{(it)^k}{k!} \right| \leq E \left( \min \left\{ \frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!} \right\} \right)$$

and therefore  $o(|t|^n)$  as  $t \rightarrow 0$ . Then in particular if  $E(X) = 0$  and  $\text{Var}(X) = \sigma^2 < \infty$

$$\phi(t) = 1 - \frac{t^2 \sigma^2}{2} + o(t^2)$$

**Proof Lemma 2.** Firstly

$$\frac{\phi(t+h) - \phi(t)}{h} = E \left( \frac{e^{i(t+h)X} - e^{itX}}{h} \right) = E \left( \frac{e^{ihX} - 1}{h} e^{itX} \right)$$

and by Taylor's theorem  $h \rightarrow 0$  (without worry about negativity since  $h > 0$ )

$$\lim_{h \rightarrow 0} \frac{e^{ihX} - 1}{h} e^{itX} = \lim_{h \rightarrow 0} \frac{ihX}{h} e^{itX} = iX e^{itX}$$

and by lemma 2 for all  $h$

$$\left| \frac{e^{ihX} - 1}{h} e^{itX} \right| \leq |iX e^{itX}| \leq |X|$$



Since  $E(|X|^n) < \infty$  all moments of  $E(|X|^j)$  for  $j < n$  exist as well, i.e. all  $|X|^j$  are integrable, in particular  $|X|$  and so by the dominated convergence theorem

$$\phi'(t) = \lim_{h \rightarrow 0} \frac{\phi(t+h) - \phi(t)}{h} = \lim_{h \rightarrow 0} E\left(\frac{e^{i(t+h)X} - e^{itX}}{h}\right) = E(iX e^{itX})$$

which by induction (and by similar arguments, i.e.  $|X^k e^{itX}| \leq |X|^k$ ) proves that

$$\phi^{(k)}(t) = i^k E(X^k e^{itX})$$

Then by Jensen's inequality

$$\begin{aligned} \left| \phi(t) - \sum_{k=0}^n E(X^k) \frac{(it)^k}{k!} \right| &= \left| E\left(e^{itX} - \sum_{k=0}^n \frac{(itX)^k}{k!}\right) \right| \\ &\leq E\left(\left|e^{itX} - \sum_{k=0}^n \frac{(itX)^k}{k!}\right|\right) \\ &\leq E\left(\min\left\{\frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!}\right\}\right) \end{aligned}$$

Then by dominated convergence again, as  $t \rightarrow 0$  (why???)

$$|t|^n E(|X|^n \min\{|tX|, 2\}) = o(|t|^n)$$

and so, as  $t \rightarrow 0$  (does that really meet the def of little o?)

$$\begin{aligned} \phi(t) &\leq \sum_{k=0}^{n-1} E(X^k) \frac{(it)^k}{k!} + E\left(\min\left\{\frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!}\right\}\right) \\ &= 1 + \frac{(it)}{1} E(X) + \frac{(it)^2}{2!} E(X^2) + E\left(\min\left\{\frac{|tX|^{3+1}}{(4)!}, \frac{2|tX|^3}{3!}\right\}\right) \\ &= 1 - \frac{t^2 \sigma^2}{2!} + \frac{|t|^3}{3!} E\left(|X|^3 \min\left\{\frac{|tX|}{4}, 2\right\}\right) \\ &= 1 - \frac{t^2 \sigma^2}{2!} + o(|t|^2) \end{aligned}$$

■

**Tiny Lemma 3:** For  $a_i$  such that  $\lim_{i \rightarrow \infty} a_i = a$

$$\lim_{n \rightarrow \infty} \left(1 + \frac{a_n}{n}\right)^n = e^a$$

**Proof Tiny Lemma 3:** TODO

Finally we can prove Theorem 2, the central limit theorem

**Proof Theorem 2.** Define  $Y_i = (X_i - \mu)/\sigma$  and let  $\phi_Y(t)$  be the characteristic function of the  $Y_i$ . Then

$$S_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i$$

and then

$$\begin{aligned}\phi_{\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}}(t) &= \phi_{\frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i}(t) \\ &= \phi_{\sum_{i=1}^n Y_i}\left(\frac{t}{\sqrt{n}}\right) \\ &= \left(\phi_Y\left(\frac{t}{\sqrt{n}}\right)\right)^n\end{aligned}$$

and by Lemma 2

$$\begin{aligned}\left(\phi_Y\left(\frac{t}{\sqrt{n}}\right)\right)^n &= \left(1 - \frac{t^2}{2n} + o\left(\left(\frac{t}{\sqrt{n}}\right)^2\right)\right)^n \\ &= \left(1 - \frac{1}{n}\left(\frac{t^2}{2} + n \cdot o\left(\frac{t^2}{n}\right)\right)\right)^n\end{aligned}$$

Now  $o\left(\left(\frac{t}{\sqrt{n}}\right)^2\right) = o\left(\frac{t^2}{n}\right)$  means there exists  $f\left(\frac{t^2}{n}\right)$  such that

$$\frac{t^2}{2} + n \cdot o\left(\frac{t^2}{n}\right) = \frac{t^2}{2} + n \cdot f\left(\frac{t^2}{n}\right)$$

and

$$\lim_{n \rightarrow \infty} \frac{f\left(\frac{t^2}{n}\right)}{t^2/n} = \lim_{n \rightarrow \infty} n \cdot \frac{f\left(\frac{t^2}{n}\right)}{t^2} = 0$$

therefore by tiny lemma 3

$$\begin{aligned}\lim_{n \rightarrow \infty} \left(\phi_Y\left(\frac{t}{\sqrt{n}}\right)\right)^n &= \lim_{n \rightarrow \infty} \left(1 - \frac{1}{n}\left(\frac{t^2}{2} + n \cdot o\left(\frac{t^2}{n}\right)\right)\right)^n \\ &= e^{-t^2/2}\end{aligned}$$

which is the characteristic function of  $\mathcal{N}(0, 1)$ .

■

Okay finally what's the point? What does the central limit buy us? Well for example suppose we were drawing iid from a population but all we know about the population is that the distribution that describes it has a finite mean  $\mu$  and finite variance  $\sigma^2$  but we're again in the same position we were that led us to Student's  $t$ . So not only do we not know the distribution of  $\bar{X}_n$  but we don't know  $\sigma^2$  either. We can then use the central limit theorem give an approximate confidence interval (a confidence confidence interval?). Since

$$\lim_{n \rightarrow \infty} \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \sim \mathcal{N}(0, 1)$$

we might be able to claim that

$$\lim_{n \rightarrow \infty} \frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} = \lim_{n \rightarrow \infty} \frac{\sigma}{S_n} \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \sim \mathcal{N}(0, 1)$$

and then we would we have that for a large enough  $n$

$$\begin{aligned} \lim_{n \rightarrow \infty} P \left( -1.96 < \frac{\sqrt{n} (\bar{X}_n - \mu)}{S_n} < 1.96 \right) &= \lim_{n \rightarrow \infty} P(-1.96 < Z < 1.96) \\ &\approx P(-1.96 < Z < 1.96) \\ &\gtrapprox .95 \end{aligned}$$

This depends on four results, one of which is the Strong Law of Large numbers.

**Theorem 3 (Kolmogorow Strong Law of Large numbers):** Let  $X_i$  be iid random variables with  $E(X_i) = \mu$ . Then

$$\bar{X}_n \xrightarrow[a.s.]{} \mu$$

**Proof:** Omitted

**Corollary:** For  $X_i$  iid with finite mean and variance the sample variance is a consistent estimator of the true variance, i.e.

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \xrightarrow[a.s.]{} \sigma^2$$

**Proof of corollary:** By Kolmogorov SLLN and the sample manipulation as in theorem 1

$$\begin{aligned} \lim_{n \rightarrow \infty} S_n^2 &= \lim_{n \rightarrow \infty} \frac{1}{n-1} \left( \sum_{i=1}^n (X_i)^2 - n (\bar{X}_n)^2 \right) \\ &= \lim_{n \rightarrow \infty} \frac{n}{n-1} \left( \frac{\sum_{i=1}^n (X_i)^2}{n} - (\bar{X}_n)^2 \right) \\ &= 1 \cdot (E(X_i^2) - \mu^2) = \sigma^2 \end{aligned}$$

**Lemma 5:**  $\sigma/S_n \xrightarrow[a.s.]{} 1$

**Proof Lemma 5:** Since continuous function of a sequence of random variables that converge to a limit itself converges to the function evaluated at that limit we have by the corollary above that  $\sqrt{S_n^2} \xrightarrow[a.s.]{} \sqrt{\sigma^2}$  and therefore  $\sigma/S_n \xrightarrow[a.s.]{} 1$ .

**Lemma 6 (Slutsky's Theorem):** If  $U_n \xrightarrow{D} U$  and  $Y_n \xrightarrow{P} a$  then  $Y_n U_n \xrightarrow{D} aU$ .

**Proof:** Omitted

Finally letting  $Y_n = \sigma/S_n$  and  $U_n = \sqrt{n} (\bar{X}_n - \mu) / \sigma$  we have

$$\lim_{n \rightarrow \infty} \frac{\sqrt{n} (\bar{X}_n - \mu)}{S_n} = \lim_{n \rightarrow \infty} \frac{\sigma}{S_n} \frac{\sqrt{n} (\bar{X}_n - \mu)}{\sigma} \sim \mathcal{N}(0, 1)$$