Lectures 13 and 14: From Samples to Populations

Content:

- Premises of the Law of Large Numbers
- Markov's inequality
- Chebyshev's inequality
- Proof of the Law of Large Numbers

Last words...

- Right now we work with samples of larger populations of data
- We measure properties of samples, like mean, standard deviation, covariance, correlation coefficient
- All these properties are also random variable and have a distribution
- Our question is therefore, what kind of distribution is the one of the correlation coefficient
- Knowing its distribution allows us to understand the relationships existing between the variables it connect

Knowing the sample . . .

- What can we infer of populations now that I know the properties of the sample?
- Now we know the mean, the standard deviation, the distribution of the sample, what would be the mean, the standard deviation, and the distribution of the population?
- Moreover, from two samples we can build a regression, what would be the regression of the population?

We start from the mean

- Now we start from the mean
- We suppose that we have an unknown population $\mathfrak{P}$ of entities on a ratio scale from which we extract n samples $\mathfrak{S}_i$ with $i \in [1 \dots n]$
- Each sample i is composed by $\mathfrak{n}_\mathfrak{i}$ elements $e_{i,j}$ with $j \in [1 \dots \mathfrak{n}_\mathfrak{i}]$
- We can compute the set of the means of each sample $\mathfrak{S}_i$, $\overline{\mathfrak{m}_i}$ with $i \in [1 \dots n]$
- $\overline{\mathfrak{m}_i}$ is a random variable, so we would like to know what is its structure
- There are two fundamental theorems in statistics that provide the distributions of such $\overline{\mathfrak{m}_i}$, the Law of Large Numbers (LLN) and the Central Limit Theorem (CLT)
- Since we are not making **any** assumption on the population Π, we can just ignore it and consider simply a sequence of random variables, which we will call x_i assuming that there is always a set that include them, which is indeed true in algebra.

LLN – Premises

- From now on, we will use the notation "iid" to denote the property of a set of random variables to be **i**ndependent and **i**dentically **d**istributed
- **Let $\{\mathfrak{Xn}_1, \mathfrak{Xn}_2, \ldots, \mathfrak{Xn}_n\}$ a set of n iid random variables drawn from a population with mean μ**
- Each $\mathfrak{Xn}_i$ could be considered the average of a sample $\mathfrak{S}_i$ of size 1, that is $\mathfrak{S}_i = \{\mathfrak{Xn}_i\}$
- **Let us consider $\overline{\mathfrak{Xn}}$, the average for this sample of size n**
- $\overline{\mathfrak{Xn}}$ is like the average of the n averages of each sample $\mathfrak{S}_i$

Source with modifications:

https://en.wikipedia.org/wiki/Law_of_large_numbers

LLN – Weak formulation

- Let $\{\mathfrak{Xn}_1, \mathfrak{Xn}_2, \ldots, \mathfrak{Xn}_n\}$ a set of n iid random variables drawn from a population with mean μ
- Let us consider $\overline{\mathfrak{Xn}}$, the average for this sample of size n
- the Law of Large Number in its weak formulation states that:

$$(\forall \epsilon \in \mathbb{R}^+) \quad \lim_{n \to \infty} \mathbb{P}(|\overline{\mathfrak{Xn}} - \mu| > \epsilon) = 0$$

- This means that $\overline{\mathfrak{Xn}}$ tends to get the value of μ probabilistically

Source with modifications:
https://en.wikipedia.org/wiki/Law_of_large_numbers

LLN – Proof (1/4)

- We are now going to prove LLN
- To do so, we need to prove two other interesting theorems:
  - The Markov's inequality
  - The Chebyshev's inequality

Source with modifications:

https://en.wikipedia.org/wiki/Law_of_large_numbers

[LLN – Proof] Markov's inequality (1/3)

- The Markov's inequality put a first boundary on the distribution of a random variable
- Let $X \geq 0$ be a random variable with mean $\mu \in \mathbb{R}$
- Then:

$$(\forall k \in \mathbb{R}^+) \;\; \mathbb{P}(X \geq k) \leq \frac{\mu}{k}$$

- Proof:

$$\mu = \int_{-\infty}^{+\infty} x f_x(x) dx$$

Source with modifications:
https://en.wikipedia.org/wiki/Markov%27s_inequality

[LLN – Proof] Markov's inequality (2/3)

- Since $X \geq 0$

$$\int_{-\infty}^{+\infty} x f_x(x) dx = \int_{0}^{+\infty} x f_x(x) dx =$$

And given $k \in \mathbb{R}^+$

$$= \int_{0}^{k} x f_x(x) dx + \int_{k}^{+\infty} x f_x(x) dx$$

Since $\int_0^k x f_x(x) dx \geq 0$

$$\mu \geq \int_{k}^{+\infty} x f_x(x) dx \geq k \int_{k}^{+\infty} f_x(x) dx = \mathbb{P}(X \geq k)$$

Source with modifications:
https://en.wikipedia.org/wiki/Markov%27s_inequality

[LLN – Proof] Markov's inequality (3/3)

- Therefore we have

$$\mu \geq k\mathbb{P}(X \geq k)$$

- And from this we conclude:

$$\mathbb{P}(X \geq k) \leq \frac{\mu}{k}$$

Source with modifications:
https://en.wikipedia.org/wiki/Markov%27s_inequality

[LLN – Proof] Chebyshev's inequality (1/3)

- The Chebyshev's inequality put a further limit on the distribution of a random variable
- Let X be a random variable with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$
- Then:

$$(\forall k \in \mathbb{R}^+) \;\; \mathbb{P}(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

- Proof:
  Let us define a new random variable

$$Y = (X - \mu)^2 \geq 0$$

  Let us define

$$h = (k\sigma)^2$$

Source with modifications:
https://en.wikipedia.org/wiki/Chebyshev%27s_inequality

[LLN – Proof] Chebyshev's inequality (2/3)

- By the Markov inequality we have for the nonnegative random variable Y and for the positive real h:

$$\mathbb{P}(Y \geq h) \leq \frac{\overline{Y}}{h}$$

- And this means:

$$\mathbb{P}((X-\mu)^2 \geq (k\sigma)^2) \leq \frac{\overline{(X-\mu)^2}}{(k\sigma)^2} = \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}$$

Source with modifications:
https://en.wikipedia.org/wiki/Chebyshev%27s_inequality

[LLN – Proof] Chebyshev's inequality (3/3)

- This can be rewritten into:

$$\mathbb{P}(|X-\mu| \geq |k\sigma|) \leq \frac{1}{k^2}$$

- Since we know that both k and σ are strictly positive:

$$\mathbb{P}(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

QED

Source with modifications:
https://en.wikipedia.org/wiki/Chebyshev%27s_inequality

LLN – Proof (2/4)

- We want to prove that:

$$(\forall \epsilon \in \mathbb{R}^+) \quad \lim_{n \to \infty} \mathbb{P}(|\overline{\mathfrak{X}\mathfrak{n}} - \mu| > \epsilon) = 0$$

- we add the additional hypothesis that $\sigma_i > 0$
- Let us consider σ_i;
  - since the variables $\mathfrak{X}\mathfrak{n}_i$ are iid

$$(\forall i, j)\ (\sigma_i = \sigma_j = \sigma)$$

  - we also assume that $\sigma > 0$
  - finally, since the variables $\mathfrak{X}\mathfrak{n}_i$ are independent of one another:

$$Var(\overline{\mathfrak{X}\mathfrak{n}}) = \frac{\sigma^2}{n} = \mathfrak{s}_{\mathfrak{n}}^2$$

Source with modifications:
https://en.wikipedia.org/wiki/Chebyshev%27s_inequality

LLN – Proof (3/4)

- Let us define:

$$k = \frac{\epsilon}{\mathfrak{s_n}}$$

k exists, since $\mathfrak{s_n}$ is strictly positive; therefore:

$$\epsilon = k\mathfrak{s_n}$$

- By Chebyshev's inequality we have:

$$\mathbb{P}(|\overline{\mathfrak{Xn}} - \mu| \geq k\mathfrak{s_n}) \leq \frac{1}{k^2}$$

- That is:

$$\mathbb{P}(|\overline{\mathfrak{Xn}} - \mu| \geq \epsilon) \leq \frac{\mathfrak{s_n}^2}{\epsilon^2}$$

Source with modifications:
https://en.wikipedia.org/wiki/Chebyshev%27s_inequality

LLN – Proof (4/4)

- Since:

$$\mathfrak{s}_{\mathfrak{n}}^2 = \frac{\sigma^2}{n}$$

- We have that

$$\lim_{n\to\infty} \frac{\mathfrak{s}_{\mathfrak{n}}^2}{\epsilon^2} = \lim_{n\to\infty} \frac{\sigma^2}{n\epsilon^2} = \frac{\sigma^2}{\epsilon^2} \lim_{n\to\infty} \frac{1}{n} = 0$$

- Therefore:

$$\lim_{n\to\infty} \left(\mathbb{P}(|\overline{\mathfrak{X}\mathfrak{n}} - \mu| \geq \epsilon)\right) \leq \lim_{n\to\infty} \frac{\mathfrak{s}_{\mathfrak{n}}^2}{\epsilon^2} = 0 \Rightarrow \lim_{n\to\infty} \left(\mathbb{P}(|\overline{\mathfrak{X}\mathfrak{n}} - \mu| \geq \epsilon)\right) = 0$$

QED

Source with modifications:

https://en.wikipedia.org/wiki/Chebyshev%27s_inequality

Content:

- Central Limit Theorem in the Linderberg-Lévy formulation
- Moment
- Moment generating function
- Proof of the Central Limit Theorem in the Linderberg-Lévy formulation
- Final comment

CLT – Lindeberg–Lévy formulation

- Let $\{\mathfrak{Xn}_1, \mathfrak{Xn}_2, \ldots, \mathfrak{Xn}_n\}$ a set of n iid random variables drawn from a population with mean μ and standard deviation σ
- Let us consider for this sample of size n:
  – the mean, $\overline{\mathfrak{Xn}}$
  – the variance, σ^2
  – the modulated difference, $\mathfrak{Dn}$, defined as:

$$\mathfrak{Dn} = \sqrt{n}(\overline{\mathfrak{Xn}} - \mu)$$

- Central Limit Theorem (Lindeberg–Lévy formulation):

$$\mathfrak{Dn} \xrightarrow{d} N(0, \sigma^2)$$

- This means that $\mathfrak{Dn}$ tends to be normal.

Source with modifications:
https://en.wikipedia.org/wiki/Central_limit_theorem

[CLT – LLf] Moment (1/2)

- To prove the CLT – LLf we need to introduce a few additional statistical concepts that could be useful also in the continuation of this course series
- We define the r^{th} **moment** of a random variable X as the expected value of the r^{th} power of X; formally:

$$\mu_X(r) = E(X^r)$$

  clearly: $\mu_X(1) = \mu_X = E(X)$
- Example:
  - If $P(X = 0) = 0.25$ and $P(X = 4) = 0.75$:
    $\mu_X(1) = 3$
    $\mu_X(2) = 12$
    $\mu_X(3) = 48$
    $\mu_X(4) = 192$

Source with modifications:

https://www.statlect.com/fundamentals-of-probability/moments

[CLT – LLf] Moment (2/2)

- We define the **central** r^{th} **moment** of a random variable X as the expected value of the r^{th} deviation of X; formally:

$$\overline{\mu_X(r)} = E((X - \mu_X)^r)$$

clearly: $\overline{\mu_X(2)} = \sigma_X^2 = E((X - \mu_X)^2)$

- Example:
  - If $P(X = 0) = 0.25$ and $P(X = 4) = 0.75$:
    $\overline{\mu_X(1)} = 0$
    $\overline{\mu_X(2)} = 3$
    $\overline{\mu_X(3)} = -6$
    $\overline{\mu_X(4)} = 21$

Source with modifications:

https://www.statlect.com/fundamentals-of-probability/moments

- Let X be a random variable defined over a set S and let f_X be its probability density function
- We define the **moment generating function** (**mgf**) M_X over X as:

$$M_X(t) = E(e^{tX}) = \int_S e^{tx} f_X(x)dx$$

  if there exists $h \in \mathbb{R}^+$ so that $E(e^{tX})$ is defined in $(-h, +h)$
- Note that:
    - The mgf may not exist
    - The mgf has interesting properties

Source with modifications:
https://onlinecourses.science.psu.edu/stat414/node/72/

- Mgf and first moment:

$$\left[\frac{dM_X(t)}{dt}\right](t=0) = \mu_X(1) = \mu_X = E(X)$$

Since:

$$\left[\frac{dM_X(t)}{dt}\right](t=0) = \left[\frac{d\int_S e^{tx} f_X(x)dx}{dt}\right](t=0) =$$

$$= \left[\int_S x e^{tx} f_X(x)dx\right](t=0) = \int_S x e^{0x} f_X(x)dx = \int_S x f_X(x)dx = \mu_X$$

Source with modifications:

https://onlinecourses.science.psu.edu/stat414/node/73/

- In general:

$$\left[\frac{d^n M_X(t)}{dt^n}\right](t=0) = \mu_X(n) = E(X^n)$$

- This comes from:

$$\frac{d^n M_X(t)}{dt^n} = \int_S x^n e^{tx} f_X(x)dx$$

- Proof. By induction, n=1, see above
- Let us assume that the proposition holds for n-1:

$$\frac{d^{n-1} M_X(t)}{dt^{n-1}} = \int_S x^{n-1} e^{tx} f_X(x)dx$$

Source with modifications:
https://onlinecourses.science.psu.edu/stat414/node/73/

- We check it holds for n:

$$\frac{d^n M_X(t)}{dt^n} = \frac{d\left[\frac{d^{n-1} M_X(t)}{dt^{n-1}}\right]}{dt} =$$

$$= \frac{d\left[\int_S x^{n-1} e^{tx} f_X(x) dx\right]}{dt} = \int_S x^n e^{tx} f_X(x) dx$$

QED

- This confirms:

$$\left[\frac{d^n M_X(t)}{dt^n}\right](t=0) = \mu_X(n) = E(X^n)$$

Source with modifications:
https://onlinecourses.science.psu.edu/stat414/node/73/

- Mgf and second moment:

$$\sigma_X^2 = E(X^2) - (E(X))^2 = \left[\frac{d^2 M_X(t)}{dt^2}\right](t=0) - \left\{\left[\frac{dM_X(t)}{dt}\right](t=0)\right\}^2$$

And if the mean is 0:

$$\sigma_X^2 = \left[\frac{d^2 M_X(t)}{dt^2}\right](t=0)$$

Source with modifications:

https://onlinecourses.science.psu.edu/stat414/node/73/

- Fundamental fact:

  > If the mgf for a random variable exists, it characterizes fully such random variable.

  Proof: omitted.
- It means that mgf and pdf are interchangeable
- We need now to determine the mgf for a normally distributed random variable $N(0, \sigma^2)$
- We will then use this to prove the CLT – LLf
- Let Z be a random variable, $Z \sim N(0,1)$ then, the mgf for Z is:

$$M_Z(t) = e^{\frac{1}{2}t^2}$$

- Proof

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

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

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

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

QED

Source with modifications:

https://www.le.ac.uk/users/dsgp1/COURSES/MATHSTAT/6normgf.pdf

Extending to the case of general Gaussian variables:

- Let X be a random variable, $X \sim N(\mu, \sigma_X^2)$, then the mgf for X is:

$$M_X(t) = e^{t\mu + \frac{1}{2}t^2 {\sigma_X}^2}$$

- We can first define $Z = \frac{X-\mu}{\sigma_X}$ and $Z \sim N(0,1)$

$$M_X(t) = E(e^{tX}) = E(e^{t(\mu+\sigma_X Z)}) = E(e^{t\mu}e^{t\sigma_X Z}) = e^{t\mu}E(e^{t\sigma_X Z}) =$$

$$= e^{t\mu}M_X(t\sigma_X) = e^{t\mu}e^{\frac{1}{2}t^2 {\sigma_X}^2} = e^{t\mu + \frac{1}{2}t^2 {\sigma_X}^2}$$

QED

Source with modifications:

https://www.quora.com/What-is-the-MGF-of-normal-distribution

The last piece of information that we miss are the following two properties:

- **Property 1: Moment of the Sum** Let $Y = \sum_{i=1}^{i=n} X_i$ where X_i are iid random variables then:

$$M_Y(t) = \prod_{i=1}^{i=n} M_{X_i}(t)$$

Proof:

$$M_Y(t) = E(e^{tY}) = E(e^{t\sum_{i=1}^{i=n} X_i}) = E(\prod_{i=1}^{i=n} e^{tX_i}) =$$

$$= \prod_{i=1}^{i=n} E(e^{tX_i}) = \prod_{i=1}^{i=n} M_{X_i}(t)$$

QED

Source with modifications: ***https://onlinecourses.science.psu.edu/stat414/node/170/***

[CLT – LLf] Mgf (10/10)

- **Property 2: Moment of the LC** Let $Y = a + bX$ where X is a random variable and $a, b \in \mathbb{R}, b \neq 0$ then:

$$M_Y(t) = e^{at} M_X(bt)$$

Proof:

$$M_Y(t) = E(e^{(a+bX)t}) = E(e^{at+bXt}) = E(e^{at}e^{bXt}) = e^{at}E(e^{bXt})$$

$$= e^{at}E(e^{btX}) = e^{at}M_X(bt)$$

QED

- Corollary: the sum of randomly iid Gaussian r.v. is still Gaussian.

Source with modifications: *https://onlinecourses.science.psu.edu/stat414/node/170/* *and* *https://www.stat.berkeley.edu/~mlugo/stat134-f11/clt-proof.pdf*

CLT – LLf – Proof (1/7)

- Remember that we want to prove that:

$$\mathfrak{Dn} \xrightarrow{d} N(0, \sigma^2)$$

- This is like proving that:

$$\frac{\mathfrak{Dn}}{\sigma} \xrightarrow{d} N(0, 1)$$

- We can rewrite $\mathfrak{Dn}/\sigma$:

$$\frac{\mathfrak{Dn}}{\sigma} = \frac{\sqrt{n}}{\sigma}(\overline{\mathfrak{Xn}} - \mu) = \frac{\sqrt{n}}{\sigma}\left[\frac{\sum_{i=1}^{i=n} \mathfrak{Xn}_i}{n} - \mu\right] = \frac{\sqrt{n}}{\sigma}\frac{\sum_{i=1}^{i=n} \mathfrak{Xn}_i - n\mu}{n} =$$

$$= \frac{\sum_{i=1}^{i=n} \mathfrak{Xn}_i - n\mu}{\sigma\sqrt{n}}$$

Source with modifications: ***https://www.stat.berkeley.edu/~mlugo/stat134-f11/clt-proof.pdf*** *and* ***https://en.wikipedia.org/wiki/Central_limit_theorem***

- Note: We can assume that $\mu = 0$. If it is not, we could define a new set of variables $\mathfrak{Y}_\mathfrak{i} = \mathfrak{X}_\mathfrak{i} - \mu$ and we would have that:

$$\sum_{i=1}^{i=n} \mathfrak{X}\mathfrak{n}_i - n\mu = \sum_{i=1}^{i=n} \mathfrak{Y}_\mathfrak{i}$$

  Preserving the same proof.

- Let now define $\mathfrak{W}_\mathfrak{n} = \mathfrak{D}\mathfrak{n}/\sigma$

$$\mathfrak{W}_\mathfrak{n} = \frac{\sum_{i=1}^{i=n} \mathfrak{X}\mathfrak{n}_i}{\sigma\sqrt{n}}$$

- We want to prove that $\mathfrak{W}_\mathfrak{n} \sim N(0,1)$ demonstrating that its moment is the same as the one of $N(0,1)$

Source with modifications: https://www.stat.berkeley.edu/~mlugo/stat134-f11/clt-proof.pdf *and* https://en.wikipedia.org/wiki/Central_limit_theorem

CLT – LLf – Proof (3/7)

- Note: We recall Property 1 (Slide 30) and 2 (Slide 31) about the momentum of combining random variables and we have:

$$M_{\mathfrak{D}\mathfrak{n}}(t) = \left[M_{\mathfrak{X}_\mathfrak{i}}(\frac{t}{\sqrt{n}})\right]^n$$

and likewise:

$$M_{\mathfrak{W}\mathfrak{n}}(t) = M_{\mathfrak{D}\mathfrak{n}}\left(\frac{t}{\sigma}\right) = \left[M_{\mathfrak{X}_\mathfrak{i}}\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$$

- In essence we need to evaluate the limit for n going to infinite of $\left[M_{\mathfrak{X}_\mathfrak{i}}\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$
- We want to prove that such limit is equal to the momentum of the standard normal distribution:

$$M_{N(0,1)}(t) = e^{\frac{1}{2}t^2}$$

Source with modifications: ***https://www.stat.berkeley.edu/~mlugo/stat134-f11/clt-proof.pdf*** *and* ***https://en.wikipedia.org/wiki/Central_limit_theorem***

- For simplicity we take the natural logarithm:

$$\ln\left[M_{\mathfrak{X}_i}\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n = n\ln\left[M_{\mathfrak{X}_i}\left(\frac{t}{\sigma\sqrt{n}}\right)\right]$$

- Now we define

$$q = \frac{1}{\sqrt{n}}$$

  Therefore n is $1/p^2$ and $n \to \infty \Rightarrow p \to 0$. This means that we want to compute:

$$\lim_{p\to 0} \frac{\ln M_{\mathfrak{X}_i}(\frac{tp}{\sigma})}{p^2} =$$

- This is an indeterminate form, so we can take the derivative of both side by the theorem of de l'Hôpital

Source with modifications: https://www.stat.berkeley.edu/~mlugo/stat134-f11/clt-proof.pdf and https://en.wikipedia.org/wiki/Central_limit_theorem

- This results to:

$$= \lim_{p\to 0} \frac{\frac{1}{M_{\mathfrak{x}_\mathfrak{i}}(\frac{tp}{\sigma})} \frac{dM_{\mathfrak{x}_\mathfrak{i}}(\frac{tp}{\sigma})}{dp} \frac{t}{\sigma}}{2p} = \frac{t}{2\sigma} \lim_{p\to 0} \frac{\frac{dM_{\mathfrak{x}_\mathfrak{i}}(\frac{tp}{\sigma})}{dp}}{pM_{\mathfrak{x}_\mathfrak{i}}(\frac{tp}{\sigma})} =$$

- This is again an indeterminate form, so we can take the derivative of both side by the theorem of de l'Hôpital

$$= \frac{t}{2\sigma} \lim_{p\to 0} \frac{\frac{d^2 M_{\mathfrak{x}_\mathfrak{i}}(\frac{tp}{\sigma})}{dp^2} \frac{t}{\sigma}}{M_{\mathfrak{x}_\mathfrak{i}}(\frac{tp}{\sigma}) + p \frac{dM_{\mathfrak{x}_\mathfrak{i}}(\frac{tp}{\sigma})}{dp} \frac{t}{\sigma}} = \frac{t^2}{2\sigma^2} \lim_{p\to 0} \frac{\frac{d^2 M_{\mathfrak{x}_\mathfrak{i}}(\frac{tp}{\sigma})}{dp^2}}{M_{\mathfrak{x}_\mathfrak{i}}(\frac{tp}{\sigma}) + p \frac{dM_{\mathfrak{x}_\mathfrak{i}}(\frac{tp}{\sigma})}{dp} \frac{t}{\sigma}}$$

- We now take the limits at numerator and denominator and we are done.

Source with modifications: https://www.stat.berkeley.edu/~mlugo/stat134-f11/clt-proof.pdf and https://en.wikipedia.org/wiki/Central_limit_theorem

CLT – LLf – Proof (6/7)

- Numerator:

$$\lim_{p\to 0} \frac{d^2 M_{\mathfrak{X}_i}(\frac{tp}{\sigma})}{dp^2} = \left[\frac{d^2 M_{\mathfrak{X}_i}(\frac{tp}{\sigma})}{dp^2}\right](0) = E({\mathfrak{X}_i}^2) =$$

$$= E(\mathfrak{X}_i)^2 + Var(\mathfrak{X}_i) = 0 + \sigma^2 = \sigma^2$$

- Denominator:

$$\lim_{p\to 0} \left[M_{\mathfrak{X}_i}(\frac{tp}{\sigma}) + p\frac{dM_{\mathfrak{X}_i}(\frac{tp}{\sigma})}{dp}\frac{t}{\sigma}\right] = M_{\mathfrak{X}_i}(0) + 0\left[\frac{dM_{\mathfrak{X}_i}(\frac{tp}{\sigma})}{dp}\frac{t}{\sigma}\right](0) =$$

$$= M_{\mathfrak{X}_i}(0) = 1$$

Source with modifications: *https://www.stat.berkeley.edu/~mlugo/stat134-f11/clt-proof.pdf* *and* *https://en.wikipedia.org/wiki/Central_limit_theorem*

CLT – LLf – Proof (7/7)

- And now we pull everything together and we obtain:

$$\lim_{p\to 0} \frac{\ln M_{\mathfrak{X}_\mathfrak{i}}(\frac{tp}{\sigma})}{p^2} = \frac{t^2}{2\sigma^2}\frac{\sigma^2}{1} = \frac{t^2}{2}$$

- And, therefore

$$\lim_{n\to +\infty} M_{\mathfrak{W}\mathfrak{n}}(t) = e^{\frac{1}{2}t^2}$$

QED

Source with modifications: ***https://www.stat.berkeley.edu/~mlugo/stat134-f11/clt-proof.pdf*** *and* ***https://en.wikipedia.org/wiki/Central_limit_theorem***

Status

- Now we know that the means of samples of a population tend to be distributed normally.
- This is an essential assumption to perform several numeric operations, like Montecarlo simulations, Bootstrap, etc.
- We can now understand the distribution of the Pearson momentum correlation coefficient of the sample
- Moreover, we have an open infinite issue on what to do if the data is NOT on a ratio scale
- This is an open issue for followup courses