CENTRAL LIMIT THEOREM
A simple proof

By
Alexander Esser
Suresh Kondati Natarajan

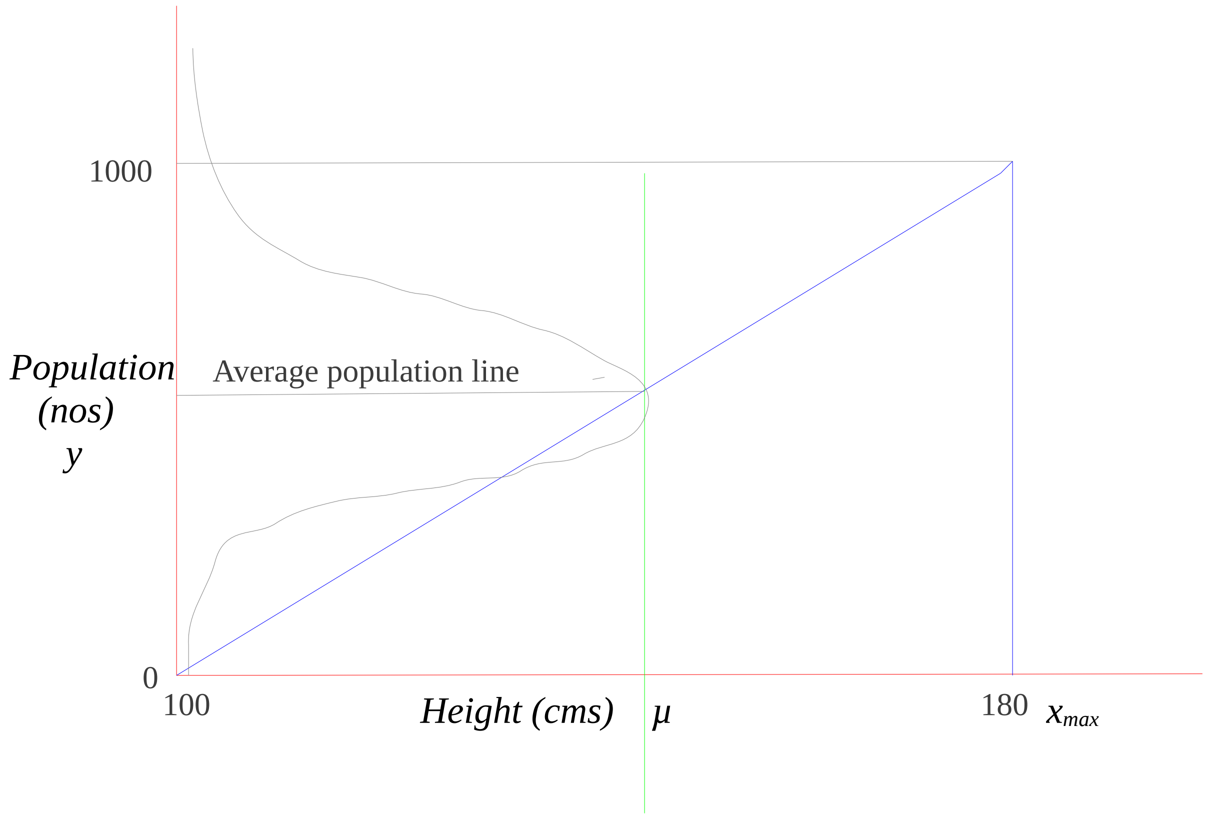
iMOS
Ruhr University – Bochum
Germany

Alexander Esser & Suresh K N

Central Limit Theorem:

*The distribution of random sampling of any quantity taken from any form of initial distribution will be **gaussian** in nature.*

*Consider a distribution with mean **μ** ,*



each value of x corresponds to a certain number of people, and our distribution here is in a ramp form.

Intuitive understanding:

*Lets say we program a **random number generator** that gives out a value between 100 and 180. Now, for every value of the random height (x) given by the program, we find the corresponding number of people (y). Importantly, The random number generator is supposed to be **unbiased** of any sort and assumed to give equal probability for every number to occur between the limits. We sum all the (y) values obtained and divide it by (N) to get the average. For every iteration, we get 1 average value from N individual values. We can **plot this point** in a graph with y axis (shown in the graph above) alone set to number of people (0 to 1000). After sufficient number of trials, when we look at the picture given by the histogram of (y) values, we find that the maximum concentration* *of this point is along the mean and it exponentially decays on either side. When we connect all edges of the bars of the histogram thus obtained, we find a **gaussian** distribution.*

Proof:

Now, lets do some random sampling and pick N number of x's from the distribution and add their population numbers(y) and sum them to get to get a value X.

$$X=\sum_{i=1}^{N} y_i \qquad (1)$$

*now, **X** is the sum of random sampling (1) and we need to find the probability distribution of this value X. We need to prove that it's a **gaussian** function, which is the central limit theorem.*

The distribution of X is denoted by $P_N(X)$.

let N=2,

$$X=(y_1+y_2) \quad (2)$$

$$y_2=X-y_1 \quad (3)$$

*now, the probability $P_2(X)$ will be given by a **convolution** of the probability functions of y_1 and y_2.*

$$P_2(X)=\int_{-\infty}^{+\infty} P(y_1).P(y_2)dy_1 dy_2 \quad (4)$$

substituting (3) in (4), we get

$$P_2(X)=\int_{-\infty}^{+\infty} P(y_1).P(X-y_1)dy_1 \quad (5)$$

*A short side note on **fourier transforms**, which is used to represent the problem in k-space where the small interactions in real world are amplified. Also, fourier transform of a convolution is the product of fourier transform of individual functions.*

f(x) $\xrightarrow{FT}$ *f(k) ; f(x)* $\xleftarrow{IFT}$ *f(k)*

similarly P(X) $\xrightarrow{FT}$ *Q(k);P(X)* $\xleftarrow{IFT}$ *Q(k)*
where, FT- fourier transform & IFT – Inverse fourier transforms.

$$Q(k)=\int_{-\infty}^{+\infty} P(X)e^{ikX}\,dX \quad (6)$$

look at any text on FT to clear how dk got replaced with dX (its a trivial step).

$$Q_2(k)=(Q(k))^2 \quad (7)$$ *from the side note about FT*

$$Q_2(k)=\left(\int_{-\infty}^{+\infty} P(X)e^{ikX}\,dX\right)^2 \quad (8)$$

here, $Q_2(k)$ is the ***moment generating function****, since, on deriving it continually with respect to k, we get all moments in X.....*

also, ***taylor expanding*** *e^{ikX} we get,*

$$e^{(ik(x))}=1+\frac{ikX}{1!}+\frac{(ik)^2X^2}{2!}+....... \quad (9)$$

$$Q_2(k)=\left[\int_{-\infty}^{+\infty}\left(1+\frac{ikX}{1!}+\frac{(ik)^2X^2}{2!}+.......\right)P(X)\,dX\right]^2 \quad (10)$$

now,

$$\int_{-\infty}^{+\infty} X\,P(X)\,dX=<X> \quad (11)$$

thus, the equation becomes (using (11))

$$Q_2(k)=\left(1+\frac{ik<X>}{1!}+\frac{(ik)^2<X^2>}{2!}+.......\right)^2 \quad (12)$$

here, the ***higher terms*** *do not contribute much,so, we can* ***neglect*** *them. For easier processing we can {exp (ln) } the while function*

$$Q_2(k)=\exp[\ln(1+\frac{ik<X>}{1!}+\frac{(ik)^2<X^2>}{2!}+.......)^2] \quad (13)$$

$$Q_2(k)=\exp[2\ln(1+\frac{ik<X>}{1!}+\frac{(ik)^2<X^2>}{2!}+.......)] \quad (14)$$

using the identity,

$$\ln(1+x)=x-\frac{x^2}{2}+....... \quad (15)$$

we get,

$$\ln(1+(\frac{ik<X>}{1!}+\frac{(ik)^2<X^2>}{2!}+.......))=ik<X>-\frac{(k)^2<X^2>}{2}-\frac{(ik)^2<X>^2}{2}-......... \quad (16)$$

we expanded only up to ***quadratic order*** *in* $<x>^2$.

$$...=ik<X>-\frac{(k)^2<X^2>}{2}+\frac{(k)^2<X>^2}{2}$$

$$...=ik<X>-\frac{k^2}{2}(<X^2>-<X>^2) \quad (17)$$

$$\mu=<X>;\sigma^2=<X^2>-<X>^2 \quad (18)$$

$$...=ik\mu-\frac{k^2\sigma^2}{2} \quad (19)$$

plugging (19) in (14) we obtain,

$$Q_2(k)=\exp[2(ik\mu-\frac{k^2\sigma^2}{2})] \quad (20)$$

This equation obtained is for N=2 , now ***generalizing*** *it, we get*

$$Q_N(k)=\exp[N(ik\mu-\frac{k^2\sigma^2}{2})] \quad (21)$$

now we can inverse transform $Q_N(k)$ *to obtain the probability* $P_N(X)$

$$P_N(X)=\frac{1}{2\pi}\int_{-\infty}^{+\infty} Q_N(k)e^{ikX}dk \quad (22)$$

$$P_N(X)=\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{[N(ik\mu-\frac{k^2\sigma^2}{2})]}e^{ikX}dk \quad (23)$$

$$P_N(X)=\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{[(ik(X-N\mu)-\frac{Nk^2\sigma^2}{2})]}dk \quad (24)$$

let $X\text{-}N\mu=X'$

$$-ikX'-\frac{Nk^2\sigma^2}{2} \quad \textit{can be written as} \quad \frac{-N\sigma^2}{2}[k-\frac{iX'}{N\sigma^2}]^2-\frac{X'^2}{2N\sigma^2}$$

$$P_N(X)=\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{(\frac{-N\sigma^2}{2}[k-i\frac{(X-N\mu)}{N\sigma^2}]^2-\frac{(X-N\mu)^2}{2N\sigma^2})}dk \quad (25)$$

keeping only the k dependent integral in and taking out all other constants,

$$P_N(X)=\frac{1}{2\pi}e^{\frac{-(X-N\mu)^2}{N\sigma^2}}\int_{-\infty}^{+\infty} e^{(\frac{-N\sigma^2k^2}{2})}dk \quad (26)$$

we know the solution of the integral $\int_{-\infty}^{+\infty} e^{(\frac{-N\sigma^2k^2}{2})}dk$ *which is* $\frac{\sqrt{\frac{\pi}{2}}}{\sqrt{N}\sigma}$

on substitution we get,

$$P_N(X)=\frac{1}{\sqrt{2\pi N}\sigma}e^{\frac{-(x-N\mu)^2}{N\sigma^2}} \quad (27)$$

here, this is a ***gaussian*** *distribution.*

This suggests that, if we took random sampling of N *values of* y *corresponding to* x *, then the distribution of the random variable (be it sum of N numbers or the average) will be of gaussian nature with the same mean μ (in case of averages plot).*